

# Differential Cross Section Measurement of the $pp \rightarrow WH \rightarrow WWW$ Process With the ATLAS Experiment

by

Moritz Hesping



Masterarbeit in Physik  
vorgelegt dem Fachbereich Physik, Mathematik und Informatik (FB 08)  
der Johannes Gutenberg-Universität Mainz  
am 10. November 2021

1. Gutachter: Prof. Dr. Volker Büscher
2. Gutachter: Prof. Dr. Stefan Tapprogge



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

Mainz, den 10.11.2021

Moritz Hesping  
ETAP  
Institut für Physik  
Staudingerweg 7  
Johannes Gutenberg-Universität D-55099 Mainz  
moritz.hesping@cern.ch

# Abstract

In this thesis, a Simplified Template Cross Section (STXS) analysis for the

$$pp \rightarrow WH \rightarrow WWW \rightarrow \nu\nu\nu\nu$$

process is presented, using final states with no lepton pairs of the same flavor and opposite charge. The analysis uses proton-proton collisions at a center of mass energy of 13 TeV with an integrated luminosity of  $139 \text{ fb}^{-1}$ , which were measured by the ATLAS detector at the Large Hadron Collider. Neural networks are used to reconstruct the transverse momentum of the associated  $W$  boson and to separate signal and background. The expected signal strengths of the STXS bins were measured to be  $\mu(0 \text{ GeV} \leq p_{\text{T}}^W < 75 \text{ GeV}) = 1.00 \pm 1.24$ ,  $\mu(75 \text{ GeV} \leq p_{\text{T}}^W < 150 \text{ GeV}) = 1.00 \pm 1.52$ ,  $\mu(150 \text{ GeV} \leq p_{\text{T}}^W < 250 \text{ GeV}) = 1.00 \pm 2.20$  and  $\mu(p_{\text{T}}^W \geq 250 \text{ GeV}) = 1.00 \pm 2.22$ .

# Contents

|                                                                        |           |
|------------------------------------------------------------------------|-----------|
| <b>1. Introduction</b>                                                 | <b>1</b>  |
| <b>2. The Standard Model</b>                                           | <b>3</b>  |
| 2.1. The Fundamental Forces . . . . .                                  | 3         |
| 2.2. The Elementary Particles . . . . .                                | 4         |
| 2.3. Parity Violation of the Weak Interaction . . . . .                | 5         |
| 2.4. QCD Phenomena at Hadron Colliders . . . . .                       | 5         |
| 2.4.1. Confinement and Hadronization . . . . .                         | 5         |
| 2.4.2. Running of the Strong Coupling Constant . . . . .               | 6         |
| 2.4.3. Proton Substructure and Parton Distribution Functions . . . . . | 7         |
| 2.5. The Gauge Principle and the Higgs Mechanism . . . . .             | 8         |
| 2.5.1. The Gauge Principle . . . . .                                   | 8         |
| 2.5.2. The Higgs Mechanism . . . . .                                   | 9         |
| 2.5.3. Properties of the Higgs Boson . . . . .                         | 10        |
| 2.6. Beyond the Standard Model . . . . .                               | 11        |
| <b>3. The ATLAS Detector at the LHC</b>                                | <b>13</b> |
| 3.1. The Large Hadron Collider . . . . .                               | 13        |
| 3.2. The ATLAS Experiment . . . . .                                    | 14        |
| 3.2.1. Inner Detector . . . . .                                        | 14        |
| 3.2.2. Calorimeters . . . . .                                          | 15        |
| 3.2.3. Muon System . . . . .                                           | 16        |
| 3.3. ATLAS Coordinate System and Kinematics . . . . .                  | 17        |
| 3.4. The ATLAS Trigger . . . . .                                       | 19        |
| 3.5. Object Reconstruction . . . . .                                   | 19        |
| 3.5.1. Tracks and Vertices . . . . .                                   | 19        |
| 3.5.2. Electrons . . . . .                                             | 20        |
| 3.5.3. Muons . . . . .                                                 | 20        |
| 3.5.4. Lepton Isolation . . . . .                                      | 21        |
| 3.5.5. Jets . . . . .                                                  | 21        |
| 3.5.6. Missing Transverse Energy . . . . .                             | 21        |
| <b>4. Statistical Methods</b>                                          | <b>22</b> |
| 4.1. Monte Carlo Simulations . . . . .                                 | 22        |
| 4.2. Maximum Likelihood Fits . . . . .                                 | 23        |
| 4.3. Hypothesis Tests and Upper Limits . . . . .                       | 24        |

## Contents

|           |                                                                       |           |
|-----------|-----------------------------------------------------------------------|-----------|
| 4.4.      | Neural Networks . . . . .                                             | 26        |
| 4.4.1.    | Nodes, Layers and Networks . . . . .                                  | 26        |
| 4.4.2.    | Activation Functions . . . . .                                        | 27        |
| 4.4.3.    | Training . . . . .                                                    | 27        |
| 4.4.4.    | Hyperparameter Optimization . . . . .                                 | 28        |
| 4.4.5.    | Overfitting and Regularization . . . . .                              | 28        |
| 4.4.5.1.  | Early Stopping . . . . .                                              | 29        |
| 4.4.5.2.  | Dropout . . . . .                                                     | 29        |
| 4.4.6.    | Validation and Testing Samples . . . . .                              | 30        |
| 4.5.      | Regression Networks . . . . .                                         | 31        |
| 4.6.      | Multiclassifier Networks . . . . .                                    | 31        |
| <b>5.</b> | <b>Signal and Background Processes</b>                                | <b>32</b> |
| 5.1.      | The Signal Process . . . . .                                          | 32        |
| 5.2.      | Spin Correlation . . . . .                                            | 32        |
| 5.3.      | Simplified Template Cross Sections . . . . .                          | 34        |
| 5.4.      | Background Processes . . . . .                                        | 34        |
| 5.4.1.    | $WZ/\gamma^*$ and Electron Charge Misidentification . . . . .         | 35        |
| 5.4.2.    | Top Pair Production and Fake Leptons . . . . .                        | 36        |
| 5.4.3.    | Triboson . . . . .                                                    | 37        |
| 5.5.      | Previous Measurement of the Signal Process . . . . .                  | 38        |
| <b>6.</b> | <b>Reconstruction of the <math>W</math> Boson Momentum</b>            | <b>39</b> |
| 6.1.      | Degrees of Freedom and Analytical Reconstruction . . . . .            | 39        |
| 6.2.      | Proxies . . . . .                                                     | 39        |
| 6.3.      | Regression Network for $p_T^W$ Reconstruction . . . . .               | 42        |
| 6.3.1.    | Training Sample . . . . .                                             | 42        |
| 6.3.2.    | Preprocessing . . . . .                                               | 42        |
| 6.3.3.    | Training and Hyperparameter Optimization . . . . .                    | 43        |
| 6.3.4.    | Input Variables . . . . .                                             | 43        |
| 6.3.5.    | Architecture . . . . .                                                | 44        |
| 6.3.6.    | Regression Network Performance . . . . .                              | 44        |
| 6.3.7.    | Multiclassifier for STXS Bins . . . . .                               | 46        |
| <b>7.</b> | <b>Analysis in the <math>Z</math> Depleted Signal Region</b>          | <b>48</b> |
| 7.1.      | Monte Carlo Samples . . . . .                                         | 48        |
| 7.2.      | $Z$ Depleted Signal Region Selections . . . . .                       | 49        |
| 7.3.      | Control Regions . . . . .                                             | 49        |
| 7.4.      | Multiclassifier for Background Separation . . . . .                   | 51        |
| 7.4.1.    | Inputs . . . . .                                                      | 51        |
| 7.4.2.    | Architecture and Training Setup . . . . .                             | 52        |
| 7.5.      | Treatment of Multiclassifier and Regression Network Outputs . . . . . | 52        |
| 7.6.      | Fits With Backgrounds . . . . .                                       | 58        |
| 7.7.      | Suppression of the Remaining $WZ$ Background . . . . .                | 59        |

## *Contents*

|                                                       |           |
|-------------------------------------------------------|-----------|
| 7.8. Retraining of the Multiclassifier . . . . .      | 61        |
| 7.8.1. Motivation . . . . .                           | 61        |
| 7.8.2. Retraining Setup . . . . .                     | 65        |
| 7.8.3. Comparison of the Training Variants . . . . .  | 65        |
| 7.9. $WW$ Validation Region . . . . .                 | 70        |
| 7.10. $WZ$ Scale Uncertainties . . . . .              | 71        |
| 7.11. Distribution of a Possible BSM Signal . . . . . | 73        |
| 7.12. Binning of the Retrained Classifier . . . . .   | 76        |
| 7.13. Results . . . . .                               | 79        |
| <b>8. Summary and Outlook</b>                         | <b>82</b> |
| <b>A. Neural Network Input Variables</b>              | <b>86</b> |
| <b>B. Cutflows</b>                                    | <b>92</b> |
| <b>References</b>                                     | <b>97</b> |

# 1. Introduction

The Standard Model of Particle Physics is among the most successful scientific theories ever developed. It aims to describe the laws of nature at the most fundamental level, i.e. the most basic constituents of matter and the interactions between them. In the decades since its development in the 1960s, its numerous predictions have been experimentally verified with great precision.

The most recent triumph of the Standard Model was the discovery of the Higgs boson. In 1964, Higgs [1], Englert and Brout [2] and others (e.g. [3]) had proposed a mechanism by which the fundamental particles, most importantly the  $W$  and  $Z$  bosons mediating the weak interaction, can acquire mass without violating fundamental symmetries of the theory. This Higgs mechanism also leads to the prediction of a massive spin 0 particle, called the Higgs boson. Almost fifty years later in 2012, the ATLAS [4] and CMS [5] collaborations at the Large Hadron Collider [6] announced the discovery of a new particle consistent with the predicted mass range of the Higgs boson [7, 8]. Later, measurements of other properties [9] of this new particle such as its spin confirmed its identity as a Higgs boson.

In spite of its tremendous success, the Standard Model is known to be an incomplete theory. Several well established phenomena, such as the matter-antimatter asymmetry of the universe and the existence of dark matter, have yet to be reconciled with the Standard Model. For this reason, numerous ongoing efforts in particle physics are dedicated to searching for new physics beyond the Standard Model, as well as scrutinizing the Standard Model in precision measurements, as discrepancies between measurement and theory would point toward such new physics.

Due to its central role in the Standard Model and extensions thereof, the properties of the Higgs boson are an important subject for such measurements. Its couplings to the mediators of the weak interaction are of special interest, as they play a key role in the Higgs mechanism. The process  $pp \rightarrow WH \rightarrow WWW$  in collisions at the Large Hadron Collider presents an opportunity for a measurement of this coupling: this process is dominated by a Feynman graph in which the Higgs boson couples only to  $W$  bosons.

In general, the effects of new physics are expected to affect the dependence of the physical processes on kinematic variables, such as the transverse momentum of the associated  $W$  boson in the above reaction. Measurements that are sensitive to this dependence are therefore more sensitive to new physics than inclusive measurements. In

## 1. Introduction

this thesis, such a measurement was developed for the  $pp \rightarrow WH \rightarrow WWW \rightarrow l\nu l\nu l\nu$  process in the Simplified Template Cross Section (STXS) scheme [10].

In chapters 2 through 4 an overview of the theoretical background, the ATLAS detector and the statistical tools used for the analysis is provided. In chapter 5 the signal process and the most important background processes are introduced. In chapter 6 the reconstruction of the momentum of the associated  $W$  boson is described, which is one of the main challenges of the STXS analysis. The other aspects of the analysis, such as the suppression of backgrounds and studies regarding the effect of systematical uncertainties and the sensitivity to physics beyond the Standard Model are presented in chapter 7. Finally, in chapter 8 the findings are summarized and an outlook is given.

## 2. The Standard Model

The Standard Model (SM) of particle physics, which was developed over the course of the 20th century, is the theory describing the fundamental building blocks of matter and the interactions between them. It is the basis of most of modern particle physics. This chapter gives a brief introduction to the Standard Model, based on the text books by Griffiths [11] and Thomson [12].

### 2.1. The Fundamental Forces

There are four fundamental interactions in nature: The electromagnetic force, the strong force, the weak force and gravity. The first three are described by the Standard Model, while gravity is described by the theory of general relativity [13], and no quantum description for it has been found as of yet.

The relative strength of the interactions and the ranges over which they take effect differ greatly. Electromagnetism and gravity are of unlimited range and thus play a role in macroscopic systems. The weak and strong forces on the other hand only work on very short scales of about 1 fm and below. The relative strengths of the forces are shown in Table 2.1. As the names suggest, at the energy and distance scales found in experiments, the strong and weak force are significantly stronger and weaker than electromagnetism, respectively. Gravity is so weak as to not play a role on microscopic scales.

Table 2.1.: The approximate relative strengths of the fundamental forces at a distance of  $10^{-15}$  m [12].

| Force            | Relative Strength |
|------------------|-------------------|
| Strong           | 1                 |
| Electromagnetism | $10^{-3}$         |
| Weak             | $10^{-8}$         |
| Gravity          | $10^{-37}$        |

## 2. The Standard Model

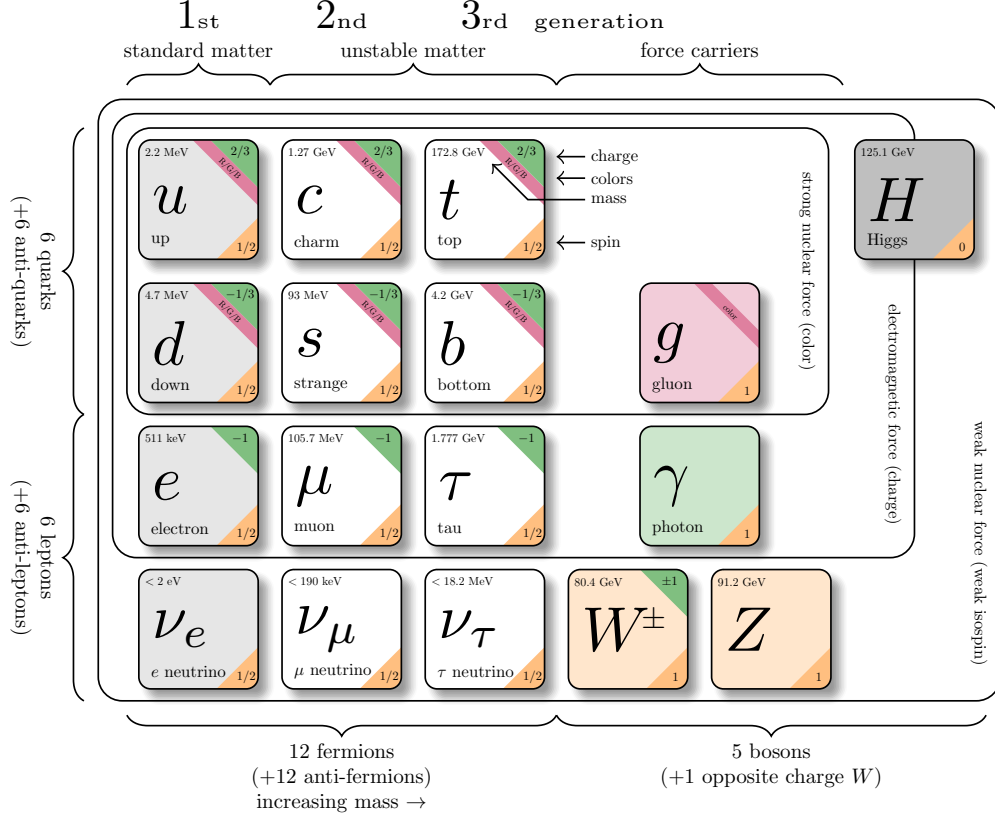


Figure 2.1.: The particles of the Standard Model. Adapted from [14], particle masses from [9].

### 2.2. The Elementary Particles

The elementary particles according to the Standard Model are divided into two groups, the fermions with spin  $\frac{1}{2}$  and the bosons with spin 0 or 1. The fermions are the constituents of matter, while the spin 1 gauge bosons are force carriers for the three fundamental forces of nature that are explained by the Standard Model. The spin 0 Higgs boson  $H$  plays a special role, which is explained in more detail in Section 2.5. An overview of the elementary particles is shown in Fig. 2.1.

The fermions can be further separated into quarks and leptons. Quarks are the constituents of protons, neutrons and other hadrons. They carry a color charge, which means they participate in the strong interaction. They also carry electric charge with a magnitude of either  $\frac{2}{3}$  or  $\frac{1}{3}$  elementary charges. Leptons do not carry a color charge and thus do not experience the strong force. The neutrinos, which belong to the leptons, are also electrically neutral and are only affected by the weak interaction.

## 2. The Standard Model

Fermions come in three generations, which only differ by the particle masses. The first generation, being the lightest, contains the particles making up all stable matter: The up quark  $u$ , the down quark  $d$  and the electron  $e$ . The heavier versions of the up quark are called the charm quark  $c$  and the top quark  $t$ . They are collectively known as up-type quarks. Similarly, the heavy down-type quarks are the strange quark  $s$  and the bottom quark  $b$ . The heavy charged leptons are the muon  $\mu$  and the tau lepton  $\tau$ . Each generation also contains a neutrino corresponding to the charged lepton:  $\nu_e$ ,  $\nu_\mu$  and  $\nu_\tau$ , respectively. Every fermion also has a corresponding anti-particle, which has the same mass, but opposite charge. An anti-up quark for example carries charge  $-\frac{2}{3}$  and an anti-color.

Each gauge boson is a force carrier that is associated with one of the fundamental forces of nature. The photon  $\gamma$  is the mediator of the electromagnetic interaction. It is charge- and massless. The gluon  $g$  carries the strong force. It also has no mass and no electric charge, but does carry color charge, one of eight possible combinations of color and anti-color. The weak interaction is mediated by the massive  $W$  and  $Z$  bosons. The  $W$  has a charge of  $\pm 1$ , while the  $Z$  is electrically neutral.

### 2.3. Parity Violation of the Weak Interaction

Charged current weak interactions, i.e. those in which a  $W$  boson is exchanged, maximally violate parity symmetry, the symmetry under spatial inversion. The  $W$  boson exclusively couples to fermions with the correct chirality, left-chiral fermions and right-chiral antifermions. These are eigenstates of the left-handed projection operator  $P_L$ . For massless particles, being left-chiral corresponds to having negative helicity and being right-chiral to positive helicity. Helicity is the projection of the particle spin on the momentum vector. Neutral current weak reactions, in which a  $Z$  is exchanged, also violate parity, but not maximally. The  $Z$  boson couples to particles of both chiralities, but with different strengths.

### 2.4. QCD Phenomena at Hadron Colliders

Particle physics experiments with proton-proton collisions are affected by two important phenomena of the strong interaction: Hadronization and the proton substructure. Because of their importance for ATLAS and the other experiments at the Large Hadron Collider, they are explained in more detail in the following.

#### 2.4.1. Confinement and Hadronization

No free quarks have ever been observed. This is explained by the concept of color confinement, which states that only color neutral objects can propagate freely, and

## 2. The Standard Model

particles with a color charge are restrained to color neutral bound states. This phenomenon is thought to be a result of gluon self interactions. They concentrate the gluon field into a tube between the quarks. This results in a potential of the form

$$V(r) \sim \kappa r, \quad (2.1)$$

where  $\kappa$  has been measured to be of the order of  $1 \text{ GeV fm}^{-1}$ . Because of the linear increase in the potential, it is energetically favorable when two quarks are pulled apart to produce a quark-antiquark pair, with which the original quarks can form color neutral pairs.

When high energy quarks or gluons are produced in a collider, quark pairs are repeatedly produced, until all quarks have low enough momentum to settle into bound states. This process is called hadronization. It results in jets of hadrons, which are detected in the experiment. In general there is one such jet for each initial high energy particle, pointing in the direction of the particle's momentum. In practice, though, this depends on the reconstruction algorithms used in a given experiment.

### 2.4.2. Running of the Strong Coupling Constant

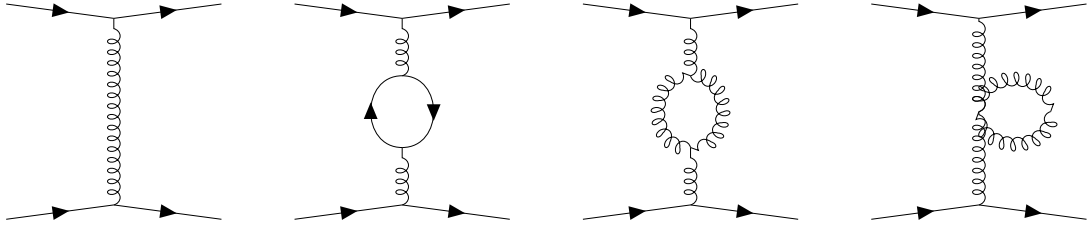


Figure 2.2.: Gluon propagators with up to one loop.

The strong coupling constant  $\alpha_S$  exhibits a running behavior: Its value depends on the momentum transfer  $q^2$  in the reaction. This is a consequence of renormalization: Higher order diagrams involve loops in the gluon propagator, such as the ones shown in Fig. 2.2. These loops lead to divergences in the matrix element, which need to be absorbed into the coupling constant. This results in the equation

$$\alpha_S(q^2) = \frac{\alpha_S(\mu^2)}{1 + B\alpha_S(\mu^2) \ln\left(\frac{q^2}{\mu^2}\right)}, \quad (2.2)$$

where  $\mu^2$  is the so-called renormalization scale. If the matrix element were calculated up to infinite order, the choice of renormalization scale would not affect the prediction. Since this is clearly impossible, however, the effect of varying choices of the renormalization scale needs to be taken into account as a systematic uncertainty.

## 2. The Standard Model

$$B = \frac{11N_c - 2N_f}{12\pi} \quad (2.3)$$

is a constant dependent on the number of colors  $N_c$  and the number of quark flavors  $N_f$ . Since it is greater than zero, the strong coupling constant decreases at higher momenta. This property, known as asymptotic freedom, is of great importance in high energy physics: At lower momenta it is impossible to use perturbation theory for QCD calculations, since  $\alpha_S$  has values greater than one. At the momenta typically reached at colliders like the Large Hadron Collider, however, the coupling constant becomes small enough that e.g. cross sections can be calculated from Feynman diagrams.

### 2.4.3. Proton Substructure and Parton Distribution Functions

The proton is a composite particle. While in the simplest description it consists of two up quarks and a down quark, with the charges adding up to the proton charge of 1, the true substructure of the proton is more complex. Apart from the three valence quarks, it contains gluons as well as quarks and antiquarks produced in pairs, known as sea quarks. Quarks and gluons are collectively known as partons.

When the momentum transfer in the scattering of two protons is sufficiently high, due to asymptotic freedom the process can be described as the scattering of one parton each from the protons. The properties of this reaction depend on the types of partons involved and on the momenta of the partons, given as a fraction  $x$  of the proton momentum. This information is encoded in the Parton Distribution Functions (PDF). They are defined such that  $f(x)\delta x$  represents the number of partons of a given type with a momentum fraction between  $x$  and  $x + \delta x$ . Since the PDFs result from the strong interactions within the proton, they cannot be calculated directly and instead need to be measured experimentally. Fig. 2.3 shows the measured PDFs of different types of partons.

Before a parton takes part in a scattering process it may radiate gluons, thereby changing its momentum [9]. The gluon momenta are mostly collinear to the parton. Since these emissions do not depend on the scattering interaction, they can be taken to be modifications to the proton structure. This procedure is called collinear factorization and involves the choice of a so-called factorization scale, below which emissions are absorbed into the PDF. Similarly to the renormalization scale, the particular choice of factorization scale affects the physical prediction of the calculation, because it is only carried out to finite order, and needs to be varied to estimate the systematical uncertainty caused by the choice of scale.

## 2. The Standard Model

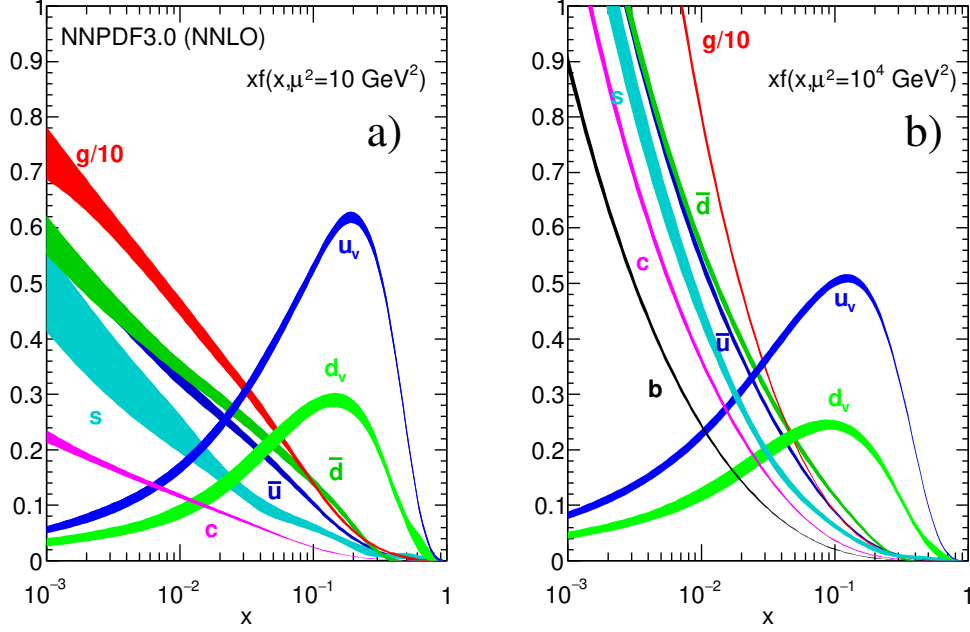


Figure 2.3.: Parton distribution functions at two different scales [9].

## 2.5. The Gauge Principle and the Higgs Mechanism

### 2.5.1. The Gauge Principle

The three fundamental forces described by the Standard Model are gauge theories, which means they can be derived by making the Lagrangian of the free fermions invariant under a local gauge transformation. This is important because gauge theories are renormalizable, as shown by 't Hooft [15].

The electromagnetic interaction, for instance, is a result of a local U(1) symmetry, i.e. a phase transformation

$$\psi(x) \rightarrow \psi'(x) = e^{iQ\chi(x)}\psi(x), \quad (2.4)$$

where  $\psi(x)$  is the fermion field and  $\chi(x)$  is the local phase. Since the derivative in the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi. \quad (2.5)$$

acts on this phase,  $\mathcal{L}$  is only invariant under the transformation after replacing

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iQeA_\mu, \quad (2.6)$$

where  $A_\mu$  is a new field with properties consistent with those of the photon and  $Q$  can now be understood as the electric charge of the fermion in units of  $e$ . Similarly, the strong interaction can be derived from the SU(3) symmetry corresponding to color and the electroweak interaction from the SU(2)<sub>L</sub>  $\otimes$  U(1)<sub>Y</sub> symmetry of weak isospin

## 2. The Standard Model

and hypercharge.

### 2.5.2. The Higgs Mechanism

The local gauge principle is theoretically well motivated and its predictions, for example in the electroweak sector, have been confirmed by experiments. However, particle mass terms break the local gauge symmetry. Therefore a mechanism has to be found to introduce these masses to the Lagrangian without breaking the gauge symmetry.

The problem of the particle masses is solved by the Higgs mechanism using spontaneous symmetry breaking. A new scalar field, the Higgs field  $\phi$ , is introduced. It is subject to the potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (2.7)$$

In the case of  $\lambda > 0$  and  $\mu^2 < 0$ , shown in Fig. 2.4, the minima of this potential satisfy  $\phi^\dagger \phi = -\frac{\mu^2}{2\lambda}$ , which means that the Higgs field has a non-zero vacuum expectation value.

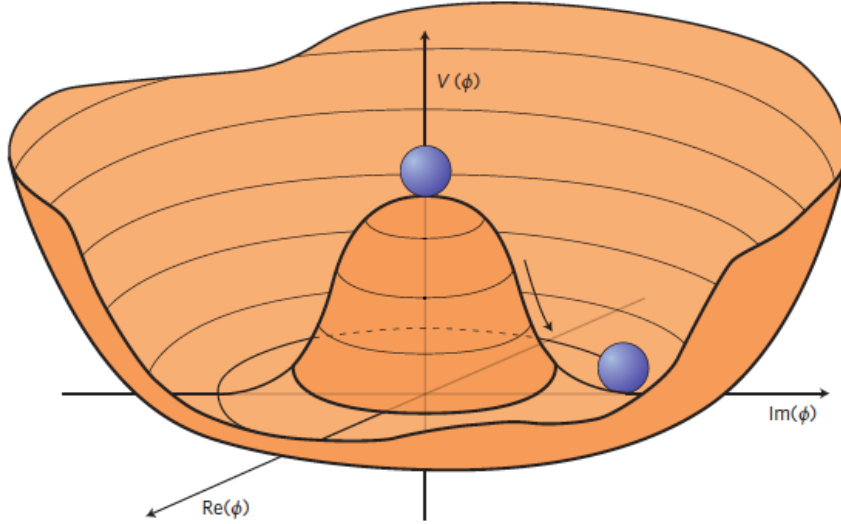


Figure 2.4.: The shape of the Higgs potential, leading to spontaneous symmetry breaking [16].

Applying the gauge principle with the  $SU(2)_L \otimes U(1)_Y$  symmetry of the electroweak interaction leads to interactions of the electroweak fields  $W_\mu^{(1)}$ ,  $W_\mu^{(2)}$ ,  $W_\mu^{(3)}$  and  $B_\mu$  to the vacuum state of the Higgs field as well as to its excitations about the vacuum state, which correspond to a new particle, the Higgs boson. The interactions with the vacuum state are then the mass terms of the gauge bosons. The breaking of the electroweak symmetry through the Higgs mechanism furthermore leads to the mixing of the  $W_\mu^{(3)}$  and  $B_\mu$  fields, resulting in the physical photon and  $Z$  boson.

## 2. The Standard Model

### 2.5.3. Properties of the Higgs Boson

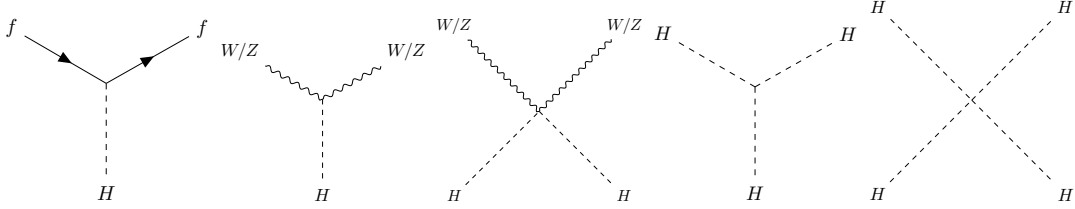


Figure 2.5.: Couplings of the Higgs boson.

The SM Higgs boson is a neutral scalar particle. Its mass was measured to be  $m_H = 125.1$  GeV [9]. The Higgs boson couples to all massive particles, with a coupling strength proportional to the fermion masses and the square of the boson masses. It has triple couplings to fermions and triple and quartic couplings to bosons. It also has triple and quartic self couplings, as shown in Fig. 2.5.

Since the coupling is proportional to mass, the Higgs boson preferentially decays into heavy particles. The most common decay is  $H \rightarrow b\bar{b}$ , since the  $b$  quark is the heaviest particle with  $2m < m_H$ . The second most common decay is  $H \rightarrow WW^*$ , where  $W^*$  is an off-shell  $W$  boson. The process is suppressed by the  $W$  being off-shell, but this is compensated for by the strength of the coupling to the large  $W$  mass. The Higgs can also decay to gluon or photon pairs via a virtual loop containing a heavy particle, as shown in Fig. 2.6. The leading decay processes and their branching fractions are listed in Table 2.2.

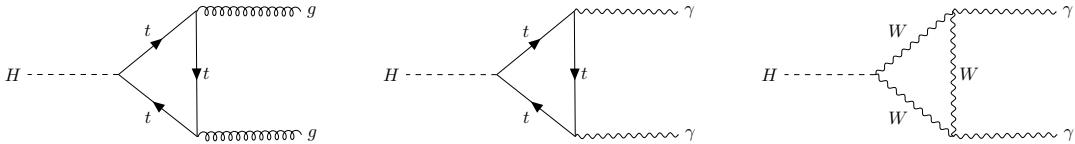


Figure 2.6.: Higgs decays to gluons and photons.

Table 2.2.: Predicted branching fractions of the leading Higgs boson decays for  $m_H = 125$  GeV [17].

| Decay Mode                   | Branching Fraction |
|------------------------------|--------------------|
| $H \rightarrow b\bar{b}$     | 58.2 %             |
| $H \rightarrow WW^*$         | 21.4 %             |
| $H \rightarrow gg$           | 8.2 %              |
| $H \rightarrow \tau^+\tau^-$ | 6.3 %              |
| $H \rightarrow c\bar{c}$     | 2.9 %              |
| $H \rightarrow ZZ^*$         | 2.6 %              |
| $H \rightarrow \gamma\gamma$ | 0.2 %              |

## 2. The Standard Model

There are four main processes in which Higgs bosons are produced in proton-proton collisions, such as at the LHC. Again, these are processes in which the Higgs boson couples to the heaviest particles, i.e. top quarks and the weak gauge bosons. In descending order of cross-section, they are gluon fusion (ggF), Vector Boson Fusion (VBF), associated production with a vector boson (VH), and associated production with a  $t\bar{t}$  pair (ttH). The Feynman graphs for these processes are shown in Fig. 2.7 and the cross-sections are listed in Table 2.3. These cross-sections depend on the parton distribution functions, since those contain the probability of a gluon or quark pair being involved in the reaction and carrying enough momentum to produce the particles in the final state.

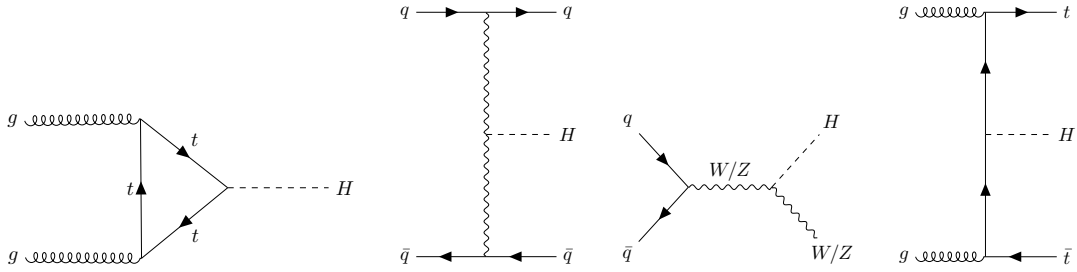


Figure 2.7.: Higgs production processes at the LHC. From left to right: ggF, VBF, VH, ttH.

Table 2.3.: Predicted cross sections of different Higgs boson production channels at the LHC [9].

| Production Process | Cross-Section at $\sqrt{s} = 13$ TeV |
|--------------------|--------------------------------------|
| ggF                | 48.6 pb                              |
| VBF                | 3.78 pb                              |
| VH                 | 1.37 pb                              |
| ttH                | 0.50 pb                              |

## 2.6. Beyond the Standard Model

While the Standard Model is a highly successful and well tested theory, there are a number of phenomena that are not explained by it. One famous example is dark matter. Several astronomical phenomena such as galaxy rotation curves and gravitational lensing would be explained by dark matter, but no suitable particle has been directly detected or produced in a collider. Other examples include gravity, dark energy, neutrino masses and the matter-antimatter asymmetry of the universe.

Many ongoing efforts in particle physics aim to find evidence of such Beyond the Standard Model (BSM) physics. In a collider experiment, BSM effects mainly manifest

## 2. The Standard Model

in the production of new particles or, if the energy of the collider is insufficient to produce such new particles, as modifications to cross sections, for example through the contributions of new virtual particles to the matrix element.

BSM effects can be described by adding additional interaction terms to the Lagrangian. Each interaction term is scaled with a parameter  $\kappa$  that describes the strength of that coupling. One such effective Lagrangian is given by the Higgs characterization model [18], which modifies the couplings between the Higgs boson and the vector bosons, as

$$\begin{aligned}
\mathcal{L} = & \{ \cos \alpha \kappa_{SM} [\frac{1}{2} g_{HZZ} Z_\mu Z^\mu + g_{HWW} W_\mu^+ W^{-\mu} \\
& - \frac{1}{4} [\cos \alpha \kappa_{H\gamma\gamma} g_{H\gamma\gamma} A_{\mu\nu} A^{\mu\nu} + \sin \alpha \kappa_{A\gamma\gamma} g_{A\gamma\gamma} A_{\mu\nu} \tilde{A}^{\mu\nu}] \\
& - \frac{1}{2} [\cos \alpha \kappa_{HZ\gamma} g_{HZ\gamma} Z_{\mu\nu} A^{\mu\nu} + \sin \alpha \kappa_{AZ\gamma} g_{AZ\gamma} Z_{\mu\nu} \tilde{A}^{\mu\nu}] \\
& - \frac{1}{4} [\cos \alpha \kappa_{Hgg} g_{Hgg} G_{\mu\nu}^a G^{a,\mu\nu} + \sin \alpha \kappa_{Agg} g_{Agg} G_{\mu\nu}^a \tilde{G}^{a,\mu\nu}] \\
& - \frac{1}{4\Lambda} [\cos \alpha \kappa_{HZZ} Z_{\mu\nu} Z^{\mu\nu} + \sin \alpha \kappa_{AZZ} Z_{\mu\nu} \tilde{Z}^{\mu\nu}] \\
& - \frac{1}{2\Lambda} [\cos \alpha \kappa_{HWW} W_{\mu\nu}^+ W^{-\mu\nu} + \sin \alpha \kappa_{AWW} W_{\mu\nu}^+ \tilde{W}^{-\mu\nu}] \\
& - \frac{1}{\Lambda} \cos \alpha [\kappa_{H\partial Z} Z_\nu \partial_\mu Z^{\mu\nu} + (\kappa_{H\partial W} W_\nu^+ \partial_\mu W^{-\mu\nu} + \text{h.c.})] \} H.
\end{aligned} \tag{2.8}$$

Here,  $\Lambda$  is the scale of the new physics. In the Standard model case,  $\cos \alpha = 1$  and  $\kappa_{SM} = 1$ , while all other parameters are 0.

## 3. The ATLAS Detector at the LHC

### 3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [6] is a proton-proton collider located near Geneva, Switzerland. It is the world's highest energy particle accelerator, with a center of mass energy of 13 TeV achieved during its second run of proton collisions from 2015 to 2018. It also has a high luminosity, with a peak luminosity of  $2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  and integrated luminosity of  $160 \text{ fb}^{-1}$  over that time span [19].

The LHC uses the 27.6 km long tunnel originally constructed for the Large Electron Positron collider (LEP) [20]. It has two separate beam pipes, which enables it to collide protons or heavy ions of the same charge. The collisions happen at four different interaction points, where the beams cross. Correspondingly, there are four major experiments located at these points:

- ATLAS (A Toroidal LHC ApparatuS) is a general purpose detector. It is detailed in Section 3.2.
- CMS (Compact Muon Solenoid) is another multi-purpose detector. Its distinguishing design features are a strong solenoid magnet, a fully silicon based inner tracker and an electromagnetic calorimeter based on homogeneous scintillating crystals [5].
- LHCb is a detector purpose built for heavy flavor physics. It is a single arm spectrometer and only covers angles close to the beam axis, as b-hadrons are predominantly produced in the forward cone [21].
- ALICE (A Large Ion Collider Experiment) is a detector for heavy ion collisions. It features a time of flight detector and a ring imaging Cherenkov counter for identification of hadrons [22].

### 3.2. The ATLAS Experiment

The ATLAS experiment [4] is a general purpose detector. It is backward-forward symmetric and has eight-fold symmetry around the beam axis. It consists of three subsystems. From the interaction point outward these are the inner detector, the calorimeters and the muon system. The different detectors are also divided into the barrel region in the center of the detector and end-cap regions at the ends of the detector. A cut-away drawing of the ATLAS detector is shown in Fig. 3.1.

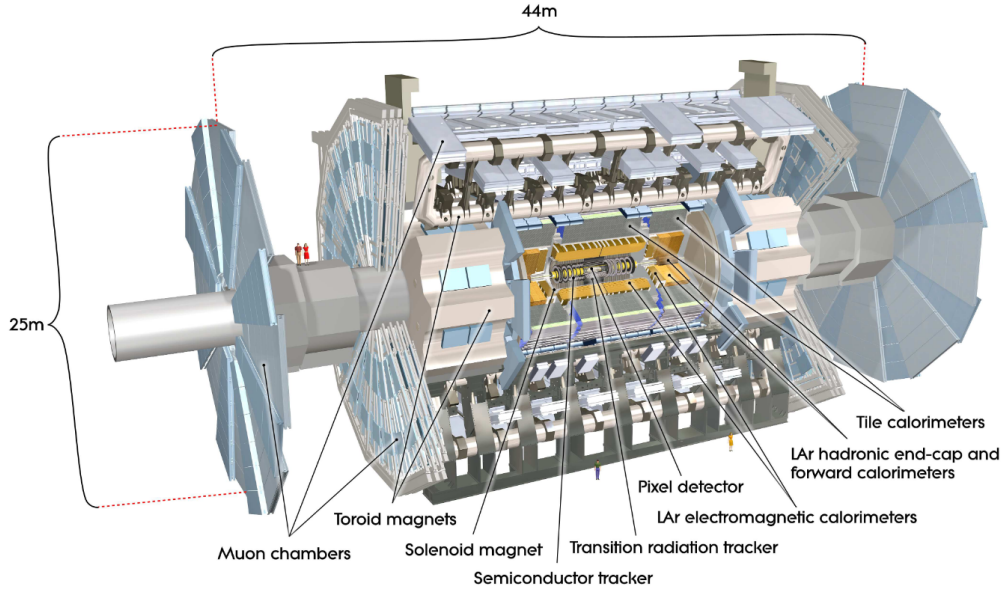


Figure 3.1.: Cut-away view of the ATLAS detector [4].

#### 3.2.1. Inner Detector

The inner detector, shown in Fig. 3.2, contains the tracking sensors, which are used for the reconstruction of charged particle tracks. They are surrounded by a 2 T solenoid magnet, which forces the particles onto helical trajectories to enable momentum reconstruction and charge identification.

The layer of the detector closest to the interaction point is the pixel detector, which has the highest granularity. It is surrounded by the semiconductor tracker. Both of these trackers can measure the position of a particle hit in both  $(R - \phi)$  and  $z$ . They are further surrounded by the straw tube Transition Radiation Tracker (TRT), which detects hits only in  $(R - \phi)$ . The TRT is used in particle identification: the intensity of the transition radiation is proportional to  $\gamma = E/m$  and can therefore be used to infer the mass of the particle [23].

### 3. The ATLAS Detector at the LHC

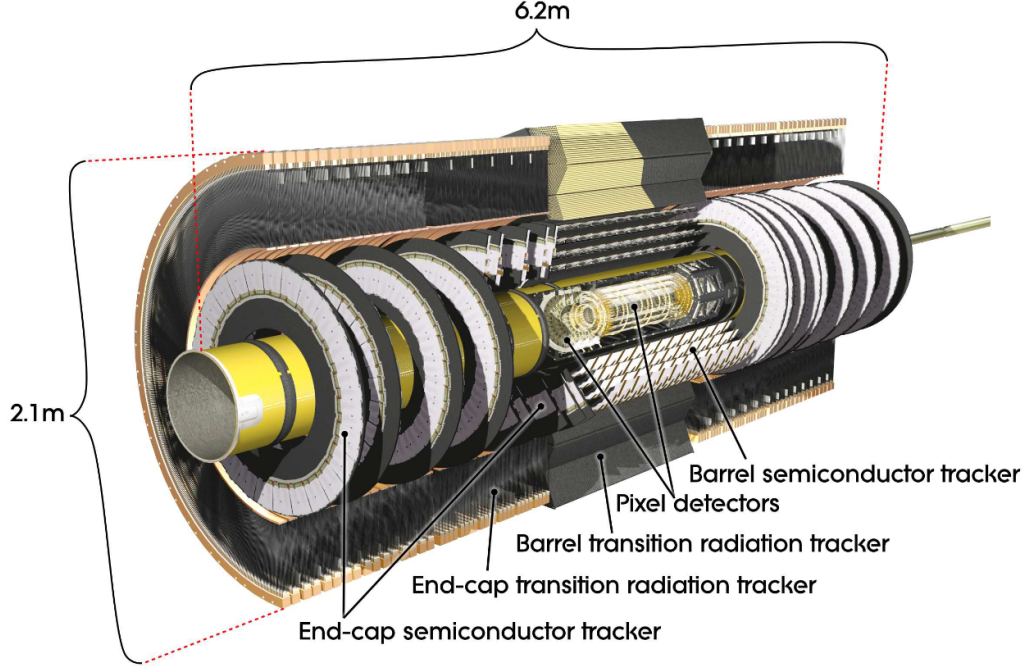


Figure 3.2.: Cut-away view of the ATLAS inner detector [4].

#### 3.2.2. Calorimeters

The inner detector is followed by the calorimeters, which measure particle energies. An illustration of the calorimeters is shown in Fig. 3.3. The innermost of these is the Liquid Argon (LAr) electromagnetic calorimeter. It has the highest granularity of the calorimeters and serves the precision measurement of photon and electron energies. The electromagnetic calorimeter is followed by the hadronic calorimeters: in the end cap region, the energy of hadrons is measured by the LAr hadronic end cap. In the barrel region, it is measured by the tile barrel surrounding the LAr electromagnetic barrel and the tile extended barrels which surround the end cap calorimeters. The LAr forward calorimeter, which is placed closest to the beam axis, serves both electromagnetic and hadronic calorimetry.

### 3. The ATLAS Detector at the LHC

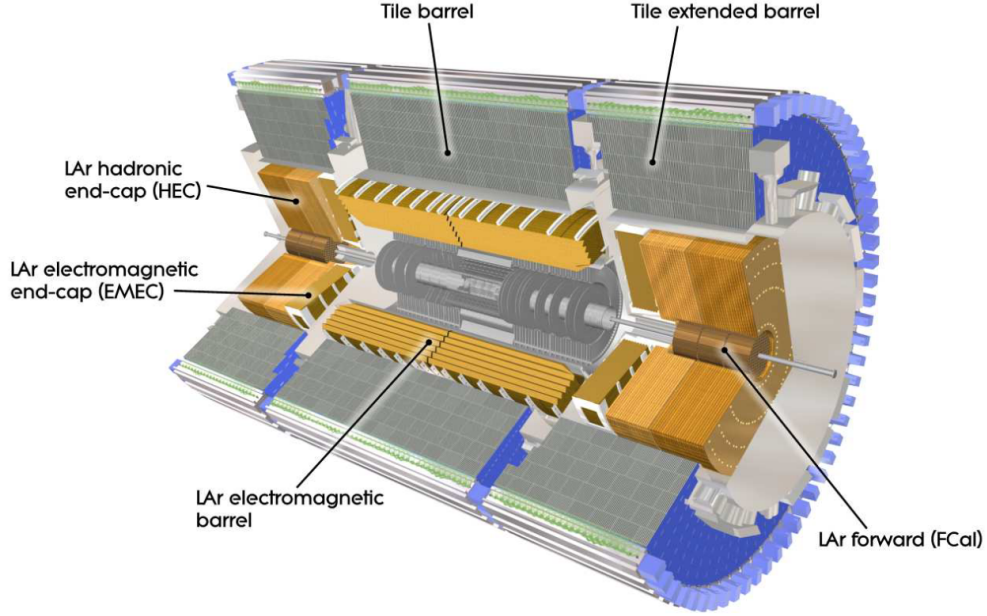


Figure 3.3.: Cut-away view of the ATLAS calorimeters (from [4]).

#### 3.2.3. Muon System

The purpose of the muon system, shown in Fig. 3.4, is to precisely measure muon momenta. Since muons are minimally ionizing, they are not stopped in the calorimeters. The muon tracks are bent by three toroidal magnets, the barrel toroid and two end-cap toroids. This magnet configuration is designed to achieve a field that is mostly perpendicular to the muon trajectories. The tracks are measured by monitored drift tubes in the low  $\eta$  region and cathode strip chambers in the high  $\eta$  region. Furthermore, resistive plate chambers and thin gap chambers are used for the trigger system.

### 3. The ATLAS Detector at the LHC

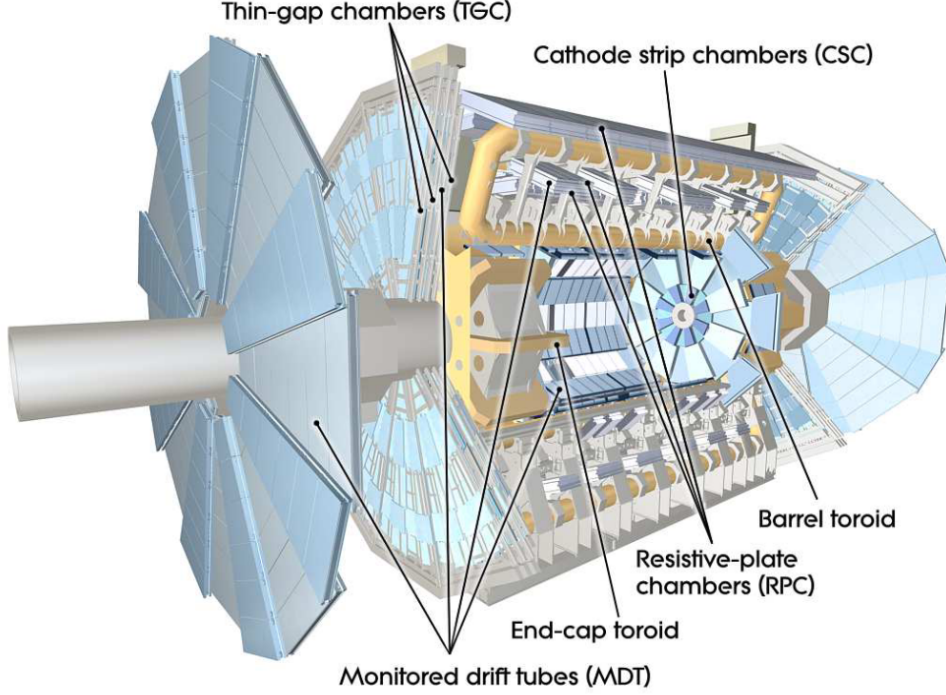


Figure 3.4.: Cut-away view of the ATLAS muon system [4].

### 3.3. ATLAS Coordinate System and Kinematics

The ATLAS experiment uses a coordinate system [4] with the interaction point as the origin, which is used throughout this thesis. The positive  $x$ -axis is defined as pointing from this origin towards the center of the LHC ring, while the  $y$ -axis points vertically upward. The  $z$ -axis then points along the beam pipe, so that the axes form a right-handed system. The azimuthal angle  $\phi$  is measured around the beam axis and the polar angle  $\theta$  is the angle to the beam axis.

Much about the kinematic variables used in the context of hadron colliders is motivated by the nature of proton-proton collisions: since the proton is a composite particle, only the momentum fraction carried by the partons involved in the process contributes to the interaction. Since in general the parton momenta are not of equal magnitude, the products of the interaction will be boosted in the  $z$ -direction, affecting the polar angles  $\theta$  of their momenta. For this reason, the pseudorapidity

$$\eta = -\ln(\tan(\theta/2)) \quad (3.1)$$

### 3. The ATLAS Detector at the LHC

is used instead. It is an approximation in the limit of  $E \approx |\vec{p}|$  of the rapidity

$$y = \frac{1}{2} \ln \left( \frac{E + |p_z|}{E - |p_z|} \right). \quad (3.2)$$

Here,  $\vec{p}$  is the momentum of a particle and  $p_z$  is its  $z$ -component. Since masses of (meta)stable particles are negligible compared to the energies involved in LHC collisions, this approximation is made throughout this thesis. Differences in rapidity are invariant under boosts in the  $z$ -direction, so the unknown boosts will not affect distributions of differences in  $\eta$ .

As a further consequence of the unknown momentum fractions, the total momentum in the reaction is unknown, and momentum conservation cannot be utilized in analyses. However, since the protons do not have a momentum component in the  $x$ - $y$  plane, it is negligible for the partons and the equation

$$\sum_i \vec{p}_T^i = 0 \quad (3.3)$$

must be fulfilled, where  $\vec{p}_T^i$  are the transverse momenta of the particles in the final state. The transverse momentum is defined as

$$\vec{p}_T = \begin{pmatrix} p_x \\ p_y \end{pmatrix}, \quad (3.4)$$

where  $p_x$  and  $p_y$  are the momentum components in the  $x$  and  $y$  directions, respectively. Conversely, they can be expressed using  $p_T = |\vec{p}_T|$  and  $\phi$ :

$$\begin{aligned} p_x &= p_T \cos \phi \\ p_y &= p_T \sin \phi. \end{aligned} \quad (3.5)$$

$p_z$  and the particle's energy  $E$  can be related to  $\eta$  using

$$\begin{aligned} E &= p_T \cosh \eta \\ p_z &= p_T \sinh \eta, \end{aligned} \quad (3.6)$$

assuming a massless particle. The transverse mass of a two particle system is defined as

$$m_T^{1,2} = \sqrt{m_1^2 + m_2^2 + 2(E_1^1 E_1^2 - \vec{p}_1^1 \vec{p}_1^2)}, \quad (3.7)$$

where the indices 1 and 2 signify the two particles. This is analogous to the invariant mass

$$m_{1,2} = \sqrt{(p_1^\mu + p_2^\mu)^2} = \sqrt{m_1^2 + m_2^2 + 2(E_1 E_2 - \vec{p}_1 \vec{p}_2)}, \quad (3.8)$$

where  $p_i^\mu$  are the momentum four-vectors of the particles. Based on the transverse

### 3. The ATLAS Detector at the LHC

momentum, one can define the transverse energy

$$E_T = \sqrt{p_T^2 + m^2}, \quad (3.9)$$

where  $m$  is the mass of the particle. For light particles this simplifies to

$$E_T \approx p_T. \quad (3.10)$$

The final state of a collision may involve particles which are not measured by the detector, for instance neutrinos. Based on Equation 3.3 one can define the missing transverse energy

$$\vec{E}_T^{\text{miss}} = - \sum_i \vec{p}_T^i \quad (3.11)$$

where  $\vec{p}_T^i$  are the transverse momenta of all measured particles.

## 3.4. The ATLAS Trigger

At the LHC, groups of protons, called bunches, collide at a rate of 40 MHz. This rate of data generation is too high for all data to be saved to memory for later analysis. Instead, it needs to be reduced to about 1 kHz. This is the responsibility of the trigger system, which selects events of physical interest to be recorded [24].

The trigger system consists of two levels, the Level-1 trigger, which consists of custom electronics and has a maximum output rate of 100 kHz. It is followed by the software based High-Level Trigger, which further reduces the rate to 1 kHz. Events are selected for storage if they are found to contain objects such as leptons, jets or missing transverse momentum, which pass criteria regarding their isolation, identification, momentum etc.

## 3.5. Object Reconstruction

The signals measured by the different detectors in ATLAS are used to reconstruct the particles produced in an interaction. The following sections briefly describe how different objects are reconstructed and identified.

### 3.5.1. Tracks and Vertices

Particle tracks in the detector are an important ingredient for all ATLAS analysis. They are reconstructed by fitting a model describing particle trajectories in the detector to particle hits in the different tracking detectors [25].

The primary vertex is the point in space at which the proton-proton collision took

### 3. The ATLAS Detector at the LHC

place. Primary vertices are reconstructed from the tracks in the event using an iterative fitting procedure [26].

#### 3.5.2. Electrons

Electrons in ATLAS are reconstructed from tracking and calorimeter information [27]. First, topological clusters in the electromagnetic calorimeter are reconstructed. These are groups of neighboring calorimeter cells, in which an amount of energy above a certain threshold has been deposited [28]. These are then matched with tracks reconstructed from particle hits in the inner detector.

The reconstructed electron candidates need to pass additional identification criteria, to distinguish them from energy deposits from jets and photons and from electrons that were produced in decays of hadrons instead of the primary interaction. The electron identification is based on properties of the electron track and the shower in the calorimeter, as well as on how well the track matches the cluster. These are combined into a discriminant, based on which electrons are accepted or rejected. It has different working points called Loose, Medium and Tight. A tighter working point leads to a higher purity, but also falsely rejects more real electrons. The efficiency is 93%, 88% and 80% for the Loose, Medium and Tight working points, respectively. The background rejection is improved by a factor of 2.0 and 3.5 relative to the Loose working point by using the Medium or Tight working point, respectively [27]. The  $WH \rightarrow WW$  analysis uses the Tight working point [29].

#### 3.5.3. Muons

In most cases, muons are reconstructed using the inner detector and the muon system [30]. First, particle tracks in both these subsystems are reconstructed independently, then a track from the inner detector and a track from the muon systems that are compatible with each other form a muon candidate. In the  $2.5 < |\eta| < 2.7$  range, which is outside of the acceptance of the inner detector, muons can be reconstructed from just the muon system, if the track can be extrapolated to originate from the interaction point.

The identification criteria for muons are based on the quality of the muon track. If the muon originates from an in-flight decay of a hadron instead of the primary interaction point, this leads to a “kink” in the track at the location of the decay, as the muon flies in a different direction than the original hadron. As a result of this, the inner detector and muon system will measure different momenta. An upper limit on the mismatch between the momentum measurements is used to reject non-prompt muons. Furthermore, there are requirements on the minimum number of hits in the muon system.

The muon identification, too, has a Loose, Medium and Tight working point with different efficiencies and background rejections. In the  $20 \text{ GeV} < p_T < 100 \text{ GeV}$

### 3. The ATLAS Detector at the LHC

range, the prompt muon efficiencies are 98.1%, 96.1% and 91.8%, while only 0.76%, 0.17% and 0.11% of non-prompt leptons are accepted, respectively [30]. The  $WH$   $H \rightarrow WW$  analysis uses the Medium working point [29].

#### 3.5.4. Lepton Isolation

In addition to the identification criteria, leptons need to pass isolation criteria to further reject non-prompt leptons. In the case of the associated production  $H \rightarrow WW$  analysis, this selection is based on the newly developed Prompt Lepton Improved Veto (PLIV) [29]. It consists of a boosted decision tree with input variables related to various properties of the lepton tracks. The sets of inputs differ between electrons and muons. The PLIV has different working points called Tight and VeryTight. The  $WH$   $H \rightarrow WWW$  analysis uses the Tight working point, which has a signal efficiency of 0.72 and a total background efficiency of 0.43, compared to 0.54 and 0.23 for the VeryTight working point [29].

#### 3.5.5. Jets

Jets are reconstructed from energy deposits in the calorimeter as well as tracks. First, topological clusters are formed from the energy deposits, then tracks are matched to the clusters. To avoid double counting the energies of charged particles in the jet, the momentum of a track is subtracted from the energy deposited in the matching clusters. This procedure is called the particle flow algorithm [31]. From the tracks matched to clusters and the remaining clusters, which originate from neutral particles, jets are then reconstructed using the anti- $k_t$  algorithm with a radius parameter of  $R = 0.4$  [32].

#### 3.5.6. Missing Transverse Energy

The reconstruction of  $E_T^{\text{miss}}$  requires the combination of all detector subsystems. It is calculated as

$$\vec{E}_T^{\text{miss}} = - \sum_{i \in \text{hard objects}} \vec{p}_T^i - \sum_{j \in \text{soft signals}} \vec{p}_T^j \quad (3.12)$$

and consists of two contributions: hard objects, i.e. leptons, jets and photons, and soft signals, i.e. signals in the detector that were not reconstructed to belong to any of the hard objects. Since all types of objects are reconstructed independently of each other, a given signal in the detector might be part of more than one object. The  $E_T^{\text{miss}}$  reconstruction uses only objects reconstructed from mutually exclusive detector signals, to avoid double counting energy [33].

## 4. Statistical Methods

This chapter explains the most important statistical methods that were used in the analysis presented here.

### 4.1. Monte Carlo Simulations

An important component of statistical analyses are probability density functions, according to which the random variables are distributed. In a complex particle physics experiment, such as ATLAS, it is generally not possible to calculate these explicitly: they depend on the matrix element of the process, the detector response, the reconstruction algorithms, the selections of the analysis etc.

This problem can be solved using Monte Carlo (MC) simulations. First, a large number of events are randomly generated according to the known theoretical distribution of the physical process. The distribution of these raw Monte Carlo events is normalized to the physical prediction by giving each event a weight, so that the sum of the weights matches the prediction.

Instead of having to explicitly calculate the shape of the distribution after it is modified for instance by the detector response or a selection in the analysis, one only needs to calculate the effect on each individual event by inputting it into a simulation of the detector or evaluating if it passes the selection criteria. One can then investigate how the MC distribution was affected by this.

The precision of the MC prediction in a given region of phase space, e.g. a histogram bin, depends on how many MC events there are in the region. The statistical uncertainty of the MC prediction can be estimated as

$$\sigma_{MC} = \sqrt{\sum_{i=1}^N w_i^2}, \quad (4.1)$$

where  $w_i$  are the weights of the MC events. This estimation is not valid in regions where no MC events are found, the resulting value of  $\sigma_{MC} = 0$  clearly does not account for the possibility that no event was generated in the region simply by chance.

## 4. Statistical Methods

### 4.2. Maximum Likelihood Fits

Maximum likelihood fits [34] are a method for estimating the value of one or more parameters  $\theta$  of a probability distribution from data. The likelihood  $L(\theta)$ , which is a function of the parameters, is the probability of observing the data, given a certain value of the parameters. In the simple case of  $N$  independent measurements  $x_i$ , which are distributed according to a probability density  $f(x, \theta)$ , the likelihood is given by

$$L(\theta) = \prod_{i=1}^N f(x_i, \theta). \quad (4.2)$$

The maximum likelihood estimate  $\hat{\theta}$  is the value of  $\theta$  at which  $L(\theta)$  has a global maximum. In practice, instead of maximizing the likelihood the negative log-likelihood  $-\ln L$  is minimized, since  $L$  can have very small values, which lead to problems in numerical calculations.

In high energy physics, the measurements are usually histogrammed data containing numbers of events in some bins. These are compared to a predicted number of signal and background events. The parameter of interest is the signal strength  $\mu = n_{\text{observed}}/n_{\text{predicted}}$ , the ratio between the observed and predicted events or equivalently between the observed and predicted cross section. The number of events in a bin follows a Poisson distribution. The likelihood for  $\mu$  is then [35]

$$L(\mu) = \prod_j \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)}, \quad (4.3)$$

where  $n_j$  are the observed events in bin  $j$ , and  $s_j$  and  $b_j$  are the signal prediction and background prediction in bin  $j$ , obtained from Monte Carlo simulations. The likelihood may include additional terms and parameters to account for systematic uncertainties. For example one might add a parameter which scales the predicted number of background events and a Poissonian term which constrains the value of this parameter, using a measurement in a different region of phase space. Parameters that need to be included in the likelihood but are not directly of interest are called nuisance parameters.

Assuming the likelihood function approximately follows a Gaussian distribution w.r.t.  $\theta$ , the covariance matrix  $V(\hat{\theta})$  of the estimated parameters can be estimated using

$$\hat{V}(\hat{\theta}) = \left( - \frac{\partial^2 \ln L(\theta)}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} \right)^{-1}. \quad (4.4)$$

If data from a real experiment is not available, one might still wish to perform fits based on the MC. For example, it is a common practice not to look at data in the regions where the signal is expected before the analysis is finalized. This is called

#### 4. Statistical Methods

blinding. It avoids subconscious biases of the analyzers skewing the analysis towards a certain result. In such cases, one can perform a so-called Asimov fit [35]: the data is simply chosen in such a way that it matches the expectation exactly. In the above example, this means setting  $n_j = s_j + b_j$ . This results in a fit result of  $\mu = 1$  with an uncertainty that is indicative of the sensitivity of the analysis.

### 4.3. Hypothesis Tests and Upper Limits

Hypothesis testing is the process of determining if a null hypothesis  $H_0$  is compatible with the observed data [34]. For this purpose, a test statistic  $t(x)$  is defined, which is a function of the data. Furthermore, a significance level  $\alpha$  of the hypothesis test is chosen. The hypothesis is rejected if the observed value of the test statistic  $t_{\text{obs}}$  is less compatible with  $H_0$  than a cut value  $t_c$ . Assuming greater values of  $t$  mean lower compatibility,  $t_c$  is related to  $\alpha$  by

$$\int_{t_c}^{\infty} g(t|H_0)dt = \alpha, \quad (4.5)$$

where  $g(t|H_0)$  is the probability density of  $t$  under the condition that  $H_0$  is true. Based on this definition, the hypothesis is falsely rejected if  $H_0$  is true with probability  $\alpha$ . One can also define an alternate hypothesis  $H_1$ . It is falsely rejected with probability

$$\beta = \int_{-\infty}^{t_c} g(t|H_1)dt. \quad (4.6)$$

Equivalently to comparing  $t_{\text{obs}}$  and  $t_c$ , one can calculate the  $p$ -value

$$p_i = \int_{t_{\text{obs}}}^{\infty} g(t|H_i)dt \quad (4.7)$$

and reject the null hypothesis if  $p_0 < \alpha$ . The  $p$ -value is the probability of measuring a value of  $t$  at least as inconsistent with the hypothesis as  $t_{\text{obs}}$ .  $p$ -values are often expressed as significances

$$Z = \Phi^{-1}(1 - p), \quad (4.8)$$

where  $\Phi$  is the cumulative distribution of the unit Gaussian.

In particle physics, the null hypothesis is often the background only hypothesis corresponding to a signal strength  $\mu = 0$ , with other values of  $\mu$  being alternative hypotheses, most notably  $\mu = 1$ , the predicted signal according to some model. In this case, the test statistic is often based on the profile likelihood ratio

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} \quad (4.9)$$

#### 4. Statistical Methods

where  $\hat{\mu}$  and  $\hat{\theta}$  are the best fit values of the parameter of interest and the nuisance parameters, respectively.  $\hat{\theta}$  are the best fit values of the nuisance parameters for a given fixed value of  $\mu$ .

If the background-only hypothesis  $\mu = 0$  cannot be excluded, one can instead calculate an upper limit for  $\mu$ . This is done by calculating  $p$ -values for a range of  $\mu$  values. Values of  $\mu$  with  $p_1 < \alpha$  can be included, values with  $p_1 > \alpha$  cannot. Thus, the value of  $\mu$  with  $p_1 = \alpha$  is an upper limit for  $\mu$  at the given confidence level. In this procedure  $p_1$  is often replaced by

$$CL_S = \frac{p_1}{1 - p_0}. \quad (4.10)$$

This prevents including values of  $\mu$ , for which there is little sensitivity for distinguishing between  $H_0$  and  $H_1$  [34].

A test statistic specifically designed for calculating upper limits is

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}. \quad (4.11)$$

It is defined in this way because data with  $\hat{\mu} > \mu$  would not be less compatible with  $\mu$  than the observed data [35]. If the parameter of interest is bounded to  $\mu \geq 0$ , this test statistic may be modified to

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad (4.12)$$

with

$$\tilde{\lambda}(\mu) = \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} \quad (4.13)$$

if  $\hat{\mu} < 0$  and  $\tilde{\lambda}(\mu) = \lambda(\mu)$  otherwise. A corresponding test statistic for the background hypothesis is

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}, \quad (4.14)$$

which is defined in such a way that downward fluctuations of the background do not lead to rejection of the hypothesis.

As seen in Equation 4.7, the probability distribution function of the test statistic is required to calculate the  $p$ -value. For the test statistics shown above, it depends on the specific likelihood function and is generally not known. Approximations can be used in the limit of large samples, in which the likelihood function can be assumed to be a Gaussian distribution [35]. In the case of small data samples so-called toy based methods need to be used instead. This means generating many sets of pseudo-data (“toys”) based on the distributions used in the construction of the likelihood function and calculating the test statistic on these. This results in a distribution of values of the test statistic, which can then be used to calculate the  $p$ -value.

## 4.4. Neural Networks

Artificial Neural Networks (ANN) are a popular machine learning method [36]. Many different types of ANNs exist. Only the simplest type, the multi-layer perceptron, will be described here, as it is the only kind of ANN used in this thesis.

### 4.4.1. Nodes, Layers and Networks

Nodes, also called neurons, are the basic constituents of ANNs. Each node has a vector of inputs  $\vec{x}$  and a scalar output  $y(\vec{x})$ , which is calculated according to

$$y(\vec{x}) = g(\vec{w}\vec{x} + b). \quad (4.15)$$

The node has a number of parameters, the weights  $\vec{w}$  and the bias  $b$ .  $g$  is the so-called activation function. A node with three inputs is illustrated in Fig. 4.1.

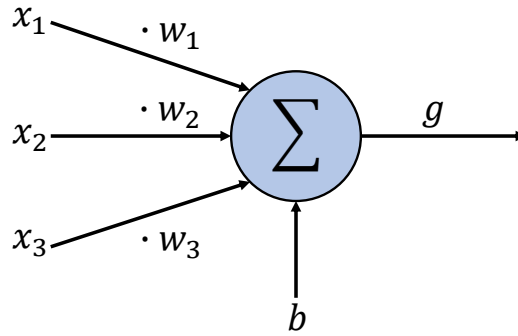


Figure 4.1.: A node with three inputs. The inputs are each multiplied with a weight, then the results and the bias are added together. Finally, the activation function is applied.

Several nodes are combined into a layer. Each node in a layer has the same inputs and the same activation function. The outputs of the nodes differ, since they have different weights and biases. Layers can be chained together to form a neural network, by passing the outputs of every node in the first layer as inputs to every node in the second layer, as illustrated in Fig. 4.2. Different layers may have different numbers of neurons. Layers that do not connect directly to the inputs and outputs of the network but instead to other layers are known as hidden layers. A neural network is effectively a function of its inputs with many parameters. By systematically tuning these parameters, a sufficiently large ANN can be made to approximate any function. This is known as the universal approximation theorem [37].

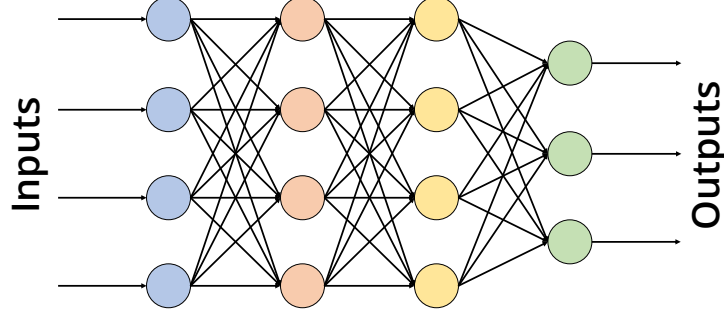


Figure 4.2.: An ANN with four layers.

#### 4.4.2. Activation Functions

A large variety of functions are used as activation functions for neural networks. Popular examples are the sigmoid function, the hyperbolic tangent and the Rectified Linear Unit (ReLU). Activation functions need to be non-linear to enable the network to approximate non-linear functions. The networks presented in this thesis use the swish activation function

$$\text{swish}(x) = \frac{x}{1 + e^{-\beta x}} \quad (4.16)$$

with  $\beta = 1$ , which was found to be the optimal activation function in the original training of the multiclassifier for signal-background separation in the  $Z$  depleted signal region and has been found to outperform other activation function in a number of different machine learning tasks [38].

#### 4.4.3. Training

The process of optimizing the parameters of an ANN is called training and proceeds as follows. First, the net is applied to a set of training data, i.e. input values for which the desired output values are known. A loss function is defined, which quantifies the difference between the ANN outputs and the desired values. The loss functions used in this thesis are the Mean Squared Error (MSE) and the categorical cross entropy. The mean squared error is defined as

$$MSE = \frac{1}{N} \sum_i \left( y_i^{\text{correct}} - y_i^{\text{predicted}} \right)^2, \quad (4.17)$$

where  $y_i^{\text{correct}}$  and  $y_i^{\text{predicted}}$  are the desired outputs for the  $i$ -th of  $N$  training events and the values produced by the ANN, respectively. The categorical cross entropy loss

#### 4. Statistical Methods

is defined for a single event as

$$C = - \sum_j y_j^{\text{correct}} \log y_i^{\text{predicted}}, \quad (4.18)$$

where  $y_j$  refers to the value of the  $j$ -th output node.

For a given set of training data, the loss function is a scalar function of the network parameters. Using numerical algorithms, a (local) minimum of the loss function can be found, corresponding to a set of parameters for which the neural network closely approximates the function underlying the training data.

The gradients necessary for the minimization are calculated using the back-propagation algorithm [36]. It works by considering the network as the composition of many small operations, e.g. the linear functions of the neurons. For each individual operation it is known how to back-propagate the results, i.e. how to calculate the value of the gradient in parameter space given some in- and outputs. By recursively applying the chain rule of calculus, the individual gradients can then be combined into the gradient of the entire network.

Once the gradient has been obtained, the parameters can be adjusted by a small step opposite to the gradient, thus reaching a configuration with a slightly smaller loss than before. By repeatedly calculating the loss and its gradient and adjusting the weights, the loss is minimized. Depending on the minimization algorithm used, this method is extended by for example adding random fluctuations to the steps or reducing the size of the steps over time. More sophisticated algorithms, such as stochastic gradient descent and the Adam algorithm used in this thesis [39], can lead to quicker convergence and avoid the minimization ending in a local minimum with poor network performance.

##### 4.4.4. Hyperparameter Optimization

When training a neural network, choices have to be made regarding the properties of the network and the training procedure, such as the network architecture, the activation functions used, the loss function, the optimization algorithm and many more. These are collectively referred to as hyperparameters. A set of hyperparameters needs to be found that leads to an ANN with good performance. There is no general prescription for how to determine the hyperparameters. However, one can systematically train a number of networks with different hyperparameters and compare them to find a good combination.

##### 4.4.5. Overfitting and Regularization

A common problem during the training of neural networks is overfitting. Given a large enough number of trainable parameters and sufficiently long training, the network can start to very precisely produce the correct results for the training data, but at the

#### 4. Statistical Methods

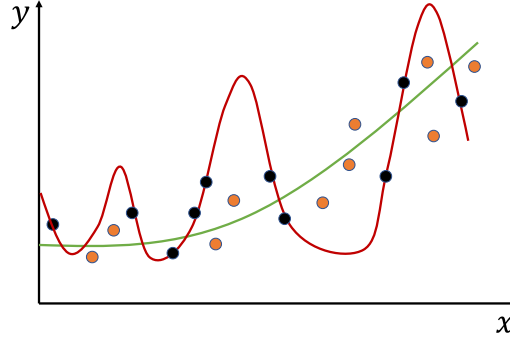


Figure 4.3.: Example of overfitting. The training data (black dots) is perfectly reproduced by the output of the overfitted network (red line). However, it fails to generalize to new inputs (orange dots). The green curve represents a better ANN, which gives good results for all data points.

cost of a loss of generality, i.e. the ability to produce good results for inputs not seen in the training. This effect is illustrated in Fig. 4.3.

Techniques designed to avoid overfitting are collectively known as regularization. Two common regularization techniques, which were used in this thesis, are explained in the following.

##### 4.4.5.1. Early Stopping

Overfitting can be recognized by defining a separate validation dataset. The data points in the validation sample are not used in the training, instead only the loss and any other metrics are computed on it and compared to the metrics on the training sample. These metrics therefore indicate how well the ANN performs on new inputs. For example, at the beginning of the training, both training loss and validation loss decrease. The training loss continues to decrease until eventually leveling off. Once the network starts overfitting, the validation loss will start to increase again. By recognizing this and stopping the training, overfitting can be avoided. This technique, which is illustrated in Fig. 4.4, is called early stopping.

##### 4.4.5.2. Dropout

Another method to prevent overfitting is dropout. During training of the ANN, each neuron has a random chance of its output being set to 0 during an iteration. This way, the network is forced to generalize better, since a given neuron cannot rely on all its inputs being available in a given training step. Outside of the training, such as when applying the ANN to the training set or in the final analysis, dropout is not

#### 4. Statistical Methods

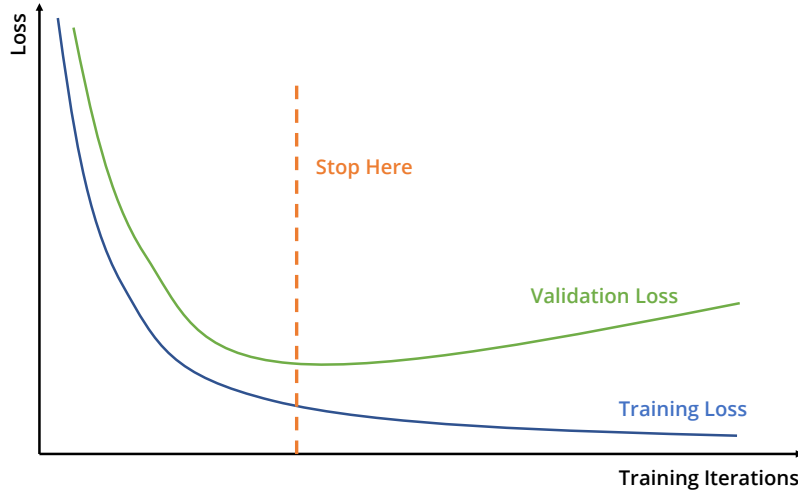


Figure 4.4.: Illustration of the principle of early stopping. Overfitting can be avoided by ending the training when the validation loss starts increasing.

used. The probabilities for a neuron to drop out are hyperparameters that can be optimized. It can differ between layers.

##### 4.4.6. Validation and Testing Samples

As described in the previous sections, a separate validation sample is useful to recognize and prevent overfitting during the network training. The easiest way of defining such a validation sample is to simply split the available training data into two parts, one part is then used as a training sample and the other as a validation sample. However, since part of the data then cannot be used for training, this can be undesirable if only limited data is available. Instead, the method of  $k$ -fold cross validation [34] can be used: the data is split into  $k$  independent subsets. Then  $k$  different ANNs are trained. In each training one of the subsets is used as a validation sample and the others as a training sample. Therefore for each event there is a network that was not trained on it. The performance of the  $k$  ANNs (“ $k$ -folds”) can then be used for the optimization of the hyperparameters.

Though the validation sample is not used directly to train the neural network, it still influences the training, since the performance of the network on the validation sample is used to optimize the hyperparameters. For this reason a third sample, the testing sample, is defined, which is only used for the final evaluation of the trained ANN. This ensures that the result is truly indicative of the neural network’s ability to generalize to new data.

## 4.5. Regression Networks

Regression networks are used to directly reconstruct some function from data, i.e. each set of inputs (in particle physics corresponding to an event) is directly associated with a value that needs to be reproduced. Most commonly, such a network has a single output node, with an activation function that lets it output a wide range of values, such as simply the identity function. A popular loss function for regression networks is the mean squared error.

## 4.6. Multiclassifier Networks

ANNs are often used in classification tasks. Each event is labeled with one of several classes, and the network is meant to predict to which class the event belongs. In the case of multiclassifiers, i.e. when there are more than two classes, the outputs of the neural network typically use one-hot encoding: each class has a separate output node, which should be 1 when the event belongs to this class and 0 otherwise. In general the output values will be between 0 and 1. Often, including in this thesis, the softmax function

$$\text{softmax}(\vec{y})_i = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (4.19)$$

is applied to the outputs  $\vec{y}$ , resulting in  $\sum_i \text{softmax}(y_i) = 1$ . The outputs can then be interpreted as a probability of the event to belong to the respective class.

## 5. Signal and Background Processes

### 5.1. The Signal Process

The analysis presented in this thesis concerns a specific combination of production process and decay mode: The Higgs boson is produced in association with a  $W$  boson and decays via  $H \rightarrow WW$ . All three  $W$  bosons then decay into a lepton-neutrino pair. In the remainder of this thesis, “lepton” is used as an umbrella term for only electrons and muons, since only they are measured directly in the detector: taus decay before reaching the detector and neutrinos pass through it undetected. Events in which a  $W$  boson decays to a tau are implicitly included in the signal, if the tau decays leptonically via  $\tau \rightarrow \nu_\tau l \nu_l$ , where  $l$  is a muon or an electron.

A Feynman diagram for this process is shown in Fig. 5.1. Due to charge conservation, the lepton charges need to add up to  $\pm 1$ . Among the three leptons, there are thus always two opposite charge pairs and one same charge pair. The leptons can be any combination of electrons and muons. There may be zero, one or two Same Flavor Opposite Sign (SFOS) lepton pairs. Based on this, two Signal Regions (SR) are defined: The  $Z$  depleted signal region (see Section 7.2 for more details) with no SFOS pairs, on which this thesis is focused, and the  $Z$  dominated region with at least one SFOS pair. This split is motivated by the  $WZ$  background, as discussed in Section 5.4.1.

### 5.2. Spin Correlation

The spin 0 nature of the Higgs boson has implications for the kinematics of its decay products, as illustrated in Fig. 5.2. The projections of the spins of the  $W$  bosons onto the Higgs boson decay axis must sum to the spin 0 of the Higgs. If one of the  $W$  boson has  $m_s = 1$ , the other must thus have  $m_s = -1$ . In this case, the spins of the lepton and neutrino from a  $W$  need to be parallel to add to  $\pm 1$ . As a result, the spins of one lepton-neutrino pair are opposite to those of the other pair.

Since the  $W$  only couples to left-chiral particles and right-chiral antiparticles and neutrinos are (nearly) massless, the neutrino from the  $W^+$  needs to have negative helicity and the antineutrino from the  $W^-$  positive helicity. Since in the case of non-zero  $W$  boson spin projections the spins of the two neutrinos are opposite, the opposite helicities imply that the neutrino momenta, and by extension the lepton momenta,

## 5. Signal and Background Processes

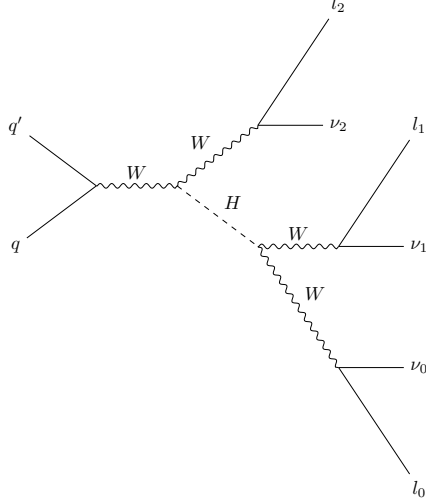


Figure 5.1.: The signal process  $qq \rightarrow WH \rightarrow WWW \rightarrow l\nu l\nu l\nu$ .

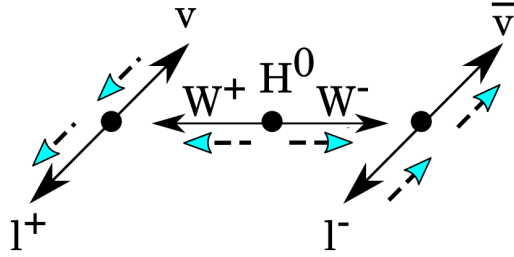


Figure 5.2.: Illustration of the spin correlation. The black arrows indicate particle momenta and the cyan arrows indicate spins. The spins and momenta of the decay products are drawn in the rest frames of the respective parent particles [40].

generally point into the same direction. In other words, the angular separation

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (5.1)$$

between the leptons originating from the Higgs boson is small.

This can be used to determine which leptons originate from the Higgs boson. Because of charge conservation, the lepton with unique charge must originate from the Higgs. It is labeled  $l_0$ . The lepton with the smallest  $\Delta R$  with  $l_0$  is then labeled  $l_1$  and assumed to also originate from the Higgs boson. The remaining lepton  $l_2$  is assumed to originate from the associated  $W$  boson. These assumptions are found to be correct in 99% of the cases for  $l_0$  and in 85% for  $l_1$  [29].

## 5. Signal and Background Processes

Since the spin correlation is specific to the  $H \rightarrow WW$  decay and not found in background processes, kinematic variables involving the angles between leptons, for example the invariant mass between leptons 0 and 1, can be used to discriminate between signal and background events.

### 5.3. Simplified Template Cross Sections

Effects of new physics beyond the Standard Model might affect not only the overall signal strength of the signal process, but also the kinematic distributions of the production and decay processes. When searching for new physics it is therefore useful to measure not only inclusive cross sections, but to also measure differential cross sections, for example in dependence of the transverse momentum  $p_T^W$  of the associated  $W$  boson. However, a finely resolved differential measurement requires a larger number of  $pp \rightarrow WH \rightarrow WWW \rightarrow \nu\nu\nu\nu\nu$  events than are expected in the  $139 \text{ fb}^{-1}$  of collision data from the LHC Run 2. As an intermediate step, one can instead measure cross sections in only a few regions of phase space. The Simplified Template Cross Section (STXS) scheme defines such bins in phase space [10]. The STXS bin definitions are updated based on the amount of available collision data, the version used here is Stage 1.2 [41].

The STXS scheme has additional benefits beyond being a coarse differential measurement. The bins are defined in such a way that they minimize the influence of theoretical uncertainties on the measurements, while maximizing the experimental sensitivity. Furthermore, the common definition of the bin allows for an easier combination of measurements across decay channels and even experiments, the STXS scheme having been adapted by both ATLAS and CMS.

The bin definitions differ between the different Higgs production processes and are inclusive in the Higgs decay channels. The scheme for VH, where the associated vector boson decays leptonically<sup>1</sup>, is shown in Fig. 5.3. The gaps between boxes represent the recommended splitting for the measurement, while the dashed lines indicate a finer splitting used for the handling of theoretical uncertainties. For the purposes of this analysis, the scheme is slightly simplified to be inclusive in the number of jets. This results in four bins in the transverse momentum  $p_T^V$  of the associated boson, as listed in Table 5.1.

### 5.4. Background Processes

The  $qq \rightarrow WH \rightarrow WWW \rightarrow \nu\nu\nu\nu\nu$  signal process is on the most basic level recognized by the presence of three leptons in the final state, the charges of which add up to  $\pm 1$ . The important background processes are therefore those that also feature such leptons.

---

<sup>1</sup>The VH process with a hadronically associated boson shares a binning scheme with VBF, as the processes have similar signatures.

## 5. Signal and Background Processes

Table 5.1.: Definition of the STXS bins used in this analysis.

| Bin | $p_T^W$ range      |
|-----|--------------------|
| 0   | 0 GeV to 75 GeV    |
| 1   | 75 GeV to 150 GeV  |
| 2   | 150 GeV to 250 GeV |
| 3   | Above 75 GeV       |

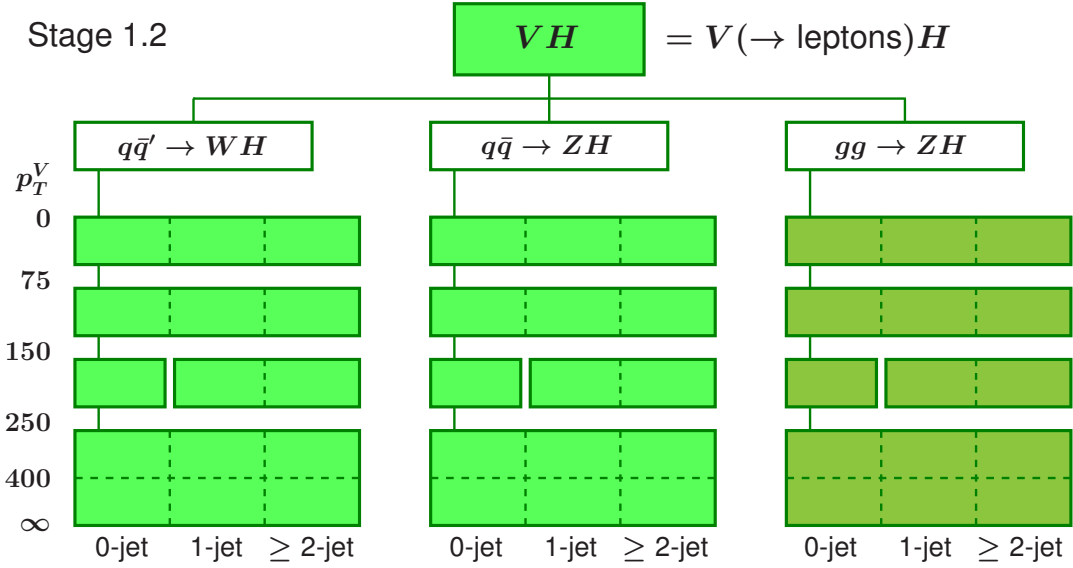


Figure 5.3.: STXS scheme for Higgs production with an associated vector boson, which decays hadronically.  $p_T^V$  is given in units of GeV.

### 5.4.1. $WZ/\gamma^*$ and Electron Charge Misidentification

The production of a  $W$  and a  $Z$  boson or virtual photon is shown in Fig. 5.4. If both bosons decay leptonically, there are three leptons in the final state, as is the case in the signal process. This background process motivates the splitting into a  $Z$  dominated and a  $Z$  depleted signal region: A lepton pair originating from a  $Z$  decay has opposite charge and the same flavor. Rejecting such SFOS pairs therefore suppresses events with a  $Z$ . This is not perfect, however: there are two mechanisms by which the  $WZ$  process can still contribute to the  $Z$  depleted signal region,  $Z \rightarrow \tau\tau$  and charge misidentification.

If the  $Z$  boson decays into a  $\tau^+\tau^-$  pair, these taus do not reach the detector, but decay after a short distance with  $c\tau = 87.03 \mu\text{m}$  [9]. Taus can decay leptonically via  $\tau^- \rightarrow l^- \bar{\nu}_l \nu_\tau$ . If one tau decays to a muon and the other to an electron, there is no SFOS pair from the  $Z$  and the event can pass the  $Z$  depleted signal region selection.

## 5. Signal and Background Processes

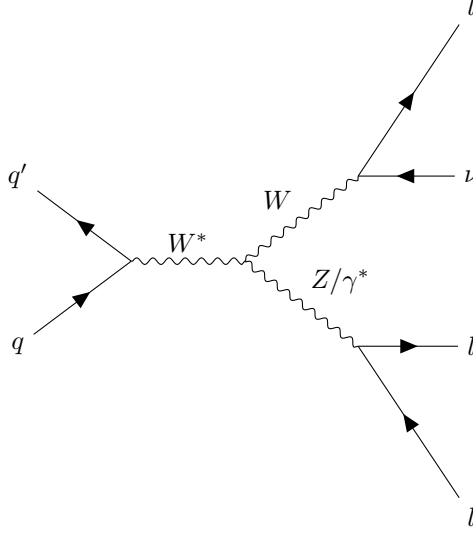


Figure 5.4.: Leading order Feynman diagram of the  $WZ$  background process.

The other way for a  $WZ$  event to have no SFOS pair is if the charge of an electron from the  $W$  and  $Z$  decays is reconstructed with the wrong sign. In ATLAS, the electron charge is determined by measuring the curvature of the electron track in the magnetic field [27]. When a high energy electron moves through matter, such as detector components, it might radiate a hard photon via bremsstrahlung. This photon, in turn, can undergo pair production in the material, leading to additional secondary electron tracks close to the primary electron. The presence of these unphysical particle tracks, which cause additional hits in the tracking detectors, can lead to particle tracks being reconstructed from hits of both primary and secondary electrons [27]. Additionally, the energy loss of the primary electrons distorts the shape of their tracks. Both of these effects can cause the curvature of the track, and thus the electron charge, to be misidentified. Since muons are minimally ionizing, they do not undergo these interactions in the detector material and are only negligibly affected by charge misidentification.

### 5.4.2. Top Pair Production and Fake Leptons

The production of top-antitop pairs is a common process at the LHC, with a cross section of 830 pb [42]. Top quarks almost exclusively decay to a  $b$  quark and a  $W$  boson, with  $\Gamma(Wb)/\Gamma(Wq(q = b, s, d)) = 0.957$  [9]. The  $b$  quark produces a jet, while the  $W$  decays further, either leptonically or hadronically. There can therefore be up to two prompt leptons in the final state.  $t\bar{t}$  production can be mistaken for the signal process if a third lepton is detected.

This can happen either if a non-prompt lepton is produced in the decay of some hadron inside a jet or if some other object, such as a photon or jet, is mistakenly identified as

## 5. Signal and Background Processes

a lepton. These effects are collectively referred to as fake leptons. Other background channels contribute to the fake lepton background, such as the production of a  $Z$  boson and jets, which also has a high cross section, but due to the 0 SFOS requirement in the  $Z$  depleted signal region  $t\bar{t}$  production is the most important process.

Since fake lepton backgrounds are not well modeled by Monte Carlo simulations, their contribution to the signal region is estimated using data driven methods. For muons, the fake factor method is used [29]. For this method a Control Region (CR) is constructed, which has the same definition as the signal region, except that one of the leptons passes only looser isolation criteria, but not those imposed on SR leptons. Such a lepton is called an anti-ID lepton, as opposed to ID leptons which pass the criteria. After the backgrounds with prompt leptons, which are determined from Monte Carlo simulations, are subtracted from the CR, the number of fake lepton background events in the SR can then be calculated by

$$N_{\text{Background,SR}} = FF \cdot N_{\text{Background,CR}}. \quad (5.2)$$

This requires the fake factor  $FF$ , which is calculated from two additional regions SR' and CR'. These are orthogonal to SR and CR and enhanced in the fake lepton background. SR' contains three ID leptons, while CR' contains two ID leptons and one anti-ID lepton. The fake factor is then calculated by subtracting the prompt backgrounds in SR' and CR' and taking the ratio of the two regions

$$FF = \frac{N_{\text{SR'}}^{\text{fake}}}{N_{\text{CR'}}^{\text{fake}}} = \frac{N_{\text{SR'}}^{\text{data}} - N_{\text{SR'}}^{\text{prompt MC}}}{N_{\text{CR'}}^{\text{data}} - N_{\text{CR'}}^{\text{prompt MC}}}. \quad (5.3)$$

Under the assumption that the ratio is the same between SR and CR as between SR' and CR', this can then be used to estimate the fake lepton background in the signal region, as described above.

For electrons, it is not possible to construct control regions sufficiently pure in the different sources of fake electrons. These are heavy flavor jets, light flavor jets and photon conversion. Thus, separate fake factors cannot be measured. Instead, a scale factor method is used, in which the factors for the three processes are simultaneously determined in a fit using three control regions and then applied to the signal region [29].

### 5.4.3. Triboson

Another background is the production of three vector bosons, where the bosons decay in such a way that there are three leptons in the final state. The most important triboson process is the production of three  $W$  bosons, as this closely resembles the signal process, with all  $W$  bosons decaying hadronically. Other triboson processes with three leptons, such as  $WZZ$  with one  $Z$  decaying hadronically, involve an SFOS pair from a  $Z$  decay and are thus suppressed in the  $Z$  depleted signal region. Feynman diagrams for  $WWW$  production are shown in Fig. 5.5.

## 5. Signal and Background Processes

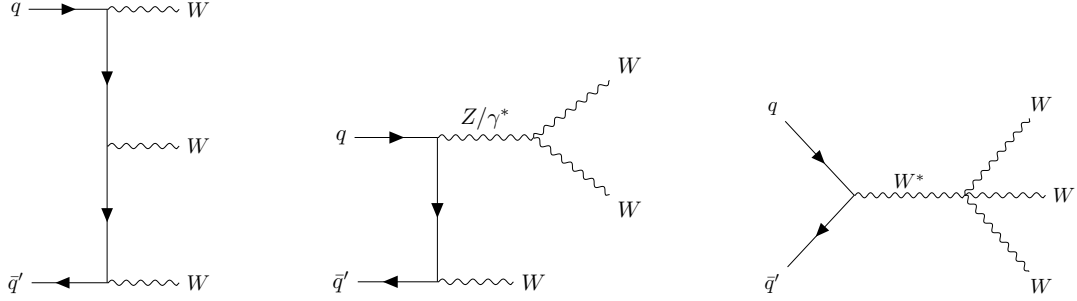


Figure 5.5.: Leading order Feynman diagrams for the  $WWW$  background process.

### 5.5. Previous Measurement of the Signal Process

The associated production of a Higgs boson with a vector boson and the subsequent decay  $H \rightarrow WW \rightarrow l\nu l\nu$  has previously been measured at the ATLAS experiment using  $36.1 \text{ fb}^{-1}$  of collision data with a significance of  $2.6 \sigma$  for the  $WH$  process and  $2.8 \sigma$  for  $ZH$  [43]. The signal strengths were  $\mu_{WH} = 2.3^{+1.2}_{-1.0}$  and  $\mu_{ZH} = 2.9^{+1.9}_{-1.3}$ , both about one standard deviation above the SM expectation. The combined result for both channels, shown in Fig. 5.6, has an observed significance of  $4.1 \sigma$  and a signal strength of  $\mu_{VH} = 2.5^{+0.9}_{-0.8}$ . This excess of about 1.5 standard deviations further motivates the STXS measurement in this channel, as it can provide additional sensitivity to possible BSM effects, which are expected to be enhanced in the higher STXS bins.

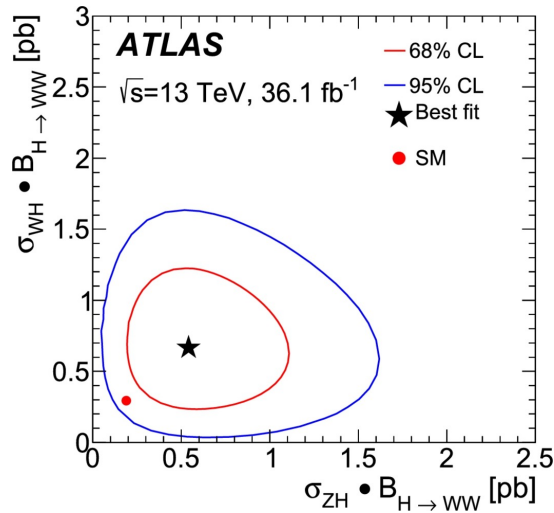


Figure 5.6.: Combined result for the  $WH$  and  $ZH$  production channels in the previous measurement [43].

## 6. Reconstruction of the $W$ Boson Momentum

### 6.1. Degrees of Freedom and Analytical Reconstruction

It is not possible to exactly reconstruct the transverse momentum of the associated  $W$  boson  $p_T^W$  in the signal process. The reason for this are the three neutrinos in the final state. The transverse momentum of the neutrino from the  $W$  would have to be known to calculate  $\vec{p}_T^W = \vec{p}_T^{l_2} + \vec{p}_T^{\nu_2}$ , but because of the other two neutrinos contributing to the missing transverse energy, it is unknown.

One may instead attempt to reconstruct  $p_T^W$  approximately by making some simplifying assumptions about the kinematics of the process. In total, there are 9 unknowns: The three neutrino momenta with three degrees of freedom each. Applying transverse momentum conservation

$$\vec{E}_T^{\text{miss}} = \vec{p}_T^{\nu_0} + \vec{p}_T^{\nu_1} + \vec{p}_T^{\nu_2} \quad (6.1)$$

removes two degrees of freedom. Three more degrees of freedom can be eliminated by imposing mass constraints

$$m_W^2 = (E_{l_i} + E_{\nu_i})^2 - (\vec{p}_{l_i} + \vec{p}_{\nu_i})^2, \quad (6.2)$$

These equations are only approximately true, since at least one  $W$  boson from the Higgs boson decay is off its mass shell. Lastly, because of spin correlation (see Section 5.2) the neutrinos from the Higgs decay can be assumed to be collinear, reducing the degrees of freedom by a further 2. This leaves 2 degrees of freedom; hence, even with these assumptions  $p_T^W$  cannot be reconstructed analytically.

### 6.2. Proxies

Since the analytical reconstruction of  $p_T^W$  is impossible, a proxy needs to be found, i.e. some observable that correlates strongly with  $p_T^W$ . For example  $p_T^{l_2}$ , the transverse momentum of the lepton from the associated  $W$ , will tend to increase when  $p_T^W$  does. The different STXS categories have different distributions in such a variable, which allows a fit to distinguish between the STXS bins and thus enables an STXS measurement. This is shown in Fig. 6.1:  $p_T^{l_2}$  correlates with  $p_T^W$ , which causes the STXS categories to have different distributions in  $p_T^{l_2}$ . The lowest  $p_T^W$  bin has a narrow

## 6. Reconstruction of the $W$ Boson Momentum

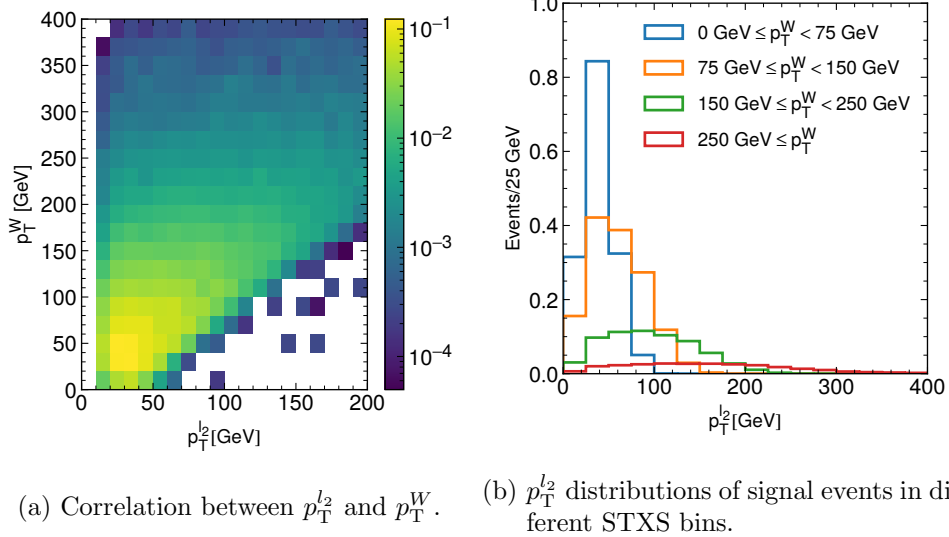


Figure 6.1.: Distributions of  $p_T^{l2}$  as a proxy for  $p_T^W$ .

distribution at low  $p_T^{l2}$ , the higher  $p_T^W$  bins have broader distributions at higher values of  $p_T^{l2}$ .

A number of proxies were tested. They all are kinematic variables involving lepton 2:

- $p_T^{l2}$ , the transverse momentum of lepton 2, which originates from the associated  $W$  boson
- $\Delta\phi_{l2,\text{MET}}$ , the difference in azimuthal angle between lepton 2 and the missing transverse energy
- $\Delta R_{l_0 l_2}$  and  $\Delta R_{l_1 l_2}$ , the angular separations between lepton 2 and the other leptons
- $m_{l_0 l_2}$  and  $m_{l_1 l_2}$ , the invariant masses of lepton 2 and one of the other leptons
- $m_{lll} = \sqrt{(p_{l_0}^\mu + p_{l_1}^\mu + p_{l_2}^\mu)^2}$ , the invariant mass of all three leptons
- $m_T^{l_0 l_2}$  and  $m_T^{l_1 l_2}$ , the transverse masses of lepton 2 and one of the other leptons. In this analysis, the transverse mass of a lepton pair is modified with respect to Equation 3.7, to approximately take into account the transverse momenta of the neutrinos [44]. It is defined as

$$m_T^{l_1 l_2} = \sqrt{\left(E_T^{l_1 l_2} + E_T^{\text{miss}}\right)^2 - \left|\vec{p}_T^{l_1} + \vec{p}_T^{l_2} + \vec{E}_T^{\text{miss}}\right|^2}, \quad (6.3)$$

where  $E_T^{l_1 l_2} = \sqrt{(\vec{p}_T^{l_1} + \vec{p}_T^{l_2})^2 + m_{l_1 l_2}^2}$

## 6. Reconstruction of the $W$ Boson Momentum

- $p_T^{l_2} \cos(\Delta\phi_{l_2, \text{MET}})$ , the projection of  $p_T^{l_2}$  on the direction of the missing transverse energy
- $p_T^{l_2} + E_T^{\text{miss}}/2$ , the transverse momentum of lepton 2 plus half the missing transverse energy as a rough estimate of the energy of neutrino 2
- $\sum E_T$ , the total transverse energy, defined as the scalar sum of the lepton transverse momenta and the missing transverse energy
- $p_T^{l_2} / \sum E_T$ , the ratio between the lepton 2 transverse momentum and the total transverse energy
- $\vec{p}_T^{l_2} \cdot \frac{\vec{p}_T^{l_0} + \vec{p}_T^{l_1}}{|\vec{p}_T^{l_0} + \vec{p}_T^{l_1}|}$ , the projection of the transverse momentum of lepton 2 on the total transverse momentum of leptons 0 and 1

To compare the performance of the different proxies, a series of fits using Asimov data [35] were performed. The likelihood used in these fits is defined as

$$L(\mu) = \prod_j \frac{\left(\sum_{i=1}^4 \mu_i s_{i,j}\right)^{n_j}}{n_j!} e^{-(\sum_{i=1}^4 \mu_i s_{i,j})}, \quad (6.4)$$

where  $s_{i,j}$  is the predicted number of events of STXS category  $i$  in bin  $j$  of the proxy. The Asimov data is given by  $n_j = \sum_i s_{i,j}$ . No background processes were entered into the fit. For every proxy the same number of bins was used, as a larger number of bins can contain more information on the shapes of the distributions and a difference in the number of bins between the proxies could therefore bias the comparison towards those that were given more bins.

Every fit results in values of 1 for each of the four parameters with uncertainties  $\sigma_\mu$ , calculated using Equation 4.4, which can be taken as a measure of the performance of the proxy. To estimate the best possible performance, the same procedure was applied to the Monte Carlo truth  $p_T^W$ , which results in the uncertainties that could be achieved with perfect reconstruction.

The uncertainties derived in this way are shown for a selection of proxies in Fig. 6.2. The best proxy was found to be  $\sum E_T$ . However, compared to the ideal case there is room for improvement, especially in the higher two STXS bins. This motivates the use of more sophisticated methods such as a regression network.

## 6. Reconstruction of the $W$ Boson Momentum

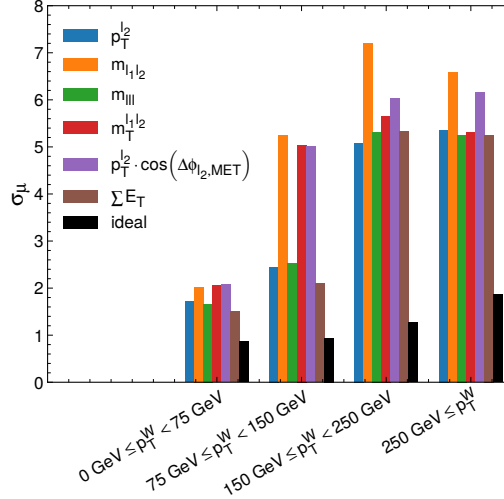


Figure 6.2.: Uncertainties derived from signal-only Asimov fits for different proxies.

## 6.3. Regression Network for $p_T^W$ Reconstruction

### 6.3.1. Training Sample

The regression network was trained on  $WH \rightarrow WWW \rightarrow l\nu l\nu l\nu$  MC samples from the  $Z$  depleted signal region. They were generated using POWHEG and PYTHIA8 [29]. 319202 raw MC events were available. 50000 of these were set aside as a testing data set and not used in the training. The MC data contains information about the true  $p_T^W$  as used by the Monte Carlo generator, which the network was trained to reconstruct.

### 6.3.2. Preprocessing

Before the training, the data was processed using the `RobustScaler` from `scikit-learn` [45]. It scales the inputs to the ANN such that the median becomes zero and the interquartile range becomes 1. Treating the input variables in such a way is called standardization and common practice in machine learning. It ensures that all input variable are of a similar size, as opposed to the large numerical differences between e.g. an angle and a momentum. While the neural network can in theory compensate for such differences by scaling the weights accordingly, standardization has practical advantages such as ensuring a good behavior of the numerical minimization and a good choice of the initial parameters of the network [46].

Since the cross section of the  $WH \rightarrow WWW \rightarrow l\nu l\nu l\nu$  process decreases with higher  $p_T^W$ , events from higher STXS bins have a smaller sum of weights. This might cause the regression network training to overly focus on the lower STXS bins. To avoid this, the training events were reweighted, so that each STXS bin has the same sum of weights.

## 6. Reconstruction of the $W$ Boson Momentum

### 6.3.3. Training and Hyperparameter Optimization

The regression network was trained using `tensorflow` [47]. Mean squared error was used as the loss function. The Adam algorithm [39] was used for optimization. To avoid overfitting, early stopping was applied, with the training stopping once the validation loss had not improved for 5 training epochs. An epoch is an iteration over the entire training sample. Validation samples were defined using 2-fold cross validation.

The choice of input parameters and the architecture were optimized by systematically training networks with different options. The ANNs were then evaluated on the validation sample. The mean squared error loss was averaged over the  $k$ -folds, giving a figure of merit for the performance of the training variant. The different variants were then compared using this figure of merit.

### 6.3.4. Input Variables

The following groups of kinematic variables were tested as inputs to the regression network:

- $p_T^{l_0}, p_T^{l_1}$  and  $p_T^{l_2}$ , the lepton transverse momenta
- $E_T^{\text{miss}}$ , the missing transverse energy
- $\Delta\phi_{l_0, \text{MET}}, \Delta\phi_{l_1, \text{MET}}$  and  $\Delta\phi_{l_2, \text{MET}}$ , the differences in azimuthal angle between the leptons and the missing transverse energy
- $\phi_{\text{MET}}$ , the azimuthal angle of the missing transverse energy
- $m_{l_0 l_1}, m_{l_0 l_2}$  and  $m_{l_1 l_2}$ , the invariant masses of the lepton pairs
- $m_{lll}$ , the invariant mass of all three leptons
- $m_T^{l_0 l_1}, m_T^{l_0 l_2}$  and  $m_T^{l_1 l_2}$ , the transverse masses of the lepton pairs, as defined in Equation 6.3
- $\Delta\eta_{l_0 l_1}, \Delta\eta_{l_0 l_2}$  and  $\Delta\eta_{l_1 l_2}$ , the differences in pseudorapidity between leptons
- $\Delta\phi_{l_0 l_1}, \Delta\phi_{l_0 l_2}$  and  $\Delta\phi_{l_1 l_2}$ , the differences in azimuthal angle between leptons
- $\Delta R_{l_0 l_1}, \Delta R_{l_0 l_2}$  and  $\Delta R_{l_1 l_2}$ , the angular separations between leptons
- $\sum_i p_T^i$ , the scalar sum of lepton transverse momenta

The lepton momenta and the missing transverse energy were always included. Each training variant additionally had 3 to 6 of the above groups of variables. The best set of input variables, as determined by comparing the MSE of the variants, is shown in Table 6.1. The distributions of all input variables are shown in the appendix.

## 6. Reconstruction of the $W$ Boson Momentum

Table 6.1.: STXS regression network input variables.

| $p_{\text{T}}^{l_0}$           | $p_{\text{T}}^{l_1}$           | $p_{\text{T}}^{l_2}$           |
|--------------------------------|--------------------------------|--------------------------------|
| $m_{l_0 l_1}$                  | $m_{l_0 l_2}$                  | $m_{l_1 l_2}$                  |
| $\Delta\phi_{l_0, \text{MET}}$ | $\Delta\phi_{l_1, \text{MET}}$ | $\Delta\phi_{l_2, \text{MET}}$ |
| $\Delta R_{l_0 l_1}$           | $\Delta R_{l_0 l_2}$           | $\Delta R_{l_1 l_2}$           |
| $m_{\text{T}}^{l_0 l_1}$       | $m_{\text{T}}^{l_0 l_2}$       | $m_{\text{T}}^{l_1 l_2}$       |
| $E_{\text{T}}^{\text{miss}}$   | $\sum_i p_{\text{T}}^{l_i}$    |                                |

### 6.3.5. Architecture

Different network architectures were tested, which differed by the number and size of the hidden layers. The networks were constructed from 3 to 8 hidden layers. Each layer could have 128, 80, 64, 48, 32 or 16 neurons. Each layer had a number of neurons smaller or equal to that of the previous hidden layer. This is based on the assumption that later layers encode high-level properties of the kinematic regions, each layer combines the information from the previous layer into fewer compound quantities. This is similar to how the result of an analytical calculation would consist of nested, more basic functions. The total size of the network was limited to at most 100000 parameters. All hidden layers used the swish activation function.

To limit the computational effort, not the entirety of these architectures was trained. Instead, for each number of hidden layers the smallest network was trained first. Then the size of the network was gradually increased, until the size increase no longer improved the network performance, as determined by the mean squared error not decreasing for 20 architectures. This method of searching for the best architecture was adapted from a previous thesis in which classifiers for signal-background separation were trained in this way [48].

### 6.3.6. Regression Network Performance

The performance of the regression network was evaluated by using it as a proxy and performing signal-only fits as described in Section 6.2. These fits were performed on the testing sample. To ensure that the results can be compared, the same sample was used for the fits with the other proxies.

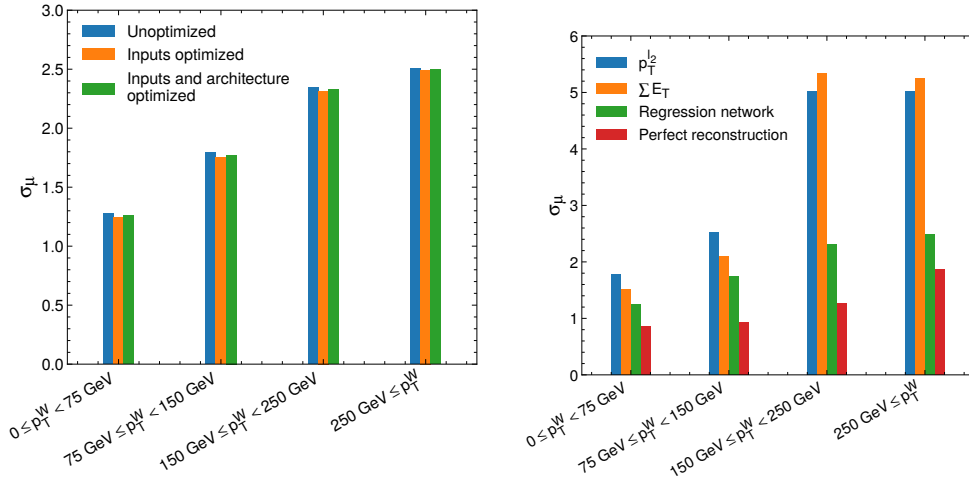
The regression network output  $p_{\text{T}}^R$  was entered into the fit before hyperparameter optimization, after optimization of the inputs but before the full optimization including the architecture and after the optimization of inputs and architecture. The result of this is shown in Fig. 6.3a. The optimization did not strongly affect the network performance in the fits. The optimization of the architecture led to a slight decrease in performance compared to the optimization of only the input variable set. This suggests that an improvement in the mean squared error, which was used as a figure of merit in the architecture optimization, does not necessarily indicate an improvement

## 6. Reconstruction of the $W$ Boson Momentum

of the performance in the fits. The network with only the inputs optimized was used in the further analysis. Its architecture is shown in Table 6.2.

As shown in Fig. 6.3b, the regression network output  $p_T^R$  significantly outperforms the previously best proxy  $\sum E_T$ . For this reason, it was used as the proxy for  $p_T^W$  in the further analysis steps.

Since the  $Z$  depleted and  $Z$  dominated signal regions differ only by the number of SFOS pairs, the kinematic properties of the signal process are the same for both regions. The regression network can therefore also be applied in the  $Z$  dominated signal region and is used in the  $Z$  dominated STXS analysis [29].



(a) Comparison between regression networks at different stages of the optimization. (b) Comparison with different proxies and perfect reconstruction.

Figure 6.3.: Performance of the regression network output  $p_T^R$  as a proxy for  $p_T^W$  in signal-only Asimov fits.

Table 6.2.: Architecture of the optimized regression network.

|                     | Input layer | Hidden layers |     |    |    |    |    |    |    | Output layer |
|---------------------|-------------|---------------|-----|----|----|----|----|----|----|--------------|
| Number of neurons   | 17          | 128           | 128 | 64 | 64 | 32 | 32 | 16 | 16 | 1            |
| Activation function | swish       | swish         |     |    |    |    |    |    |    | identity     |

## 6. Reconstruction of the $W$ Boson Momentum

### 6.3.7. Multiclassifier for STXS Bins

As an alternative to the regression network, the use of a multiclassifier network to distinguish between STXS bins was investigated. Its input variables and architecture are shown in Table 6.3 and Table 6.4. The multiclassifier network has four outputs. The highest output indicates that an event belongs to the corresponding STXS bin.

To enter the four-dimensional output of this network into the signal-only fit, the following binning was applied: First the events were sorted by which output was the highest. Then they were further divided by the value of the highest output node, i.e. how “certain” the network was about the choice of STXS bin. The resulting distribution is shown in Fig. 6.4.

The fit results of the multiclassifier were slightly better than those of the unoptimized regression network, as shown in Fig. 6.5. However, since the difference was small and the regression network has the advantage of having a one-dimensional output, which significantly simplifies the following steps of the analysis, it was decided not to use a multiclassifier for the STXS reconstruction. Neither network had had its parameters optimized when this comparison was made.

Table 6.3.: STXS multiclassifier input variables.

| $p_{\text{T}}^{l_0}$          | $p_{\text{T}}^{l_1}$          | $p_{\text{T}}^{l_2}$          |
|-------------------------------|-------------------------------|-------------------------------|
| $\Delta\phi_{l_0,\text{MET}}$ | $\Delta\phi_{l_1,\text{MET}}$ | $\Delta\phi_{l_2,\text{MET}}$ |
|                               | $E_{\text{T}}^{\text{miss}}$  |                               |

Table 6.4.: Architecture of the STXS multiclassifier network.

|                     | Input layer | Hidden layers |    |    |    |    |    | Output layer |
|---------------------|-------------|---------------|----|----|----|----|----|--------------|
| Number of neurons   | 17          | 64            | 64 | 32 | 32 | 16 | 16 | 4            |
| Activation function | swish       | swish         |    |    |    |    |    | softmax      |

## 6. Reconstruction of the $W$ Boson Momentum

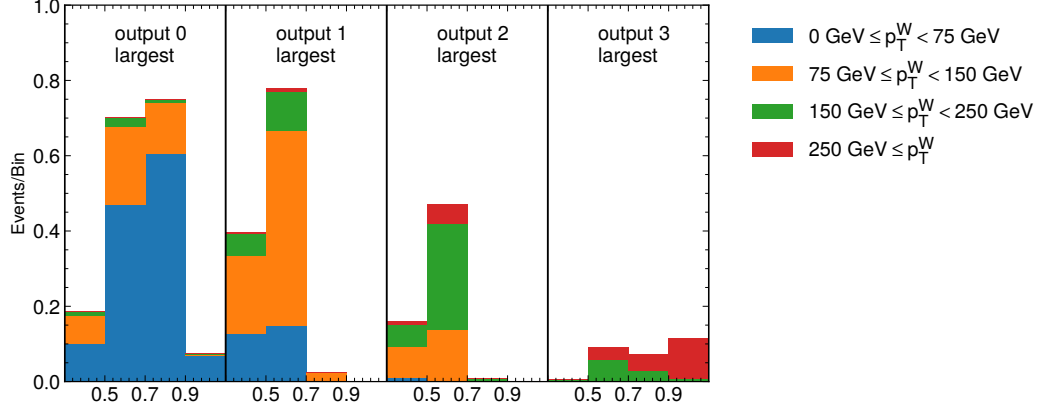


Figure 6.4.: Distribution of the outputs of the STXS multiclassifier that was entered into the fit. The events are first binned based on which output node is highest and then by the value of the highest node.

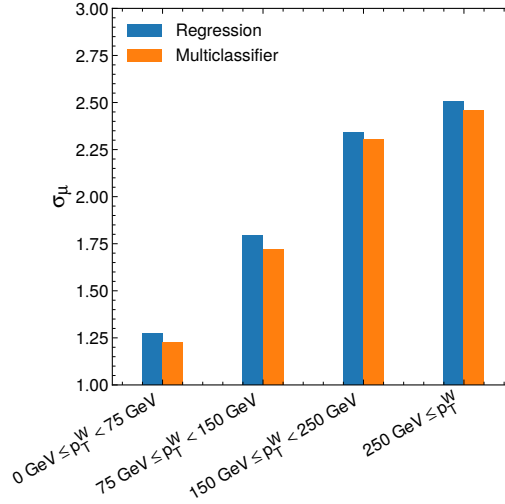


Figure 6.5.: Comparison of the fit results of the STXS multiclassifier and the unoptimized regression network.

## 7. Analysis in the $Z$ Depleted Signal Region

### 7.1. Monte Carlo Samples

A number of different Monte Carlo (MC) generators were used in the  $VH H \rightarrow WW$  analysis, depending on the process. The POWHEG [49] and MADGRAPH5\_aMC@NLO [50] programs generate the hard scattering process at the parton level. Their output is then processed by PYTHIA8 [51], which generates the parton shower, i.e. the emission of additional partons, hadronization, and the underlying event, i.e. the processes apart from the hard scatter. SHERPA [52] generates the entire event.

Table 7.1 lists which generators are used for which processes. Also listed are the products of cross section and branching ratio  $\sigma \cdot \text{Br}$  of the processes, as well as the precision to which these were calculated, i.e. Leading Order (LO), Next to Leading Order (NLO) etc. The samples are processed through a simulation of the ATLAS detector and then reconstructed.

Table 7.1.: MC generators used for different processes in the  $VH H \rightarrow WW$  analysis [29].

| Process                                                                   | Generator                  | $\sigma \cdot \text{Br}$ (pb) | Precision $\sigma_{\text{incl.}}$ |
|---------------------------------------------------------------------------|----------------------------|-------------------------------|-----------------------------------|
| $WH H \rightarrow WW$                                                     | POWHEG +PYTHIA8 (MINLO)    | 0.293                         | NNLO                              |
| $ZH H \rightarrow WW$                                                     | POWHEG +PYTHIA8 (MINLO)    | 0.189                         | NNLO                              |
| $ggF H \rightarrow WW$                                                    | POWHEG +PYTHIA8 NNLOPS     | 10.4                          | N3LO+NNLL                         |
| $VBF H \rightarrow WW$                                                    | POWHEG +PYTHIA8            | 0.808                         | NNLO                              |
| Inclusive $Z/\gamma^{(*)} \rightarrow ll$ ( $40 \geq m_{ll} \geq 10$ GeV) | SHERPA 2.2.1               | $6.80 \times 10^3$            | NNLO                              |
| Inclusive $Z/\gamma^{(*)} \rightarrow ll$ ( $m_{ll} \geq 40$ GeV)         | SHERPA 2.2.1               | $2.107 \times 10^3$           | NNLO                              |
| $(W \rightarrow l\nu)\gamma$ ( $p_T^\gamma > 7$ GeV)                      | SHERPA 2.2.8               | 1071                          | NLO                               |
| $(Z \rightarrow ll)\gamma$ ( $p_T^\gamma > 7$ GeV)                        | SHERPA 2.2.8               | 353                           | NLO                               |
| $t\bar{t}$ di-leptonic ( $e, \mu, \tau$ )                                 | POWHEG +PYTHIA8            | 87.6                          | NNLO+NNLL                         |
| $Wt$ leptonic                                                             | POWHEG +PYTHIA8            | 7.55                          | NLO                               |
| $t\bar{t}V$ leptonic                                                      | MADGRAPH5_aMC@NLO +PYTHIA8 | 0.67                          | NLO                               |
| $tZ$ leptonic                                                             | MADGRAPH5_aMC@NLO +PYTHIA8 | 0.21                          | NLO                               |
| $tWZ$ leptonic                                                            | MADGRAPH5_aMC@NLO +PYTHIA8 | 0.016                         | NLO                               |
| $q\bar{q}/g \rightarrow WW \rightarrow l\nu l\nu$                         | SHERPA 2.2.2               | 12.7                          | NNLO                              |
| $q\bar{q} \rightarrow WW q\bar{q} \rightarrow l\nu l\nu jj$               | SHERPA 2.2.2               | 0.18                          | NNLO                              |
| $gg \rightarrow WW \rightarrow l\nu l\nu$                                 | SHERPA 2.2.2               | 0.60                          | LO                                |
| $gg \rightarrow ZZ \rightarrow llll$                                      | SHERPA 2.2.2               | 0.02                          | LO                                |
| $q\bar{q}/g \rightarrow WZ/W\gamma^{(*)} \rightarrow l\nu ll$             | SHERPA 2.2.2               | 7.54                          | NLO                               |
| $q\bar{q}/g \rightarrow ZZ/Z\gamma^{(*)} \rightarrow llll$                | SHERPA 2.2.2               | 2.71                          | NLO                               |
| $VVV$                                                                     | SHERPA 2.2.2               | 0.11                          | NLO                               |

## 7. Analysis in the $Z$ Depleted Signal Region

### 7.2. $Z$ Depleted Signal Region Selections

The  $Z$  depleted signal region is defined by the following selections [29]: firstly, there need to be exactly three isolated leptons, with their charges adding up to  $\pm 1$ . Each lepton is required to have a transverse momentum greater than 15 GeV. Single lepton triggers were used. One of the leptons needs to match the triggering lepton. These requirements strongly suppress any backgrounds that would require two fake leptons, since the simultaneous misidentification of two leptons is exceedingly unlikely.

Background processes involving the decays of  $Z$  bosons can be suppressed by requiring that there are no lepton pairs with the same flavor and opposite charge (SFOS). However, this also removes about 3/4 of the signal events. This can be seen from Table 7.2: out of the eight possible combinations of lepton flavors in the signal process, only the two in which lepton 0 has a unique flavor pass the 0 SFOS pair requirement. For this reason, the analysis is split into a  $Z$  dominated signal region with 1 or 2 SFOS pairs and a  $Z$  depleted signal region with 0 SFOS pairs. The analysis in each region can then be adapted to treat the respective dominant backgrounds. This thesis is concerned with the  $Z$  depleted analysis only.

Table 7.2.: Possible combinations of lepton flavors and charges. Lepton 0 has unique charge, leptons 1 and 2 have the same charge.

| Lepton |       |       | Number of<br>SFOS pairs |
|--------|-------|-------|-------------------------|
| 0      | 1     | 2     |                         |
| $e$    | $e$   | $e$   | 2                       |
| $e$    | $e$   | $\mu$ | 1                       |
| $e$    | $\mu$ | $e$   | 1                       |
| $\mu$  | $e$   | $e$   | 0                       |
| $\mu$  | $\mu$ | $e$   | 1                       |
| $\mu$  | $e$   | $\mu$ | 1                       |
| $e$    | $\mu$ | $\mu$ | 0                       |
| $\mu$  | $\mu$ | $\mu$ | 2                       |

### 7.3. Control Regions

To assess the agreement between the Monte Carlo simulations and real measurements, Control Regions (CR) and Validation Regions (VR) are defined. These are regions in phase space which are enhanced in a certain background process. Crucially, the signal expectation in a control or validation region is negligibly small, so that one may look at the experimental data in the region while the analysis is blinded. A CR is additionally used to constrain the normalization of a background process in a fit. For the  $Z$  depleted  $WH \rightarrow WWW \rightarrow l\nu l\nu l\nu$  analysis, two such regions are defined, the  $WZ$  CR and the Top VR [29]. Both share the requirement of the SR of three leptons

## 7. Analysis in the $Z$ Depleted Signal Region

with a minimum momentum and the correct total charge, but then invert one of the further selections of the signal region to enhance the background of interest instead of suppressing it.

The  $WZ$  control region requires at least one SFOS pair, with an invariant mass differing from the  $Z$  mass by at most 25 GeV, thus only events are selected which have a lepton pair that likely originates from a  $Z$  decay. Further requirements are that no opposite charge lepton pair has an invariant mass below 12 GeV, that the missing transverse energy is greater than 30 GeV and that there is at most one jet. Top related backgrounds are suppressed in the  $WZ$  CR by requiring that there are no jets identified to originate from a  $b$  quark.

The Top VR has the same 0 SFOS requirement as the signal region, to avoid contributions from the  $WZ$  background. It is then simply defined by requiring the presence of a  $b$  jet.

Fig. 7.1 shows the distributions of the regression network output  $p_T^R$  in these regions. The agreement between data and MC is found to be good, though in the Top VR the comparison is made difficult by the low number of events in the region. The agreement in the  $WZ$  CR is likely to improve somewhat after the full statistical treatment when the  $WZ$  normalization has been scaled using this control region; this is beyond the scope of this thesis, however. Tables with the yields of the different processes are shown in Appendix B.

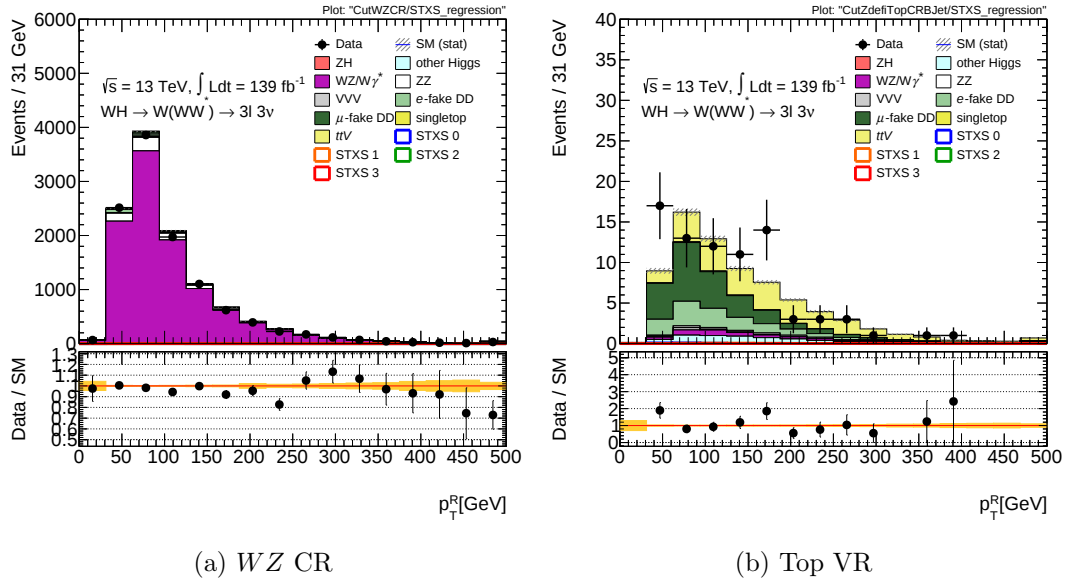


Figure 7.1.: Distributions of the regression network output  $p_T^R$  in the control and validation regions.

## 7.4. Multiclassifier for Background Separation

A multiclassifier separating the signal from the most important background processes was previously trained for the  $Z$  depleted inclusive analysis [29, 48], which from here on out will be referred to as the old multiclassifier. It was trained using  $WH$ ,  $WZ$  and  $t\bar{t}$  samples. Correspondingly, it has three output nodes, one for each of the processes.

### 7.4.1. Inputs

The inputs of the old multiclassifier include all the kinematic variables also used for the regression network, as well as the following additional variables. The input variables are summarized in Table 7.3 and their distributions are shown in Appendix A.

- $\Delta\eta_{l_0l_1}$ ,  $\Delta\eta_{l_0l_2}$  and  $\Delta\eta_{l_1l_2}$ , the differences in pseudorapidity between leptons
- $m_{lll}$ , the invariant mass of all three leptons
- $d_0/\sigma(d_0)_{l_0}$ ,  $d_0/\sigma(d_0)_{l_1}$  and  $d_0/\sigma(d_0)_{l_2}$  are the transverse impact parameters of the leptons, divided by their uncertainties.
- $z_0 \sin \theta_{l_0}$ ,  $z_0 \sin \theta_{l_1}$  and  $z_0 \sin \theta_{l_2}$ , the longitudinal impact parameters of the leptons. They are multiplied with  $\sin \theta$  to account for the fact that the resolution of  $z_0$  worsens as the lepton is emitted closer to the beam axis.
- $p_T^{\max}(\text{jet})$ , the highest transverse momentum of any jet in the event.
- $n_{\text{jets}}$ , the number of jets in the event.
- $n_{b\text{-jets}}$ , the number of jets identified as originating from a  $b$  quark using the DL1r algorithm [53] at the 85% efficiency working point.
- $S/E_T^{\text{miss}}$ , the significance of the missing transverse energy divided by  $E_T^{\text{miss}}$ .
- $\phi_{\text{MET}}$ , the azimuthal angle of the missing transverse energy.
- $\max(|1 - E^{l_i}/p^{l_i}|)$ , the highest mismatch between the energy of a lepton as measured by the calorimeter and its momentum. For muons, which pass through the calorimeter,  $E/p$  is set to 1.
- $F_{\alpha_1}$ , a variable designed to suppress the  $WZ$  background with  $Z \rightarrow \tau\tau$ . It is obtained by reconstructing the  $Z$  mass from the lepton pair assumed to originate from the  $Z$ , under the approximation that the lepton from a  $\tau$  decay is emitted in the same direction as the  $\tau$ . There are not enough kinematic constraints to determine all unknowns in this calculation. One variable, chosen to be  $\alpha_1 = E_\tau^{(1)}/E_l^{(1)}$ , the ratio between the energy of the lepton with lower  $p_T$  from the pair and the energy of the  $\tau$  it originates from, remains unconstrained. Several values for this variable are tried, only for some of which the system of equations can be solved and yields physical results.  $F_{\alpha_1}$  is then the fraction of attempts that was successful [29].

## 7. Analysis in the $Z$ Depleted Signal Region

Table 7.3.: Input variables of the old multiclassifier.

|                                 |                                 |                                 |                                |
|---------------------------------|---------------------------------|---------------------------------|--------------------------------|
| $p_T^{l_0}$                     | $p_T^{l_1}$                     | $p_T^{l_2}$                     | $E_T^{\text{miss}}$            |
| $\Delta R_{l_0 l_1}$            | $\Delta R_{l_0 l_2}$            | $\Delta R_{l_1 l_2}$            | $\phi_{\text{MET}}$            |
| $\Delta \eta_{l_0 l_1}$         | $\Delta \eta_{l_0 l_2}$         | $\Delta \eta_{l_1 l_2}$         | $S/E_T^{\text{miss}}$          |
| $\Delta \phi_{l_0, \text{MET}}$ | $\Delta \phi_{l_1, \text{MET}}$ | $\Delta \phi_{l_2, \text{MET}}$ | $p_T^{\text{max}}(\text{jet})$ |
| $d_0/\sigma(d_0)_{l_0}$         | $d_0/\sigma(d_0)_{l_1}$         | $d_0/\sigma(d_0)_{l_2}$         | $n_{\text{jets}}$              |
| $z_0 \sin \theta_{l_0}$         | $z_0 \sin \theta_{l_1}$         | $z_0 \sin \theta_{l_2}$         | $n_{\text{b-jets}}$            |
| $m_T^{l_0 l_1}$                 | $m_T^{l_0 l_2}$                 | $m_T^{l_1 l_2}$                 | $\sum_i p_T^{l_i}$             |
| $m_{l_0 l_1}$                   | $m_{l_0 l_2}$                   | $m_{l_1 l_2}$                   | $m_{lll}$                      |
|                                 | $F_{\alpha_1}$                  | $\max( 1 - E^{l_i}/p^{l_i} )$   |                                |

### 7.4.2. Architecture and Training Setup

The multiclassifier was not trained on the MC samples used in the  $Z$  depleted analysis in order to preserve statistical independence. For the  $WH$  and  $t\bar{t}$  processes, which do not involve  $Z$  bosons, events from the  $Z$  dominated signal region were used. For the  $WZ$  process, a dedicated MC sample was used. A looser isolation working point, FCLoose [54], was applied to the training samples to increase their size.

The architecture of the old multiclassifier is shown in Table 7.4. It was optimized in a gridsearch. As a loss function, the categorical cross entropy was used. The training used 5-fold cross validation. Overfitting was avoided by employing dropout. Like in the training of the regression network, the training events were preprocessed using the **RobustScaler**. The samples were reweighted, so that each process has the same sum of weights. This avoids biasing the network toward one class.

Table 7.4.: Architecture of the old multiclassifier.

|                     | Input layer | Hidden layers |     |      | Output layer |
|---------------------|-------------|---------------|-----|------|--------------|
| Number of neurons   | 34          | 80            | 32  | 16   | 3            |
| Dropout rate        | -           | 0.37          | 0.3 | 0.26 | -            |
| Activation function | swish       | swish         |     |      | softmax      |

## 7.5. Treatment of Multiclassifier and Regression Network Outputs

Applying both the regression network and the multiclassifier network to an event results in a total of four new variables:  $p_T^R$ , the  $WH$  node, the  $t\bar{t}$  node and the  $WZ$  node. These can be used either to define additional selections in order to further reduce the backgrounds or to define bins to be used in a fit.

The outputs of the old multiclassifier are shown in Fig. 7.2. The  $t\bar{t}$  node clearly separates the backgrounds involving top quarks and fake leptons from the signal. It

## 7. Analysis in the $Z$ Depleted Signal Region

can therefore be used to simply define a cut, discarding all events above a certain  $t\bar{t}$  node threshold. A value of 0.3 was chosen for this threshold; this value was optimized for the inclusive analysis.

For the  $WH$  and  $WZ$  nodes, such a simple cut is not used, since the signal distributions are much wider. Instead, the output nodes are combined: while the multiclassifier has three output nodes, the use of a softmax activation ensures that the outputs always add to 1. This makes the output effectively two-dimensional. A new quantity called the  $WH - WZ$  node is constructed by subtracting the  $WZ$  node from the  $WH$  node. This reduces the  $WH$  and  $WZ$  nodes to a single discriminant with better discrimination than the individual nodes.

Fig. 7.3 shows the distributions of the total  $WH$  signal, the sum of all background processes and the ratio between total signal and total background  $S/B$  in the plane of  $p_T^R$  and the  $WH - WZ$  node. The outputs of the two ANNs appear to be largely uncorrelated. The regression network does not distinguish between signal and background: both peak at low  $p_T^R$ , while  $S/B$  is approximately flat in  $p_T^R$ . Conversely, the multiclassifier does not distinguish between the different STXS bins, as can also be seen in Fig. 7.4.

Because of this, the networks can simply be treated separately. First  $p_T^R$  is used to split the signal region into four bins. In each of these bins one of the four STXS signals is dominant. The bin borders were defined as 90 GeV, 180 GeV and 270 GeV. Within each of these bins, the  $WH - WZ$  node is then used as a discriminant and the combination of the four  $p_T^R$  regions is entered into the fit.

A preliminary binning of the  $WH - WZ$  node in the different  $p_T^R$  bins was defined by first applying a fine binning and then manually merging bins with a similar signal to background ratio  $S/B$ . It is shown in Fig. 7.5. The purpose of the merging is to increase the expected number of both expected events and raw Monte Carlo events in the bins, which decreases the relative statistical uncertainties on data and the MC prediction, respectively. However, since some information about the shape of the distribution is lost when merging, this somewhat increases the uncertainties of the fit results.

## 7. Analysis in the Z Depleted Signal Region

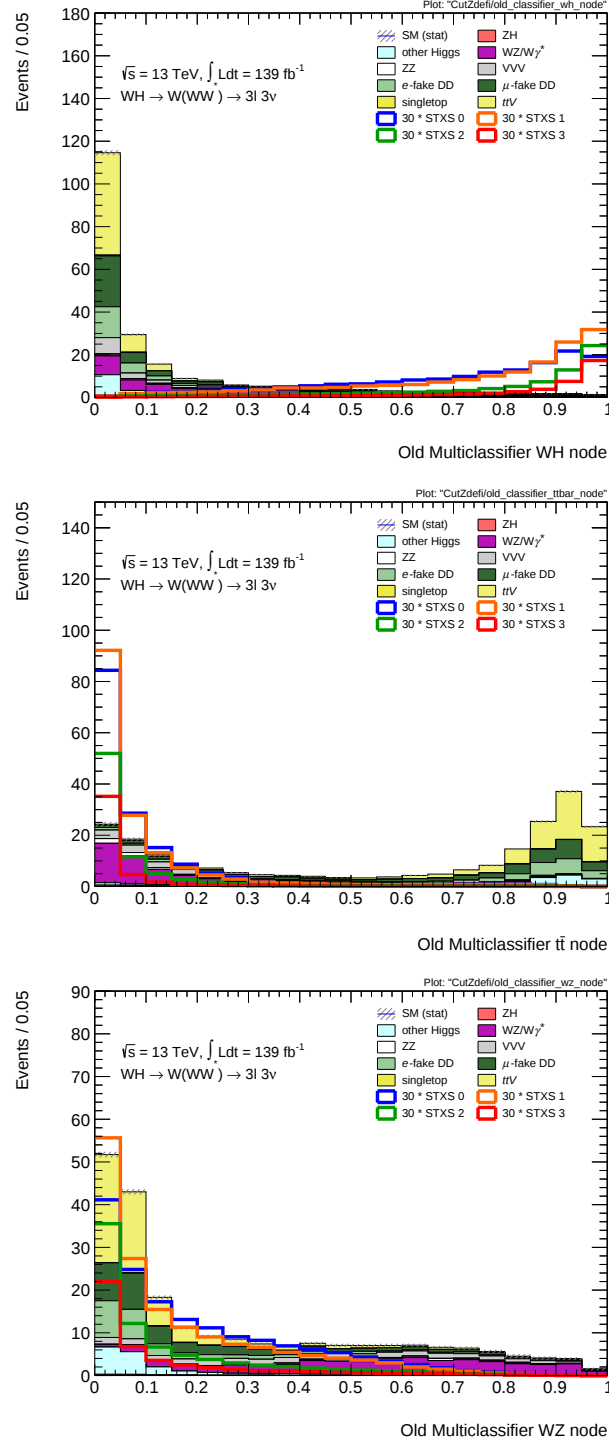


Figure 7.2.: Output distributions of the old multiclassifier network. The 0 SFOS requirement is applied.

### 7. Analysis in the $Z$ Depleted Signal Region

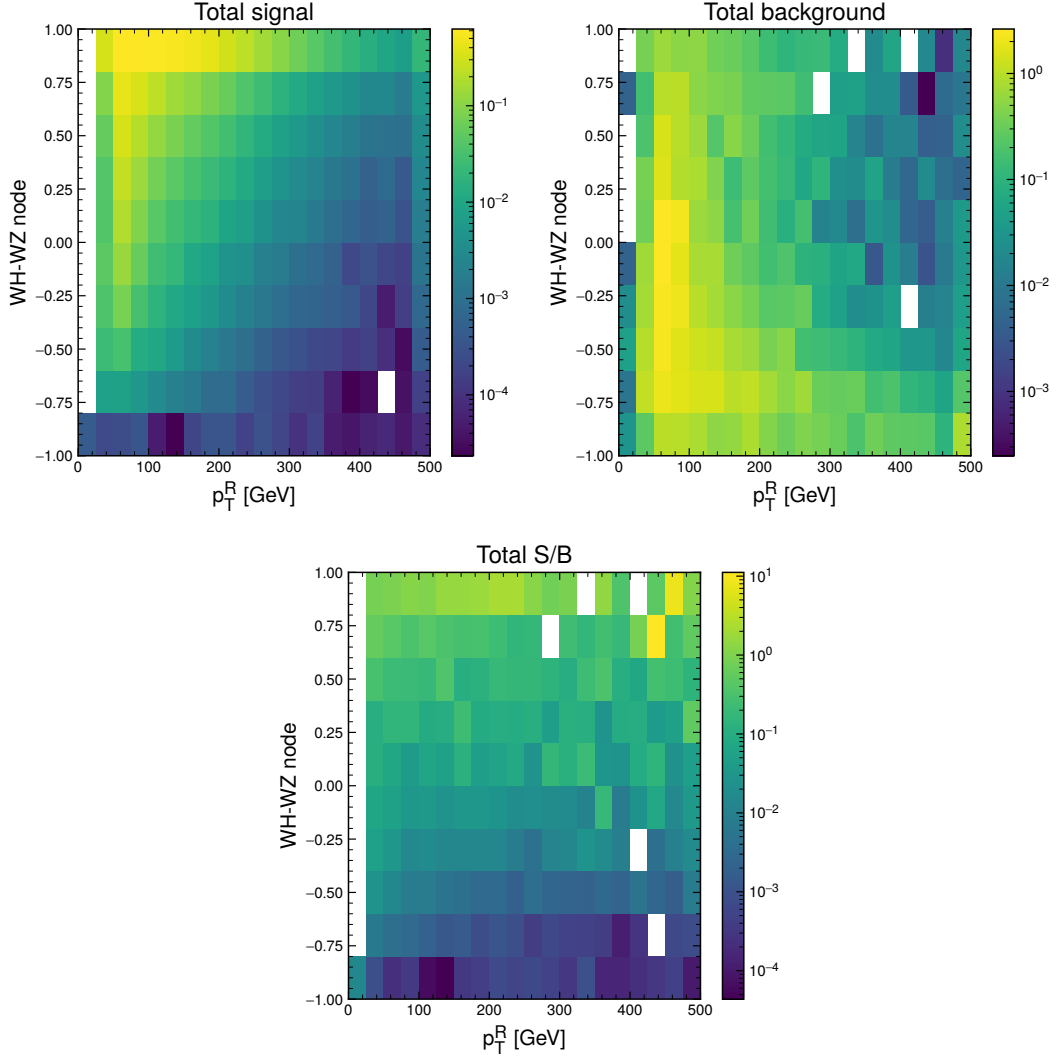


Figure 7.3.: Distributions of the total signal, the total background, and the ratio between the two in the plane spanned by  $p_T^R$  and  $WH - WZ$  node. A cut of  $t\bar{t}$  node  $\leq 0.3$  has been applied.

## 7. Analysis in the $Z$ Depleted Signal Region

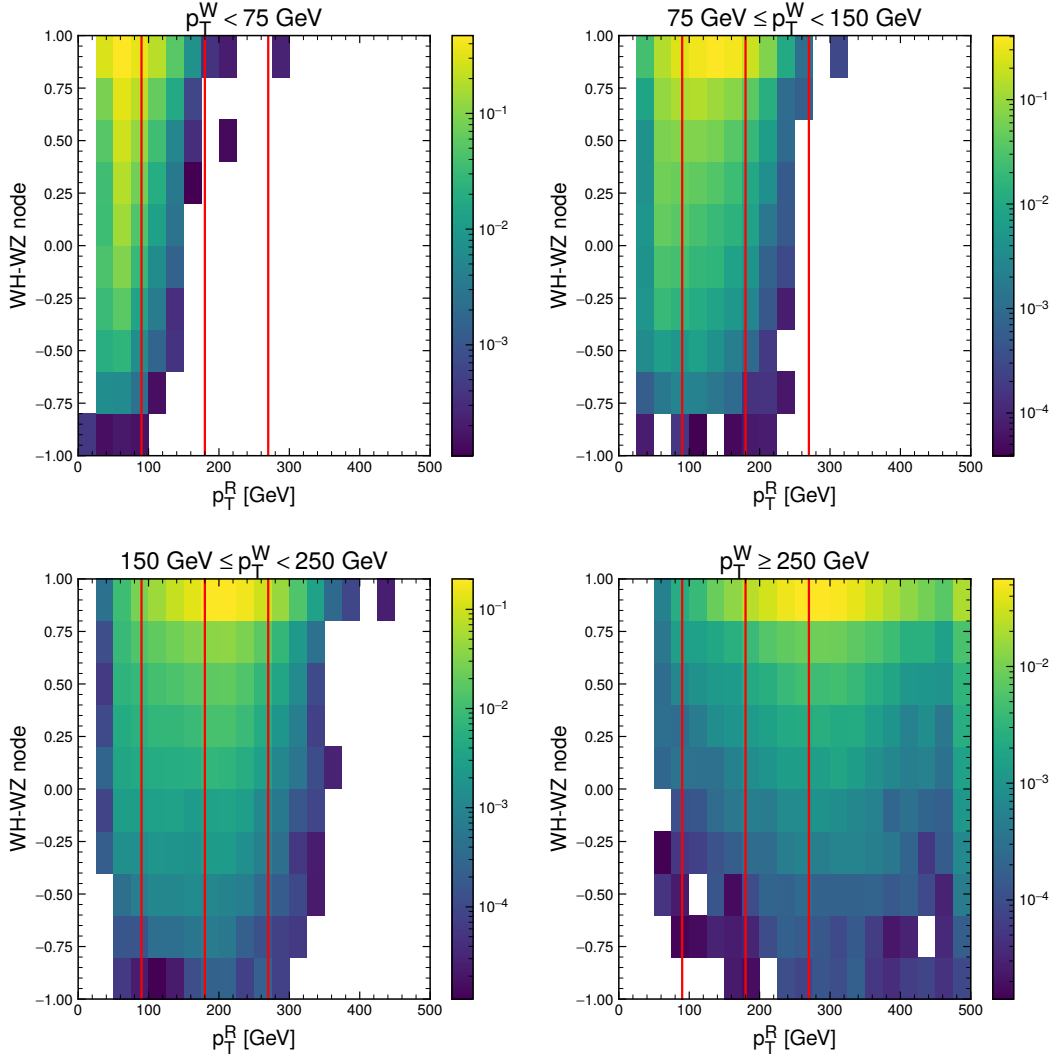


Figure 7.4.: Distributions of the signal in the plane spanned by  $p_T^R$  and  $WH - WZ$  node, split by STXS bins. The red lines indicate the definitions of the  $p_T^R$  bins with borders at 90, 180 and 270 GeV. A cut of  $t\bar{t}$  node  $\leq 0.3$  has been applied.

### 7. Analysis in the Z Depleted Signal Region

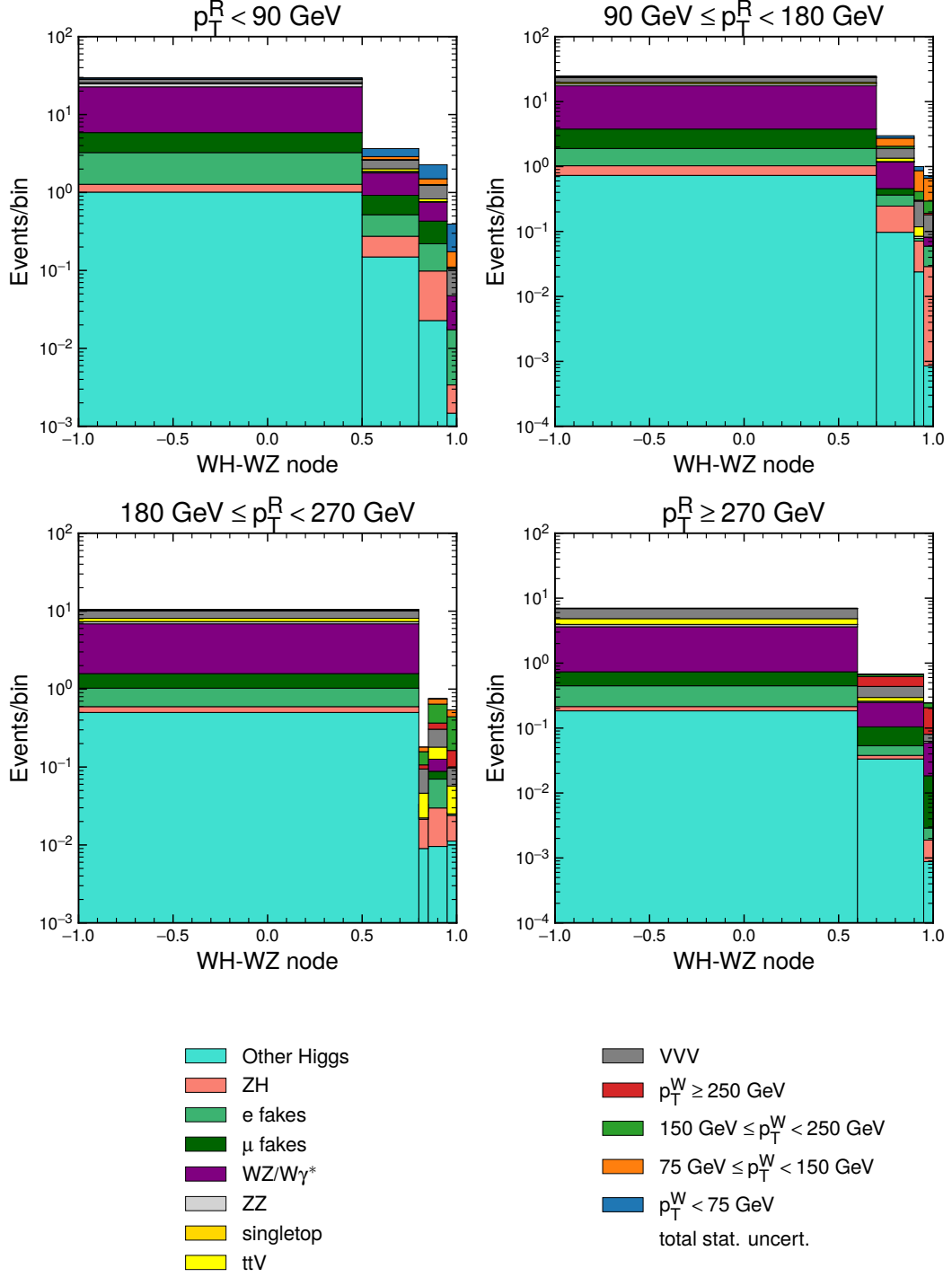


Figure 7.5.: Binning of the old multiclassifier output in the  $p_T^R$  bins.

## 7.6. Fits With Backgrounds

To quantify the performance of the multiclassifier in STXS measurements, Asimov fits including the background processes were performed. The four STXS bins were given a signal strength  $\mu_i$ , while the background is not scaled. The likelihood is

$$L(\mu) = \prod_j \frac{\left(\sum_{i=1}^4 \mu_i s_{i,j} + b_j\right)^{n_j}}{n_j!} e^{-(\sum_{i=1}^4 \mu_i s_{i,j} + b_j)}, \quad (7.1)$$

where  $s_{i,j}$  is the signal prediction for the  $i$ -th STXS category in the  $j$ -th bin of the classifier output. The fit results in an uncertainty for each  $\mu$  which can be used to judge the performance of the classifier.

While the uncertainties are a useful figure of merit for the multiclassifier, they are somewhat flawed. The `pyhf` software [55] which was used for these fits calculates them based on Equation 4.4. However, this approximation assumes large sample sizes, which is not the case here, especially in the higher  $p_T^R$  bins. The uncertainties are symmetrical and larger than 1, which in the downward direction is unphysical, since  $\mu$  cannot be less than 0.

For this reason, upper limits were calculated as an additional figure of merit. They are motivated by the use of the STXS analysis as a probe for BSM physics and are meant to indicate the sensitivity of the analysis to a possible signal excess caused by BSM effects. For this purpose, the signal strengths in the likelihood were replaced by

$$\mu_i \rightarrow 1 + \Delta\mu_i, \quad (7.2)$$

where  $\Delta\mu_i$  is the excess of the signal over the standard model expectation. This method does not account for any shape differences in the classifier output between the SM and BSM contributions, instead the SM prediction is simply scaled. It is intended only as a simple approximation of a possible BSM sensitivity, as no MC samples with predictions of specific BSM theories were available.

The limits are upper limits on  $\Delta\mu$  with a confidence level of 95%, using  $CL_S$ . The hypothesis tests were performed using toys instead of the asymptotic approximation. Such a limit was calculated for each STXS category, treating the other three parameters as nuisance parameters. The results of the fit and upper limit calculations using the distributions in Fig. 7.5 are shown in Table 7.5.

## 7. Analysis in the $Z$ Depleted Signal Region

Table 7.5.: Fit results for the old multiclassifier with the binning shown in Fig. 7.5.

|                                                         | Uncertainty | Upper limit |
|---------------------------------------------------------|-------------|-------------|
| $0 \text{ GeV} \leq p_{\text{T}}^W < 75 \text{ GeV}$    | 1.236       | 2.834       |
| $75 \text{ GeV} \leq p_{\text{T}}^W < 150 \text{ GeV}$  | 1.476       | 3.156       |
| $150 \text{ GeV} \leq p_{\text{T}}^W < 250 \text{ GeV}$ | 1.913       | 4.778       |
| $p_{\text{T}}^W \geq 250 \text{ GeV}$                   | 2.072       | 6.919       |

### 7.7. Suppression of the Remaining $WZ$ Background

While the 0 SFOS criterion greatly reduces the contribution of the  $WZ$  background, it still contributes to the  $Z$  depleted signal region via  $Z \rightarrow \tau\tau$  decays and charge misidentification, as described in Section 5.4.1. To study the contributions of these two processes to the  $WZ$  background, the reconstructed charges of the leptons in a Monte Carlo event were compared to the truth values. It was found that out of 47.42 expected  $WZ$  events 18.31 featured a charge misidentification, making up 38.6 % of the background. As shown in Fig. 7.6, charge misidentification events become dominant in the high  $p_{\text{T}}^R$  region. Since this region is expected to be sensitive to BSM signals, it is desirable to further suppress the charge flip background.

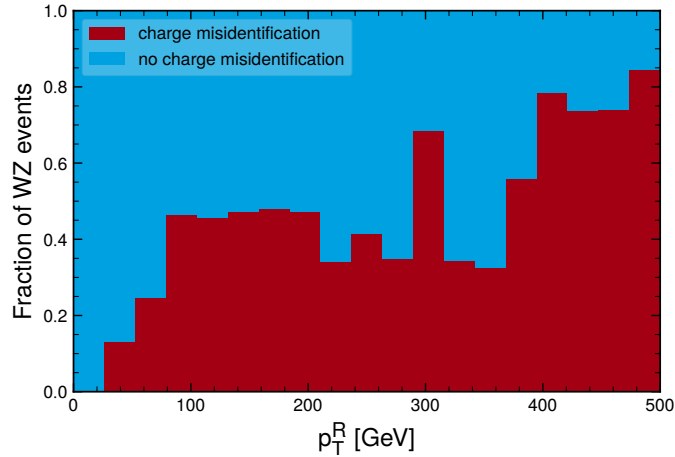


Figure 7.6.: Relative contributions to the  $WZ$  background by events with a charge misidentification and events without one, assumed to be  $Z \rightarrow \tau\tau$  decays.

For this purpose, the Electron Charge ID Selection tool (ECIDS) [27] was tested. It is a boosted decision tree which uses properties of the electron tracks to distinguish between electrons with correctly and incorrectly identified charge. It was applied by requiring all electrons to pass the ECIDS selection. The result of this is shown in Fig. 7.7: applying ECIDS significantly reduces the contribution of events with charge misidentification, while signal events and  $WZ$  events without charge misidentification

### 7. Analysis in the $Z$ Depleted Signal Region

are largely unaffected. Notably, the distribution of  $WZ$  events with charge misidentification peaks at a lower value of the  $WH - WZ$  node than that without charge misidentification. This indicates that the neural network is also capable of recognizing and suppressing events with a charge misidentification because it receives the impact parameters and  $\max(|1 - E^{l_i}/p^{l_i}|)$  as inputs, which are sensitive to charge misidentification. The effect of the ECIDS requirement on the signal and  $WZ$  yields is shown in Table 7.6.

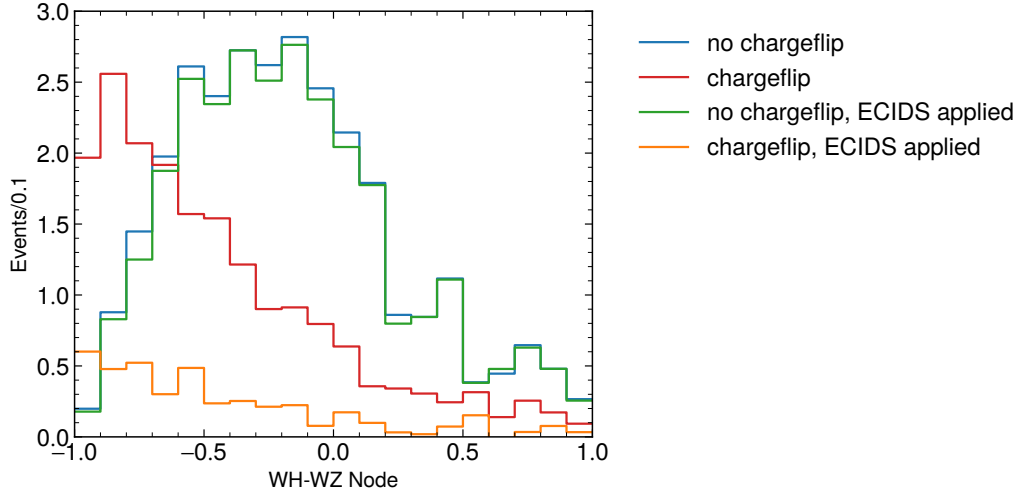


Figure 7.7.: Distribution of  $WZ$  events with and without charge flips in the  $WH - WZ$  output node, before and after applying ECIDS.

Table 7.6.: Total yields for the  $WH$  signal and the  $WZ$  background in the  $Z$  depleted signal region, before and after applying ECIDS.

|                       | $WH$  | $WZ$  |                               |                                  |
|-----------------------|-------|-------|-------------------------------|----------------------------------|
|                       |       | Total | With charge misidentification | Without charge misidentification |
| Before applying ECIDS | 14.44 | 47.42 | 18.31                         | 29.11                            |
| After applying ECIDS  | 14.10 | 32.21 | 4.04                          | 28.17                            |

## 7.8. Retraining of the Multiclassifier

### 7.8.1. Motivation

The multiclassifier for signal-background separation was retrained for the following reasons. Because of the symmetry of the physics processes and the approximate symmetry of the detector in  $\phi$  and because the  $\phi$  of the leptons is given relative to the missing transverse energy, the variable  $\phi_{\text{MET}}$  has no discriminating power and should therefore be removed from the input variables. Furthermore, it is desirable to remove the variables  $\max(|1 - E^{l_i}/p^{l_i}|)$ ,  $d_0/\sigma(d_0)_{l_i}$  and  $z_0 \sin \theta_{l_i}$ . They are poorly modeled in the MC simulations. Comparisons between data and Monte Carlo simulation for the variables are shown in Fig. 7.8, Fig. 7.9 and Fig. 7.10. Furthermore, their use in the multiclassifier conflicts with the lepton fake estimation (see Section 5.4.2). Because these variables are used in the lepton isolation algorithms, they differ between events with an anti-ID lepton and events with only ID leptons, and thus might cause a difference in the multiclassifier output between these types of events. If there is such a shape difference, one cannot extrapolate the fake lepton contribution by simply scaling with the fake factor. For the purposes of the retraining, they are collectively referred to as impact parameters. The multiclassifier had originally been optimized to maximize its power [48], but the later adoption of the fake factor method and discovery of the mismodelling necessitated the retraining.

As shown in Section 7.7, applying ECIDS helps to suppress  $WZ$  events with a charge misidentification. It is unclear whether ECIDS should also be applied to the training samples for the multiclassifier: removing events with a charge misidentification might cause the network to focus on events with  $Z \rightarrow \tau\tau$ , instead of also learning to discriminate against events which will be removed by ECIDS anyway. However, this possible benefit might be outweighed by the reduction of the training sample size leading to a worse performance of the network against  $WZ$  overall, since there are fewer examples for the network to learn the characteristics of the  $WZ$  process that are not specific to  $Z \rightarrow \tau\tau$ . Since it is unclear a priori which of these two options is better, both were tested.

After the old classifier had been trained, a larger MC sample of the triboson process became available. Including this sample in the training of the multiclassifier is desirable for two reasons. Firstly, it was found that in the most signal-like region of the old multiclassifier, the triboson process is a dominant background, similar in magnitude to the  $WZ$  process. Including triboson events might help to better discriminate against this background. Secondly, a recent measurement of the process  $pp \rightarrow WWW$ , which includes  $pp \rightarrow WH \rightarrow WWW$ , measured an excess of events, with a signal strength  $\mu_{WWW} = 1.66 \pm 0.28$  [56]. It will be of interest to compare this result to that of the  $WH \rightarrow WWW$  analysis. For this purpose it is useful to define a region of phase space that is similar to the signal region, but that contains mostly triboson events and very few signal events. Such a region could be defined by training a multiclassifier with a separate output node for this process and selecting events for which this node has

## 7. Analysis in the $Z$ Depleted Signal Region

high values. The triboson sample can either be combined with the  $WZ$  sample, which might help discriminate against  $WWW$  but not lead to an additional output node, or be used separately, which does lead to a new output.

Since the old multiclassifier was optimized for the inclusive analysis, it needs to be investigated if its performance can be improved upon for the STXS measurement. One possible approach to this is a reweighting of the training samples, similar to the reweighting done for the training of the regression network. Since no information on the true  $p_T^W$  is available for background processes, especially of course for  $t\bar{t}$  which does not involve a  $W$  boson at all, the reweighting is done in the  $p_T^R$  bins previously defined instead of the STXS bins. This is done in addition to the reweighting of the classes for the training. After the reweighting all classes have the same sum of weights, and so do all  $p_T^R$  bins.

A second method of possibly improving the performance of the classifier for the STXS measurement is adding  $p_T^R$  to the input variables. This might enable the ANN to become more sensitive to the kinematic properties of the processes in different  $p_T^R$  regimes. The input variables of the regression network are a subset of those of the multiclassifier, so the multiclassifier could in principle implicitly reconstruct  $p_T^W$  on its own. However, passing  $p_T^R$  explicitly could remove the need for this implicit reconstruction and “free up” parts of the ANN, thus improving the overall performance. Similarly, variables such as invariant masses were provided to the network, even though they could be reconstructed from the lepton momenta.

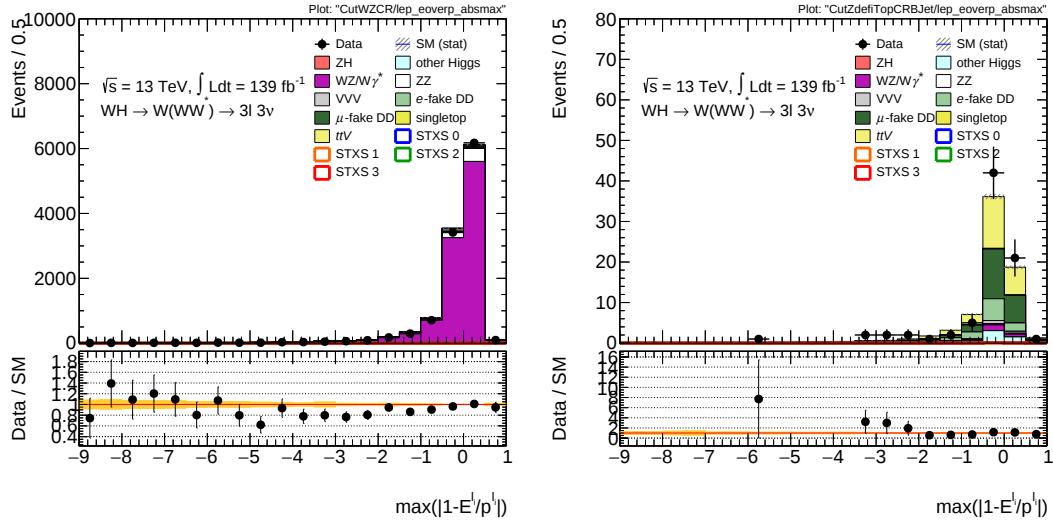


Figure 7.8.: Data-MC comparison for  $\max(|1 - E^l/p^l|)$  in the  $WZ$  CR and Top VR.

## 7. Analysis in the Z Depleted Signal Region

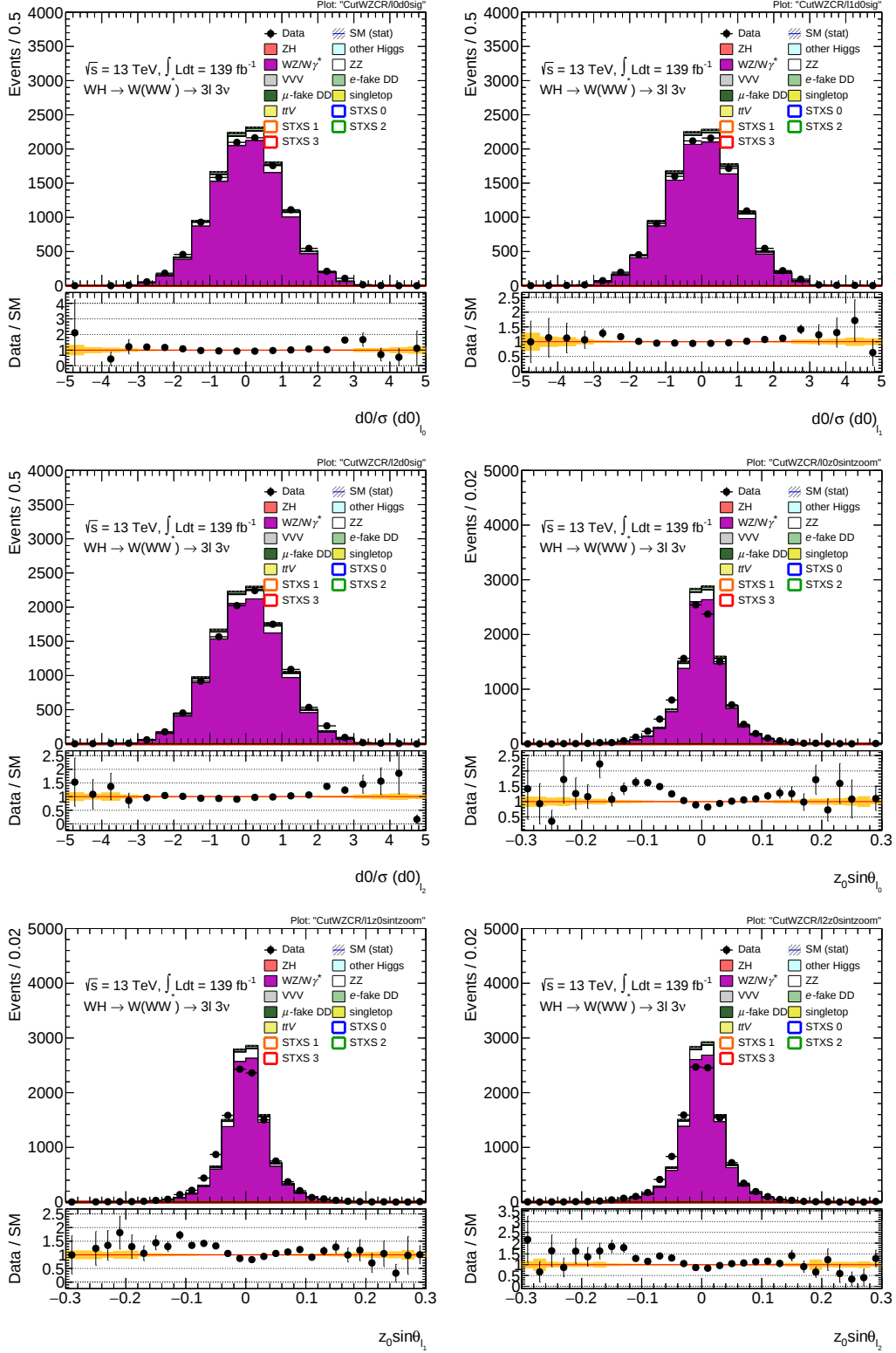


Figure 7.9.: Data-MC comparisons for the impact parameters in the WZ CR.

## 7. Analysis in the Z Depleted Signal Region

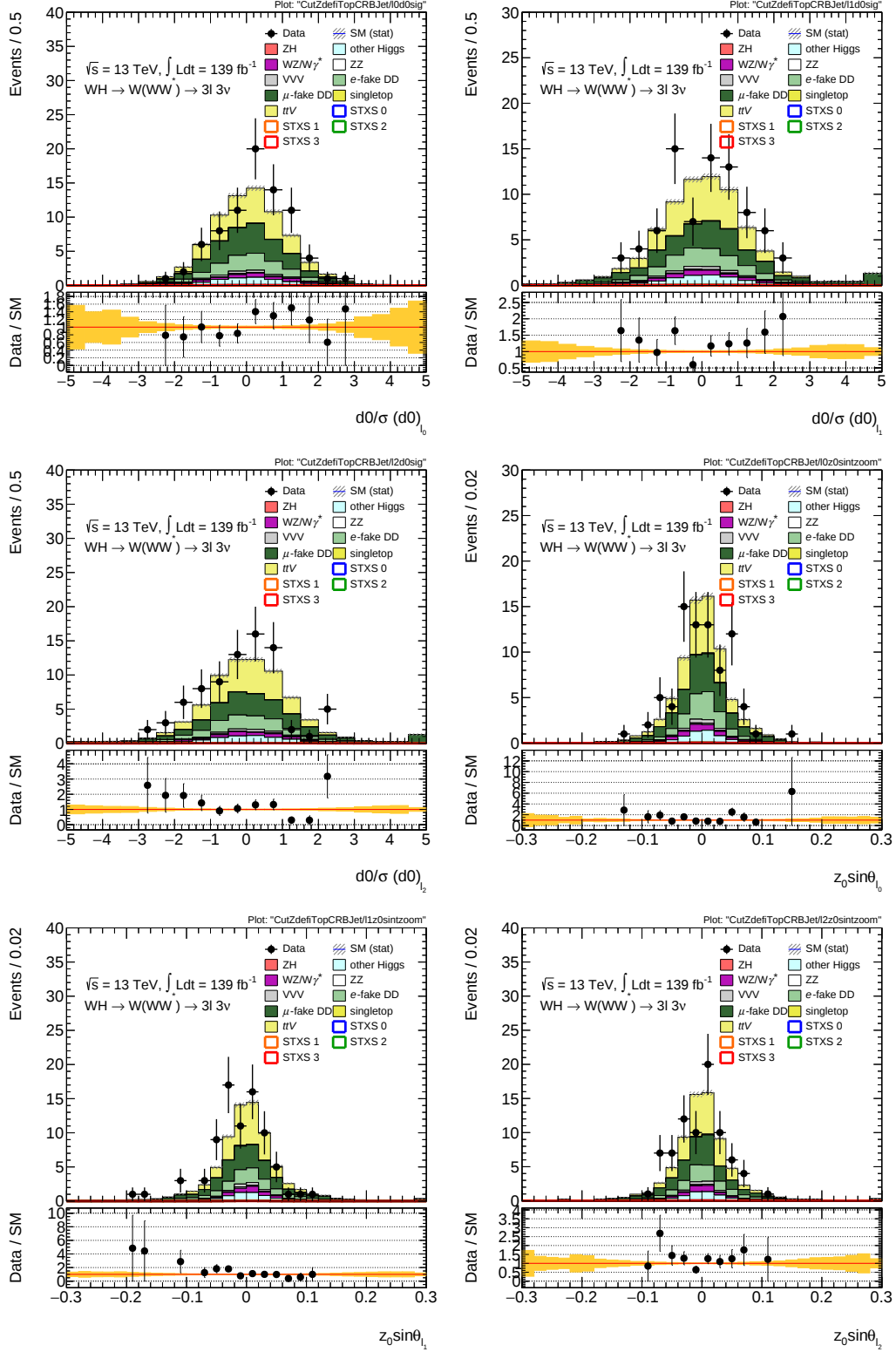


Figure 7.10.: Data-MC comparisons for the impact parameters in the Top VR.

## 7. Analysis in the $Z$ Depleted Signal Region

### 7.8.2. Retraining Setup

The training setup for the retraining of the classifier was mostly identical to the original training. The same architecture was used, except for the changes in the number of input variables and the possible addition of a  $WWW$  output node. The hyperparameters of the training were not changed. The same training samples were used as for the old multiclassifier, with the addition of the triboson sample. Events from the  $Z$  dominated region were used for the  $WWW$  training.

Table 7.7 lists the options for the different aspects of the retraining, as previously discussed. All combinations of these options were tested, leading to a total of 48 variants.  $\phi_{\text{MET}}$  was always removed from the input variables. For every variant a network with 5  $k$ -folds was trained.

Table 7.7.: Summary of changes to the classifier tested during the retraining.

| $WWW$ sample            | ECIDS                       | Impact parameters | $p_{\text{T}}^R$ | Weights                         |
|-------------------------|-----------------------------|-------------------|------------------|---------------------------------|
| Not included            | Not applied before training | Not included      | Not included     | No $p_{\text{T}}^R$ reweighting |
| Merged with $WZ$ sample | Applied to training sample  | Included          | Included         | Reweight in $p_{\text{T}}^R$    |
| Included separately     |                             |                   |                  |                                 |

### 7.8.3. Comparison of the Training Variants

To compare the different retrained ANNs, they were used as discriminants in likelihood fits. Their output was treated as described in Section 7.5: the  $t\bar{t}$  node was used to define a cut, while the other nodes were combined into a single discriminant. In the case of networks with an additional  $WWW$  node, this new node was defined by subtracting both the  $WZ$  and the  $WWW$  node from the  $WH$  node, creating the  $WH - WZ - WWW$  node. Like the  $WH - WZ$  node, it takes values between -1 and 1. As before, the events were first divided into  $p_{\text{T}}^R$  bins and then binned in the  $WH - WZ(-WWW)$  node within these.

Since it was not feasible to develop a custom binning of the multiclassifier output for each of the over 200 individual networks, the same relatively fine binning was used for all ANNs. From the binning studies for the old multiclassifier it was known that the signal contributions are mostly confined to values of the  $WH - WZ$  node above 0.5, the bin borders were thus chosen to be  $(-1, 0.5, 0.6, 0.7, 0.8, 0.9, 1)$ . ECIDS was applied to the samples used for the fit. A fit as described in Section 7.6 was performed for every network, resulting in four uncertainties and four upper limits each. These figures of merit were averaged over the  $k$ -folds for every training variant.

## 7. Analysis in the Z Depleted Signal Region

The training variants with and without impact parameters are compared visually in Fig. 7.11. An overall trend is visible in which variants with impact parameters, shown in blue, tend to outperform variants without, shown in orange. However, some networks without impact parameters do achieve performances similar to those without. Considering the reasons for removing these variables discussed above, it was therefore decided to use a classifier without impact parameters.

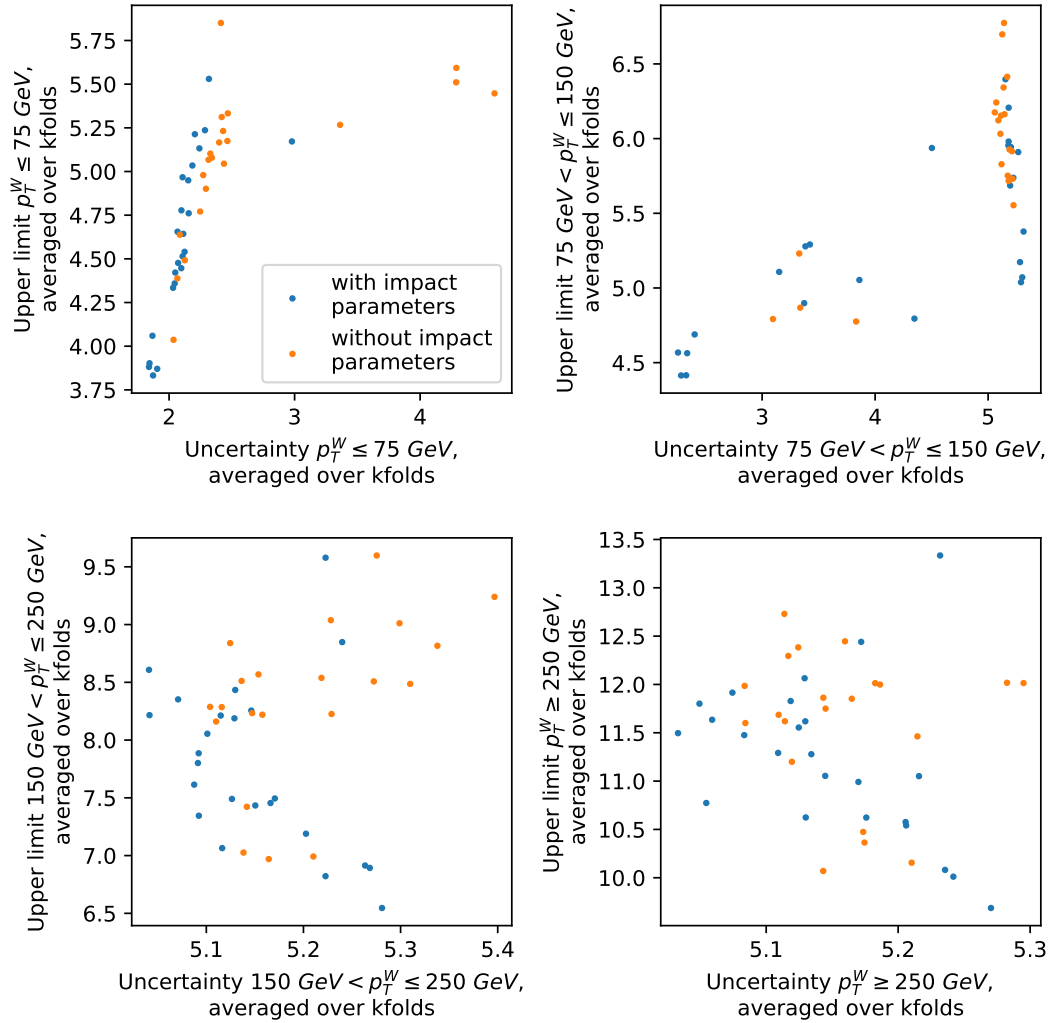


Figure 7.11.: Uncertainties and upper limits for training variants with and without impact parameters.

## 7. Analysis in the $Z$ Depleted Signal Region

Fig. 7.12 shows the variants without impact parameters, color coded by whether the  $WWW$  sample was merged with the  $WZ$  sample, included separately, or left out. In the best performing variants the sample was merged. However, in light of excess in the recent  $WWW$  measurement [56] it was decided to use a network with a separate  $WWW$  node because of its possible use in defining a region pure in the  $WWW$  process. The loss in performance incurred when choosing such a network was found not to be too large, especially in the more important high STXS bins.

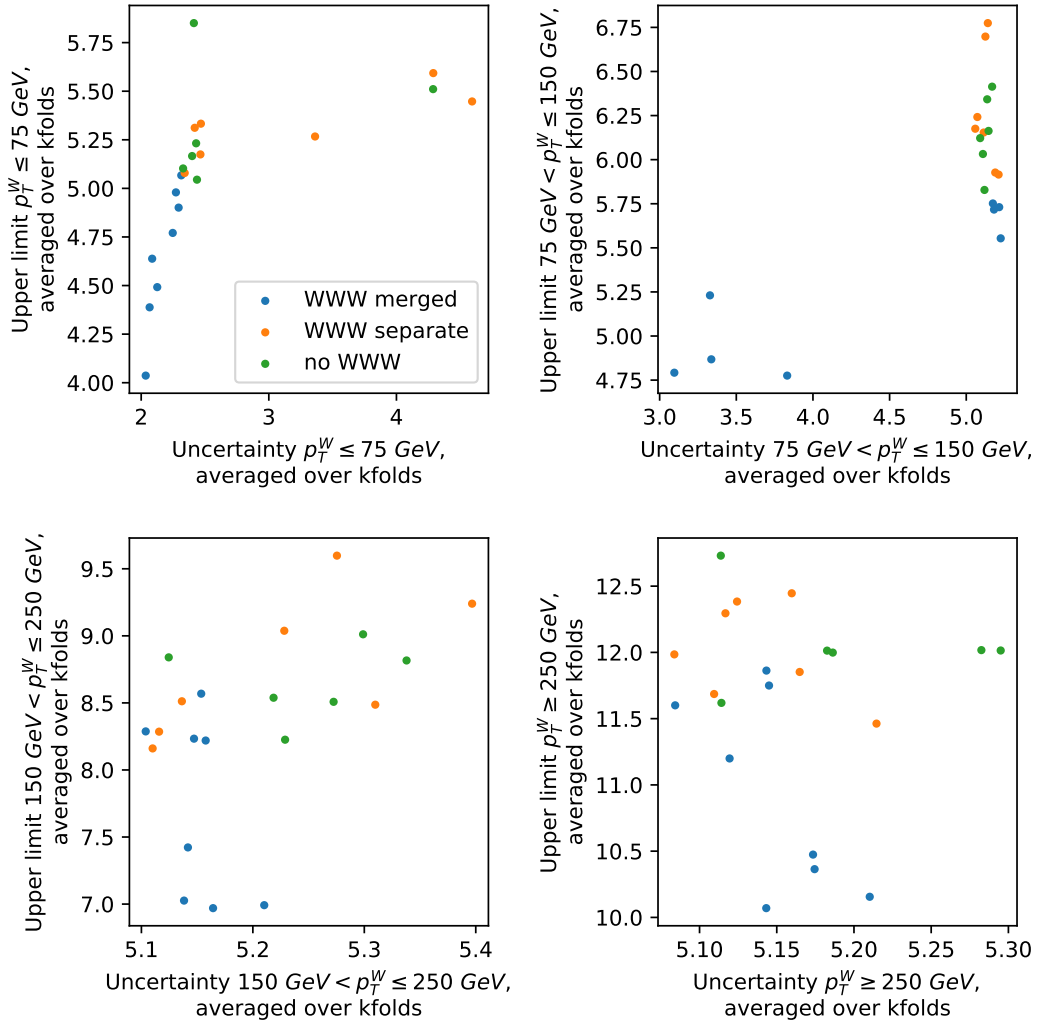


Figure 7.12.: Uncertainties and upper limits for training variants with different treatments of the  $WWW$  sample. Only variants without impact parameters are shown.

## 7. Analysis in the $Z$ Depleted Signal Region

The remaining variants after enforcing a separate  $WWW$  node and the removal of the impact parameters are shown in Fig. 7.13. None of these variants clearly outperforms the others. Ultimately, the variant was chosen which is marked with arrows in the figure. It is among the variants with small values for all figures of merit except for the uncertainty of the highest STXS bins, where the spread of the variants is small at about 0.13 between the highest and lowest value. Furthermore, this variant was also found to perform well in a separate study for the inclusive analysis, in which fits were performed to determine the significance of the SM hypothesis versus the background only hypothesis using the different networks [29].

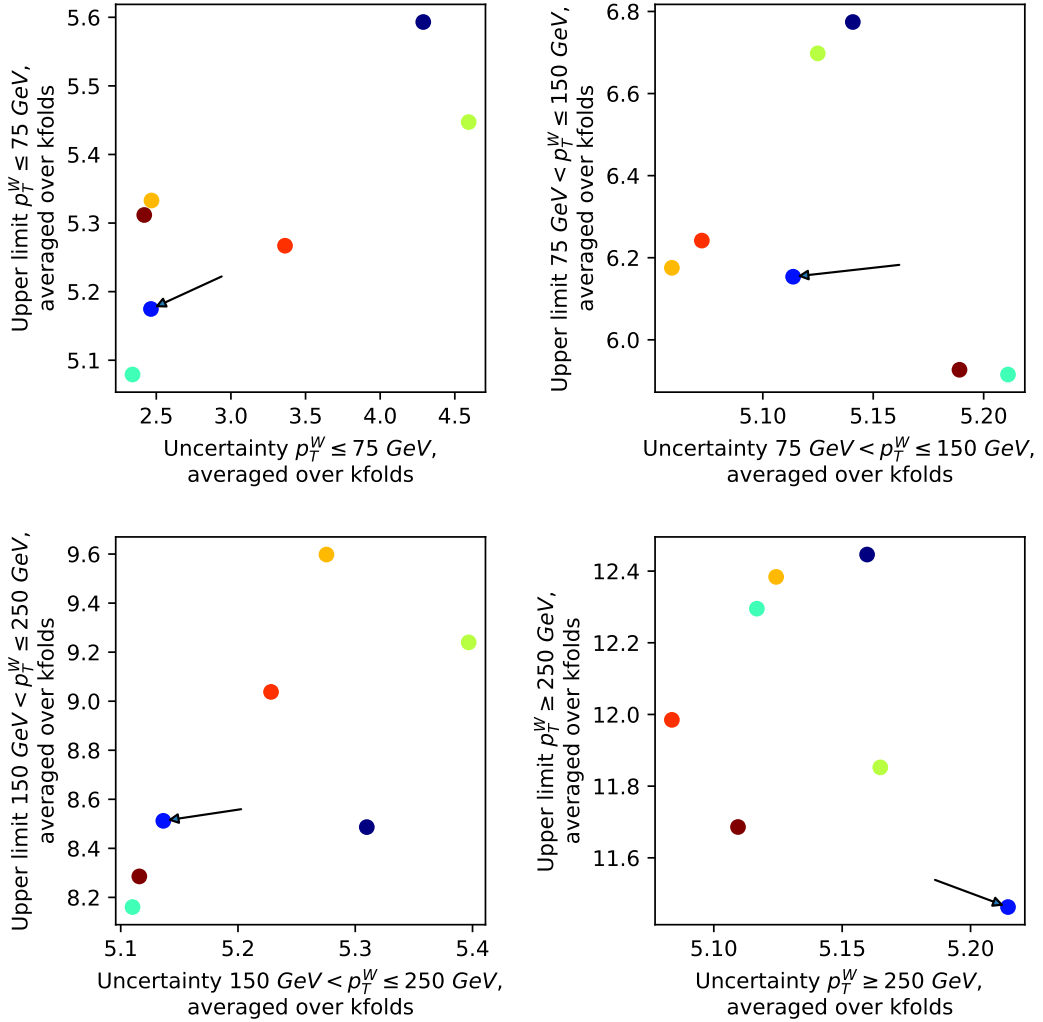


Figure 7.13.: Uncertainties and upper limits for training variants with a separate  $WWW$  node and no impact parameters. The chosen variant is marked with arrows.

## 7. Analysis in the $Z$ Depleted Signal Region

Table 7.8.: Training options used for the final retrained multiclassifier.

| WWW sample           | ECIDS                      | Impact parameters | $p_T^R$      | Weights                |
|----------------------|----------------------------|-------------------|--------------|------------------------|
| Separate output node | Applied to training sample | Not included      | Not included | No $p_T^R$ reweighting |

Table 7.9.: Input variables of the retrained multiclassifier.

| $p_T^{l_0}$                     | $p_T^{l_1}$                     | $p_T^{l_2}$                     | $E_T^{\text{miss}}$            |
|---------------------------------|---------------------------------|---------------------------------|--------------------------------|
| $\Delta R_{l_0 l_1}$            | $\Delta R_{l_0 l_2}$            | $\Delta R_{l_1 l_2}$            | $\sum_i p_T^{l_i}$             |
| $\Delta \eta_{l_0 l_1}$         | $\Delta \eta_{l_0 l_2}$         | $\Delta \eta_{l_1 l_2}$         | $S/E_T^{\text{miss}}$          |
| $\Delta \phi_{l_0, \text{MET}}$ | $\Delta \phi_{l_1, \text{MET}}$ | $\Delta \phi_{l_2, \text{MET}}$ | $p_T^{\text{max}}(\text{jet})$ |
| $m_T^{l_0 l_1}$                 | $m_T^{l_0 l_2}$                 | $m_T^{l_1 l_2}$                 | $F_{\alpha_1}$                 |
| $m_{l_0 l_1}$                   | $m_{l_0 l_2}$                   | $m_{l_1 l_2}$                   | $m_{lll}$                      |
|                                 | $n_{\text{jets}}$               | $n_{\text{b-jets}}$             |                                |

The chosen variant uses the training options shown in Table 7.8 and the inputs shown in Table 7.9. The best performing  $k$ -fold of this training variant was chosen to be the new multiclassifier for the  $Z$  depleted analysis, which will be referred to as the retrained classifier from here on out.

As a point of comparison, a fit with the same setup as for the comparison between the training variants was performed for the old multiclassifier. The results are compared with those of the retrained classifier in Table 7.10. The retrained classifier appears to outperform the old one. This comparison is biased, however, since the new classifier was selected specifically based on these fits, the binning that was used might simply be disadvantageous for the old multiclassifier. An unbiased comparison can only be done using binnings specifically developed for the respective classifiers. This also applies to the comparison of the training variants with each other. It is possible that with a different binning a different training variant would have been chosen, which might ultimately perform better than the retrained classifier. Ideally, one would first optimize a binning for each ANN, but this was not feasible within the scope of this thesis.

Table 7.10.: Comparison between the old and retrained multiclassifier with the binning used for the training variant comparison.

|               |                                                | Old classifier | Retrained classifier |
|---------------|------------------------------------------------|----------------|----------------------|
| Uncertainties | $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 1.448          | 1.236                |
|               | $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 1.696          | 1.502                |
|               | $150 \text{ GeV} \leq p_T^W < 250 \text{ GeV}$ | 2.245          | 1.955                |
|               | $p_T^W \geq 250 \text{ GeV}$                   | 5.281          | 2.166                |
| Limits        | $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 3.249          | 2.642                |
|               | $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 3.310          | 2.758                |
|               | $150 \text{ GeV} \leq p_T^W < 250 \text{ GeV}$ | 5.654          | 4.190                |
|               | $p_T^W \geq 250 \text{ GeV}$                   | 9.939          | 6.205                |

### 7.9. $WWW$ Validation Region

The dedicated  $WWW$  output node of the retrained multiclassifier network is used to define a  $WWW$  validation region. This region shares all requirements of the  $Z$  depleted SR, i.e. 0 SFOS pairs, the ECIDS requirement, the  $b$  veto and a  $t\bar{t}$  node value below 0.3. In addition to this, it is required that the  $WWW$  node is greater than 0.5. This ensures that this region is free of signal events, as well as most events from other backgrounds.

Fig. 7.14 shows the distribution of the  $WWW$  node in this region. This region is clearly enriched in  $WWW$  events. There appears to be an excess of data events in the center bin. Because of the difference in shape between the data and the  $WWW$  prediction, this does not appear to be an excess in  $WWW$  events. Possible sources of this discrepancy, such as background processes that were not accounted for, will need to be investigated, but this is outside the scope of this thesis.

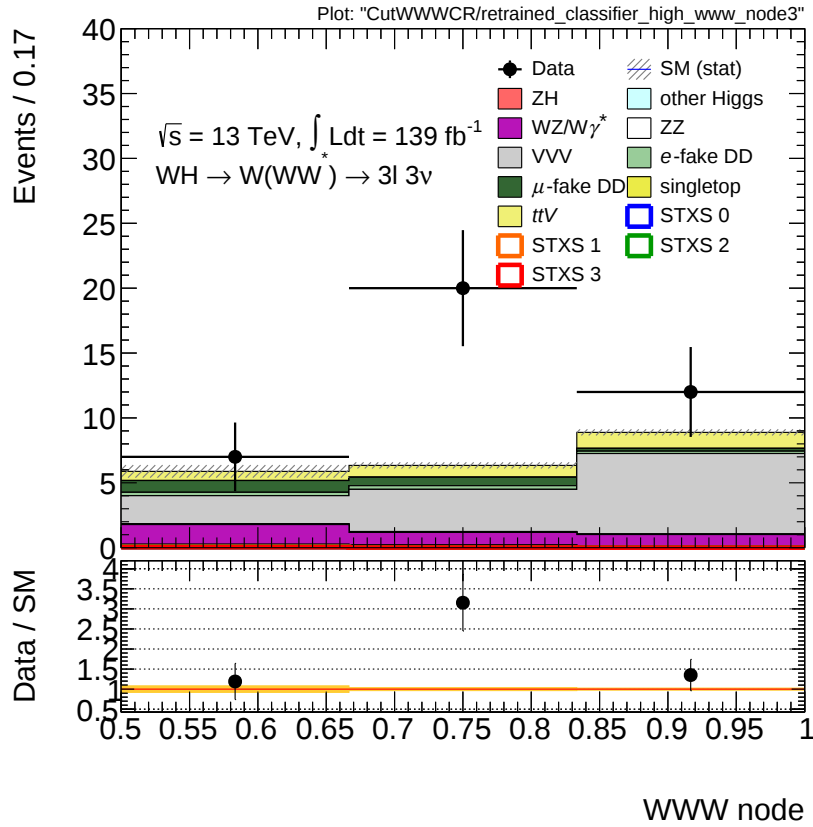


Figure 7.14.: Distribution of the  $WWW$  output node in the  $WWW$  validation region.

## 7.10. $WZ$ Scale Uncertainties

The uncertainties of the previous inclusive  $WH \rightarrow WWW \rightarrow l\nu l\nu l\nu$  analysis [43] were dominated by the statistical uncertainties of the data. This is still expected to be the case in the full Run 2 analysis, especially for the STXS measurement. To estimate the impact of systematic uncertainties on the STXS measurement, a subset of the systematic uncertainties of the prediction for the  $WZ$  background process were studied. It was the most important background process in terms of uncertainties in the previous analysis, followed by lepton fake related backgrounds, which after the introduction of the PLIV isolation working point (see Section 3.5.4) are rejected better compared to the previous analysis and thus less important.

The systematical uncertainties that were investigated were the choice of the renormalization and factorization scales in the MC simulation of the  $WZ$  background. To study their impact on the shape of the predicted  $WZ$  contribution, samples of the process  $pp \rightarrow WZ$  were generated using MADGRAPH5\_aMC@NLO [50], with up to 2 jets in the final state. The different scale choices are the scalar sum of transverse momenta  $\sum p_T$ , the scalar sum of jet transverse momenta  $H_T$ ,  $H_T/2$ , the center of mass energy  $\sqrt{s}$  and the central  $m_T^2$  scale  $\sqrt{(m_W^2 + (p_T^W)^2)(m_Z^2 + (p_T^Z)^2)}$ . The scale is then multiplied with the scale factors  $\mu_R$  and  $\mu_F$  to calculate the renormalization and factorization scale, respectively. They can take values of 0.5, 1 or 2, leading to a total of 45 variations. The variation with the central  $m_T^2$  scale and  $\mu_R = \mu_F = 1$  is taken to be the nominal distribution.

The distributions of the different variations were normalized, so that only their shapes differ, not the total yield. The overall yield of the  $WZ$  background is a separate systematic uncertainty and is constrained by the  $WZ$  control region in the full statistical treatment [29], but is not considered here. The  $p_T^W$  spectra of the different variations are shown in Fig. 7.15. The height of the peak of the distribution does not differ much between variations. Instead, the difference is mostly in the positions of the peak. Most variations are higher than the nominal distribution at low  $p_T^W$  and lower at high  $p_T^W$  or vice versa, corresponding to a shift toward lower  $p_T^W$  or higher  $p_T^W$ , respectively. The variations with the largest difference to the nominal distribution are the one with the central  $m_T^2$  scale and  $\mu_R = \mu_F = 0.5$  and the one with  $\sqrt{s}$  and  $\mu_R = \mu_F = 2$ .

The variations in the STXS bins are shown in Fig. 7.16. For every STXS bin, a relative uncertainty on the  $WZ$  yield is extracted from the variations. This is done by simply taking the difference between the highest and lowest variation in each bin and dividing it by the nominal value. The values of the relative uncertainties are listed in Table 7.11. They are largest in the highest STXS bin.

In the fit, the relative uncertainties are assigned based on the  $p_T^R$  bins corresponding to the STXS bins, that is all multiclassifier bins within the lowest  $p_T^R$  bin are assigned the relative uncertainty on the  $WZ$  contribution calculated in the lowest STXS bin and so forth. In every multiclassifier bin, the  $WZ$  contribution is scaled by a factor

## 7. Analysis in the $Z$ Depleted Signal Region

corresponding to the uncertainty that is independent of the other bins and constrained by a Poissonian term. The uncertainties extracted from fits with and without the  $WZ$  scale systematics are compared in Table 7.12. The increase in the uncertainties after including the systematics is tiny. The STXS analysis is clearly dominated by statistical uncertainties. The  $WZ$  scale systematics were included in all fits and limit calculations using the retrained classifier (see Section 7.13). Because of the time constraints of the thesis and because systematic uncertainties were found to be negligible, no other systematical uncertainties were included.

Table 7.11.: Relative systematic uncertainties of the  $WZ$  yield caused by scale variations.

| STXS bin                                                | Uncertainty |
|---------------------------------------------------------|-------------|
| $0 \text{ GeV} \leq p_{\text{T}}^W < 75 \text{ GeV}$    | 0.0475      |
| $75 \text{ GeV} \leq p_{\text{T}}^W < 150 \text{ GeV}$  | 0.0517      |
| $150 \text{ GeV} \leq p_{\text{T}}^W < 250 \text{ GeV}$ | 0.1497      |
| $p_{\text{T}}^W \geq 250 \text{ GeV}$                   | 0.2523      |

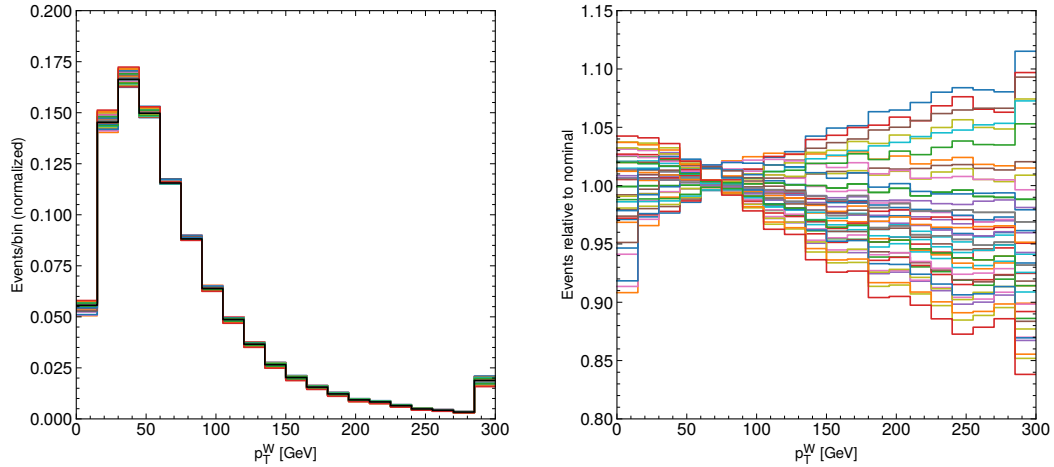


Figure 7.15.: Left: normalized  $p_{\text{T}}^W$  distributions of the  $WZ$  background with different scale variations. The nominal distribution is shown in black. Right: scale variations relative to the nominal distribution.

## 7. Analysis in the $Z$ Depleted Signal Region

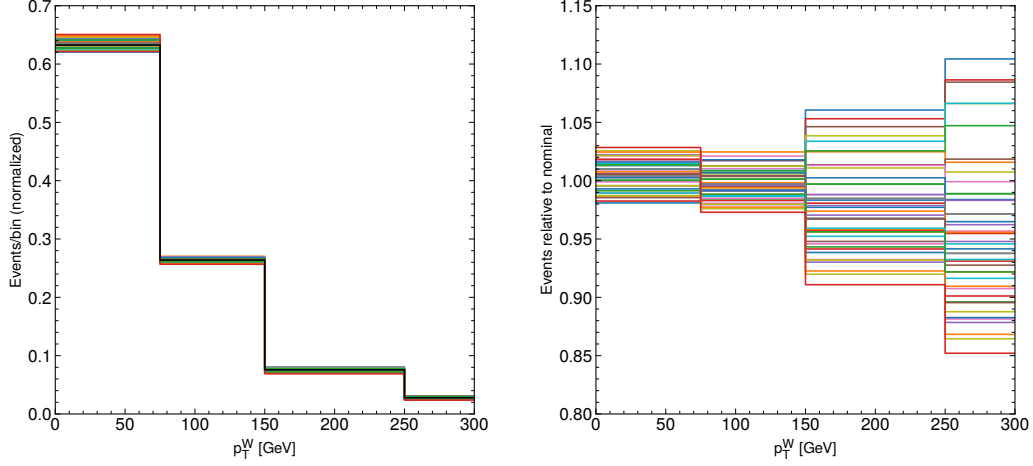


Figure 7.16.:  $WZ$  scale variations in the STXS bins.

Table 7.12.: Uncertainties on the signal strengths in the four STXS bins with and without  $WZ$  scale systematics included in the fit.

|                                                | Statistical uncertainties only | With $WZ$ scale systematics |
|------------------------------------------------|--------------------------------|-----------------------------|
| $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 1.240                          | 1.240                       |
| $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 1.518                          | 1.519                       |
| $150 \text{ GeV} \leq p_T^W < 250 \text{ GeV}$ | 1.981                          | 1.983                       |
| $150 \text{ GeV} \leq p_T^W < 250 \text{ GeV}$ | 2.215                          | 2.220                       |

### 7.11. Distribution of a Possible BSM Signal

To better quantify the sensitivity of the analysis to possible BSM signals beyond the upper limits on the STXS signal strengths described in Section 7.6, the difference between the shape in  $p_T^W$  of the SM  $WH$  process and of an example of a BSM process was studied. This was done at generator level, using MADGRAPH5\_aMC@NLO [50].

Two samples of the  $pp \rightarrow WH \rightarrow WWW \rightarrow \nu\nu\nu\nu$  were generated. One of these used the Standard Model. The BSM sample was generated using the Higgs characterization model (see Section 2.6). The scale  $\Lambda$  was set to 1 TeV and  $\cos \alpha$  was set to 1.  $\kappa_{H\partial Z}$  and  $\kappa_{H\partial W}$  were set to 1 and all others parameters including  $\kappa_{SM}$  were set to 0.

Fig. 7.17 shows the  $p_T^W$  spectra of the SM and BSM processes. Both distributions are normalized to the same total yield. There is a clear shape difference between the processes. The BSM distribution is much wider, with contributions at high values of  $p_T^W$ , where only minuscule SM yields are expected. As such, most of the BSM contribution is found in the highest STXS bin.

## 7. Analysis in the Z Depleted Signal Region

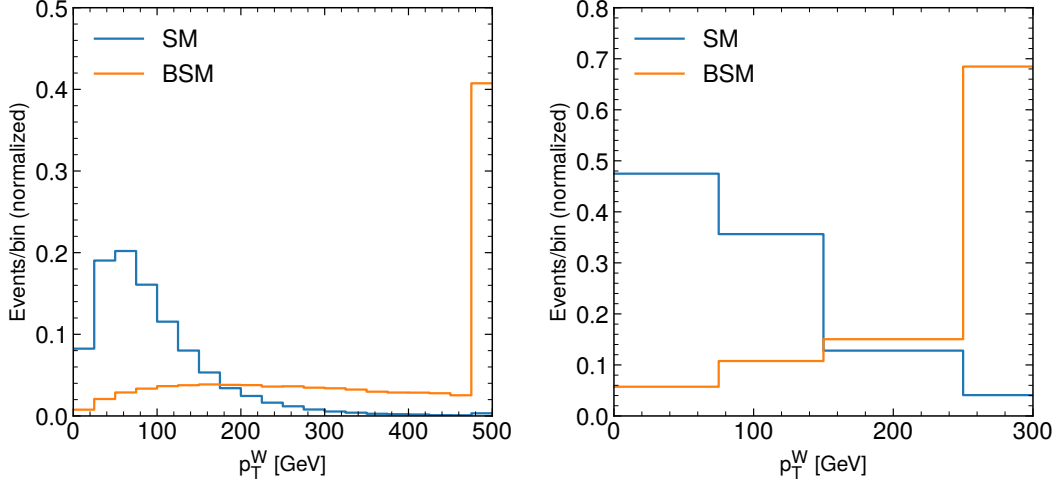


Figure 7.17.: Normalized  $p_T^W$  distributions of the SM and BSM samples in different binnings. Left: equal bins of 25 GeV width. Right: STXS bins.

In order to test the sensitivity of the analysis to this exemplary BSM signature, the generator-level distribution is translated to an approximate detector-level distribution. This is done by scaling the SM signal according to the shapes of the SM and BSM distributions. In each STXS bin, the ratio between the truth level BSM and SM contributions is calculated. Then, these factors are applied to the reconstruction level SM samples in the respective STXS bins. This method is illustrated in Fig. 7.18. The left figure shows the reconstruction level distributions of the STXS categories, as well as their sum, the total SM signal. On the right the STXS distributions are shown after being scaled by the factors extracted from the  $p_T^W$  distributions at truth level. Their sum is taken to be the BSM signal for the fit. This method works under the assumption that there are no differences in acceptance between the SM and BSM processes, i.e. that the ratios are unaffected by the cuts in the analysis. Furthermore, this is a simple approximation of a BSM effect in that it does not take into account interferences between the SM and BSM contributions to the  $pp \rightarrow WH \rightarrow WWW$  process.

An upper limit on the signal strength  $\mu_{BSM}$  of this BSM signal can be calculated to determine the sensitivity of the analysis to such a signal. Because the ratios between SM and BSM were calculated from normalized distributions, they only encode shape differences. The cross sections calculated by MADGRAPH are 0.095 pb for the SM process and 1.496 pb for the BSM process. The BSM distribution for the fit is scaled such that the ratio between the total yields of the SM and BSM signals is equal to that of these cross sections. A result of  $\mu_{BSM} = 1$  thus corresponds to a number of BSM events about 15.7 times higher than the SM expectation.

## 7. Analysis in the $Z$ Depleted Signal Region

The above method for estimating a BSM contribution was repeated while splitting the  $p_T^W$  spectra into only three bins, i.e. merging the two highest STXS bins. The resulting distribution is shown in Fig. 7.19. This alternative distribution can be used to compare the sensitivity achieved with four or only three bins.

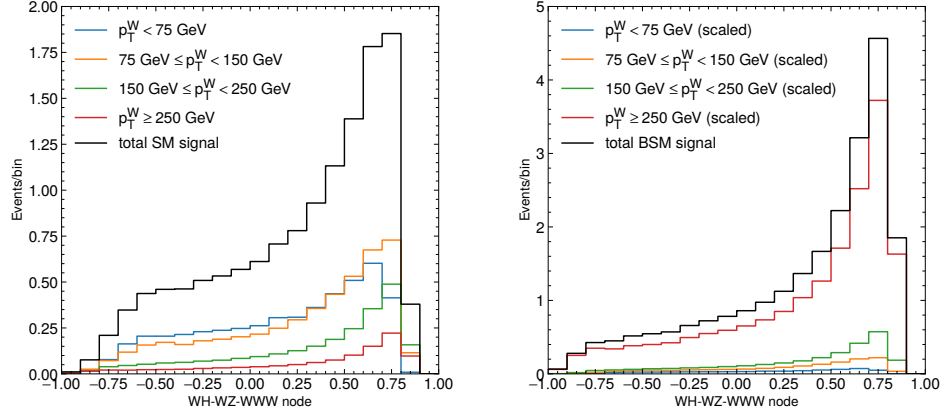


Figure 7.18.: SM and BSM signal distributions in the  $WH - WZ - WWW$  node. ECIDS,  $b$ -veto and  $t\bar{t}$  node cut were applied.

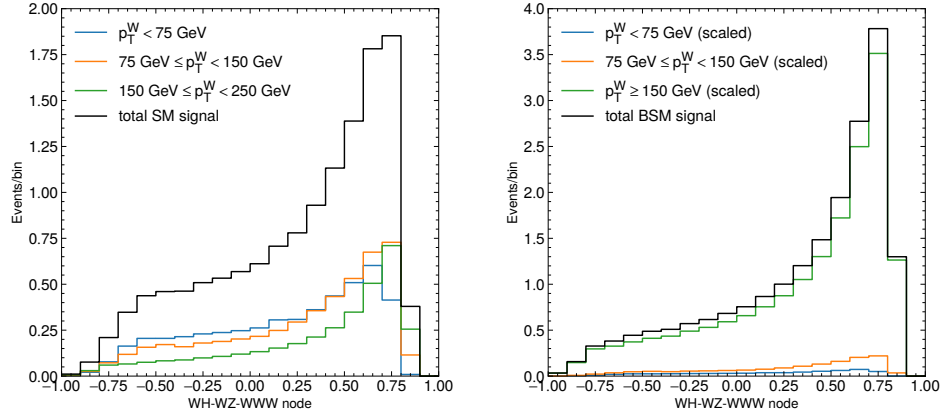


Figure 7.19.: SM and BSM signal distributions in the  $WH - WZ - WWW$  node, with only three STXS bins. ECIDS,  $b$ -veto and  $t\bar{t}$  node cut were applied.

## 7. Analysis in the $Z$ Depleted Signal Region

### 7.12. Binning of the Retrained Classifier

The output of the retrained classifier is shown in Fig. 7.20. It was treated the same way as that of the old classifier in Section 7.5. The events are first separated into  $p_T^R$  bins with borders at 90, 180 and 270 GeV. Within these bins, the events are binned in the  $WH - WZ - WWW$  node of the retrained multiclassifier. Fig. 7.21 shows the distributions of the multiclassifier output. Neighboring bins with similar  $S/B$  were merged. To maintain a high sensitivity of the fit, merging bins with very high  $S/B$  was avoided. The resulting binning is shown in Fig. 7.22. Some of the most sensitive bins have expected yields below 1. This is deemed acceptable as the statistical uncertainty of the MC prediction, drawn as a shaded region, is small compared to the prediction. It does, however, prevent the use of asymptotic methods when doing hypothesis tests. The yields in the  $p_T^R$  bins are shown in Table B.1.

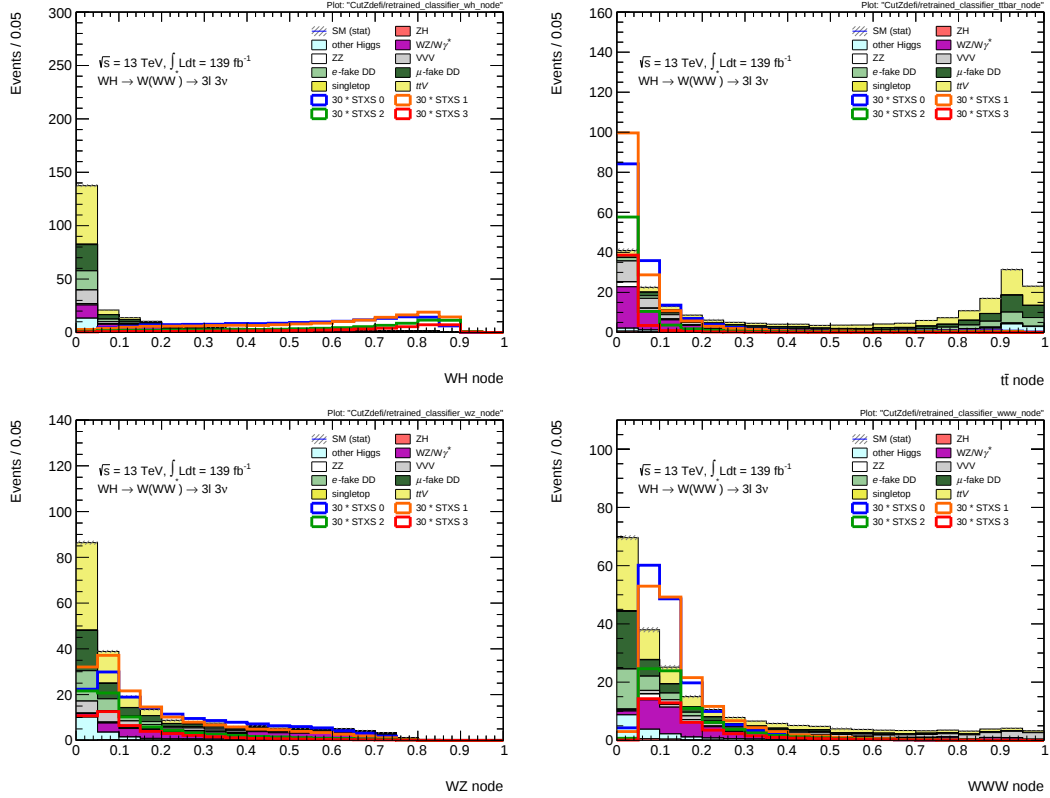


Figure 7.20.: Output distributions of the retrained multiclassifier network. The 0 SFOS requirement is applied but the  $b$  veto and ECIDS are not.

### 7. Analysis in the Z Depleted Signal Region

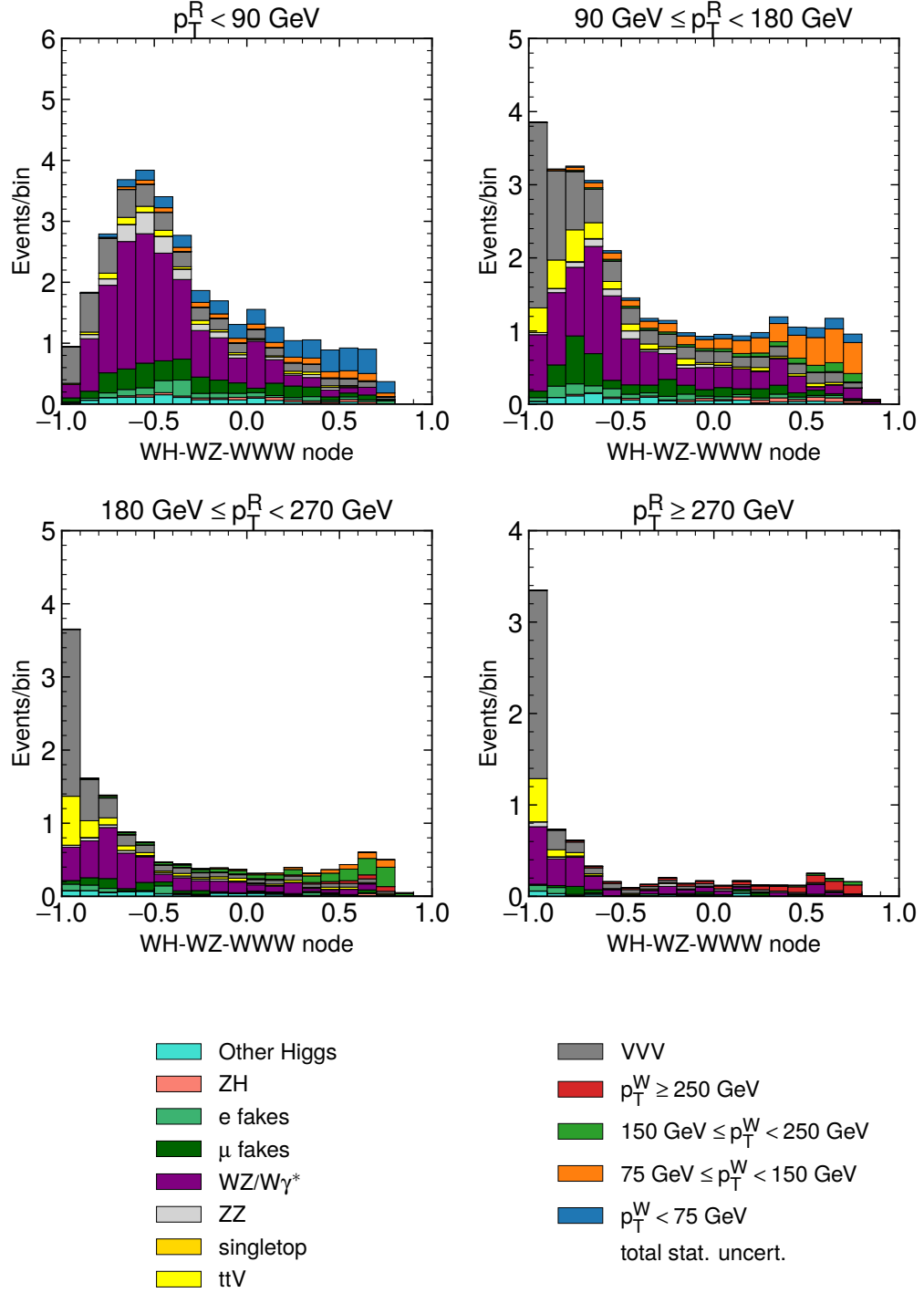


Figure 7.21.: Fine binning of the retrained multiclassifier output in the  $p_T^R$  bins.

### 7. Analysis in the Z Depleted Signal Region

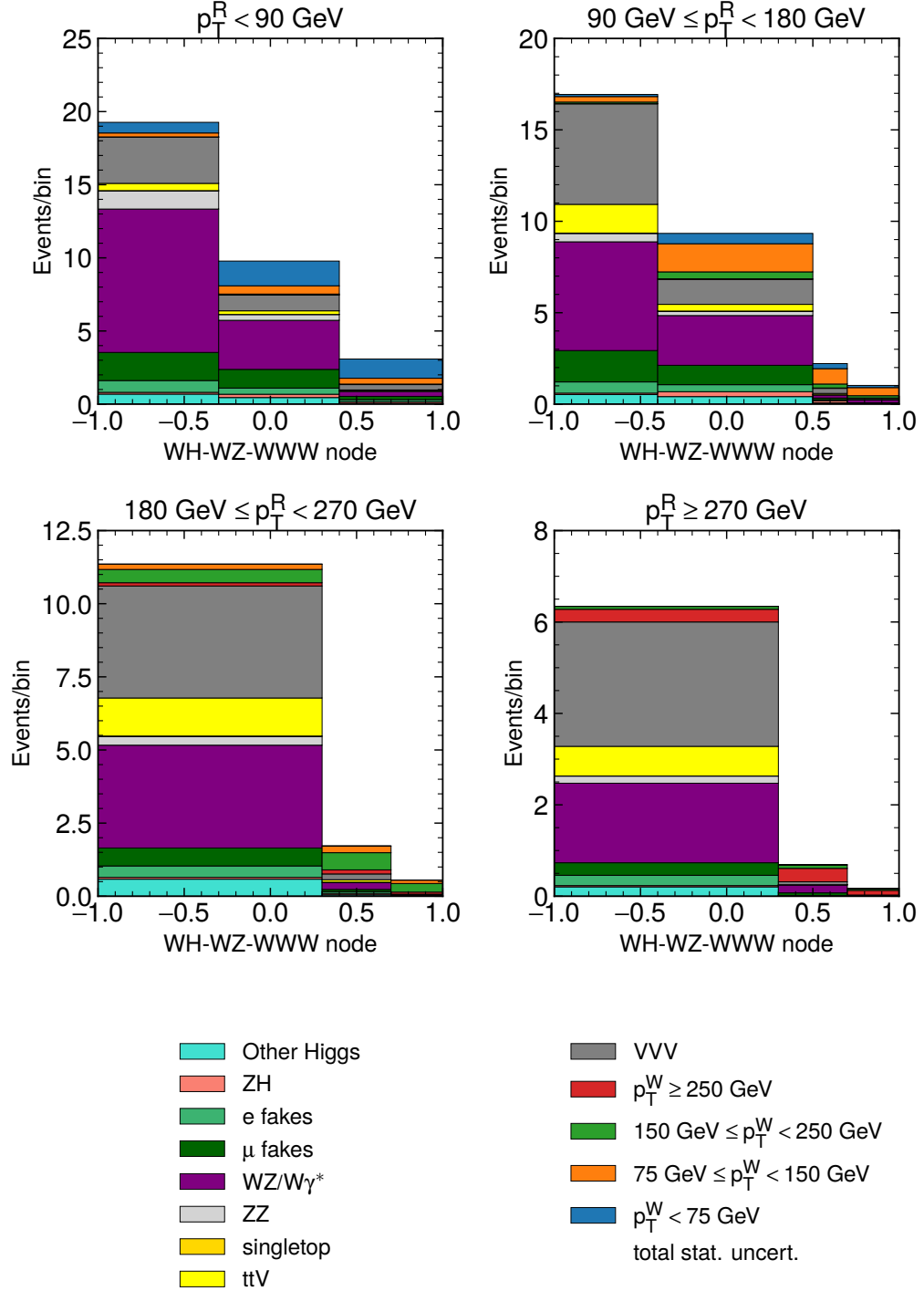


Figure 7.22.: Binning of the retrained multiclassifier output in the  $p_T^R$  bins.

## 7. Analysis in the $Z$ Depleted Signal Region

### 7.13. Results

In the following, the results of the analysis are summarized. Only results of Asimov fits can be given, as the  $pp \rightarrow WH \rightarrow WWW$  analysis this thesis is part of was still blinded at the time of writing [29].

Asimov fits were performed for both the old and the retrained classifier, and the expected 95%  $CL_S$  upper limits on the signal strengths of the STXS categories were calculated. The results of this are listed in Table 7.13. The retrained classifier performs slightly worse than the old one in terms of the uncertainties. This is likely a result of the removal of the impact parameters, which was already observed to negatively impact the performance of the classifier when comparing training variants in Section 7.8.3. In spite of this, it is desirable to use the new classifier due to the problems caused by using the impact parameters and the utility of the additional  $WWW$  output node. In terms of the upper limits, the retrained classifier performs slightly better than the old one in the lowest STXS bin and somewhat worse in the highest. All in all, it can be concluded that the retraining efforts were a success, in that the desired changes to the ANN could be achieved without significantly impacting the classifier performance.

Table 7.13.: Comparison between the old and retrained multiclassifier using the final binning.

|               |                                                | Old classifier,<br>no ECIDS and $b$ veto | Old classifier,<br>with ECIDS and $b$ veto | Retrained classifier |
|---------------|------------------------------------------------|------------------------------------------|--------------------------------------------|----------------------|
| Uncertainties | $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 1.236                                    | 1.203                                      | 1.240                |
|               | $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 1.477                                    | 1.468                                      | 1.518                |
|               | $150 \text{ GeV} \leq p_T^W < 250 \text{ GeV}$ | 1.913                                    | 1.924                                      | 1.981                |
|               | $p_T^W \geq 250 \text{ GeV}$                   | 2.072                                    | 2.047                                      | 2.215                |
| Limits        | $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 2.9                                      | 2.8                                        | 2.7                  |
|               | $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 3.1                                      | 3.1                                        | 3.1                  |
|               | $150 \text{ GeV} \leq p_T^W < 250 \text{ GeV}$ | 4.8                                      | 4.8                                        | 4.8                  |
|               | $p_T^W \geq 250 \text{ GeV}$                   | 6.8                                      | 6.6                                        | 7.3                  |

While the analysis presented here was optimized for the use of four STXS bins defined in the transverse momentum of the associated  $W$  boson, some of these bins could be merged. The advantage of this would be having a larger number of events for the merged STXS bin. A smaller number of bins generally makes fits simpler and more stable. The drawback, however, is that information of the shape of the  $WH$  signal in  $p_T^W$  is lost, which lowers the sensitivity to BSM effects.

Table 7.15 shows the fit results for different configurations of the STXS bins, both with and without the  $WZ$  scale uncertainties. As expected, the uncertainties are reduced when a smaller number of STXS bins is chosen. Table 7.14 shows the upper limits on the BSM signal derived in Section 7.11. The limits on  $\mu_{BSM}$  are very small, which is a result of the large total yield  $N_{BSM}$  to which the BSM distribution was normalized. As more meaningful numbers the corresponding upper limits on the number of additional BSM events  $\mu_{BSM} \cdot N_{BSM}$  are listed.

## 7. Analysis in the $Z$ Depleted Signal Region

Table 7.14.: Upper limits on the BSM signal for different STXS bin configurations, with and without  $WZ$  scale uncertainties.

|             |             | Statistical<br>uncertainties only | With $WZ$ scale<br>uncertainties |
|-------------|-------------|-----------------------------------|----------------------------------|
| 4 STXS bins | $\mu_{BSM}$ | 0.0383                            | 0.0392                           |
|             | Events      | 7.92                              | 8.11                             |
| 3 STXS bins | $\mu_{BSM}$ | 0.0475                            | 0.0488                           |
|             | Events      | 9.82                              | 10.28                            |

According to these results, the analysis is sensitive to BSM signals with a slightly smaller yield than the SM signal with 13.19 predicted events in the  $Z$  depleted signal region. These limits show the benefits of using a larger number of STXS bins to make the analysis more sensitive to BSM effects: The upper limit with 4 STXS bins is more than 20% lower than when the higher two STXS bins are merged.

When including the  $WZ$  scale uncertainties, the uncertainties on the signal strengths increase only very slightly, at the permille level. While only a small subset of all systematic uncertainties was included in the fit, since the effect of one of the dominant systematic uncertainties from the previous analysis in the channel is so small, it can be concluded that this analysis is dominated by statistical uncertainties and that the results reported here are close to what would be obtained with a more complete treatment.

The effect of the  $WZ$  systematics is more pronounced for the upper limits, at the 10% level. However, these differences are of similar size to the fluctuations in the results of the toy based upper limit calculations, so it cannot be concluded that the systematic uncertainties have a major effect on the upper limits.

The different tested STXS bin configuration include a fit with a single bin. This is done as a validation of the STXS analysis, as it should be similar to the result of the inclusive analysis. The Asimov fit result including experimental and theoretical systematic uncertainties for the inclusive  $Z$  depleted analysis is  $\mu_{WH} = 1^{+0.51}_{-0.42}$ , with an expected significance of  $2.72 \sigma$  [29]. This is well replicated by the result of this analysis with  $\mu_{WH} = 1 \pm 0.45$  including the  $WZ$  scale systematical uncertainties for a single STXS bin with an expected significance of  $2.99 \sigma$ . The small differences result from the different distributions entered into the fit. In the inclusive analysis the  $WH - WZ - WWW$  node is directly binned using an automated binning strategy [29], in the STXS analysis it is first split in  $p_T^R$  regions and then binned in the multiclassifier output as described in Section 7.5. The fact that the inclusive result using the full systematic uncertainties is so similar to the single bin STXS result with only the  $WZ$  scale systematical uncertainties is further evidence that the analysis is dominated by statistical uncertainties and that the results presented here are indicative of the performance of the STXS analysis, even after a full treatment of the uncertainties.

## 7. Analysis in the $Z$ Depleted Signal Region

Table 7.15.: Fit uncertainties and upper limits for different numbers of STXS bins.

|               |                                                | Statistical<br>uncertainties only | With $WZ$ scale<br>uncertainties |
|---------------|------------------------------------------------|-----------------------------------|----------------------------------|
| Uncertainties | $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 1.240                             | 1.240                            |
|               | $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 1.518                             | 1.519                            |
|               | $150 \text{ GeV} \leq p_T^W < 250 \text{ GeV}$ | 1.981                             | 1.983                            |
|               | $p_T^W \geq 250 \text{ GeV}$                   | 2.215                             | 2.218                            |
| Limits        | $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 2.7                               | 2.8                              |
|               | $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 3.1                               | 3.1                              |
|               | $150 \text{ GeV} \leq p_T^W < 250 \text{ GeV}$ | 4.8                               | 4.9                              |
|               | $p_T^W \geq 250 \text{ GeV}$                   | 7.3                               | 7.3                              |
| Uncertainties | $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 1.219                             | 1.220                            |
|               | $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 1.370                             | 1.371                            |
|               | $p_T^W \geq 150 \text{ GeV}$                   | 1.118                             | 1.119                            |
| Limits        | $0 \text{ GeV} \leq p_T^W < 75 \text{ GeV}$    | 2.7                               | 2.8                              |
|               | $75 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  | 3.1                               | 3.1                              |
|               | $p_T^W \geq 150 \text{ GeV}$                   | 2.9                               | 3.0                              |
| Uncertainties | $0 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$   | 0.607                             | 0.608                            |
|               | $p_T^W \geq 150 \text{ GeV}$                   | 1.040                             | 1.041                            |
| Limits        | $0 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$   | 1.5                               | 1.4                              |
|               | $p_T^W \geq 150 \text{ GeV}$                   | 2.9                               | 2.9                              |
| Uncertainty   |                                                | 0.449                             | 0.449                            |
| Limit         | inclusive                                      | 1.0                               | 1.0                              |

The results using two STXS bins can be compared to an existing measurement, which was performed by the CMS collaboration [57]. In the  $WH \rightarrow WWW \rightarrow \nu\ell\nu\ell\nu\ell$  channel, they used  $p_T^{l_2}$  as a proxy for  $p_T^W$ . The expected significances of the measurement, combined over all  $WH \rightarrow WWW$  channels, were reported by CMS as  $1.24 \sigma$  for  $0 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  and  $0.83 \sigma$  for  $p_T^W \geq 150 \text{ GeV}$ . As a comparison, hypothesis tests were performed in the 2 STXS bin configuration, using only statistical uncertainties. When performing the test for  $\mu_{0 \text{ GeV} \leq p_T^W < 150 \text{ GeV}}$ , the parameter  $\mu_{p_T^W \geq 150 \text{ GeV}}$  was profiled, and vice versa. The background-only hypothesis  $\mu_i = 0$  can be excluded at an expected significance of  $2.02 \sigma$  for  $0 \text{ GeV} \leq p_T^W < 150 \text{ GeV}$  and  $1.29 \sigma$  for  $p_T^W \geq 150 \text{ GeV}$ . While a more meaningful comparison would need to include the full statistical treatment of the STXS analysis including e.g. systematical uncertainties, considering that the analysis presented here is dominated by statistical uncertainties and that the sensitivity will improve when combining the  $Z$  depleted channel with the  $Z$  dominated channel and the two and four lepton channels, it can be concluded that it can match or exceed the CMS result in terms of the STXS sensitivity.

## 8. Summary and Outlook

This thesis is concerned with the production of a Higgs boson in association with a  $W$  boson and the subsequent decay of the Higgs boson to two  $W$  bosons, with a fully leptonic final state. In the  $Z$  depleted signal region a Simplified Template Cross Section (STXS) analysis was developed, i.e. a measurement in different bins of  $p_T^W$ , the transverse momentum of the associated  $W$  boson. The analysis uses  $139 \text{ fb}^{-1}$  of proton-proton collision data collected by the ATLAS experiment at a center of mass energy of 13 TeV.

Since  $p_T^W$  cannot be reconstructed analytically due to the presence of three neutrinos from the  $W$  decays, proxies were studied, i.e. variables that strongly correlate with  $p_T^W$ . First, different kinematic observables were tested and compared based on expected uncertainties obtained from signal-only likelihood fits. Of these simple proxies, the scalar sum of the lepton transverse momenta and the missing transverse energy was found to perform best.

To improve on the simple proxies, a regression neural network was trained to reconstruct  $p_T^W$  and its architecture and input variables were optimized. Its output  $p_T^R$  was found to significantly outperform the simple proxies. Good agreement between MC simulation and data was found for  $p_T^R$ .

A multiclassifier that had previously been developed as a discriminant in the  $Z$  depleted signal region was retrained with the goal of removing problematic input variables and adding an additional output node, which can be used to define a region enhanced in the  $WWW$  background. These changes could be made with only minor impacts on the performance of the multiclassifier, while improving the robustness and capabilities of the analysis.

A fit strategy was developed, in which the signal region is split into  $p_T^R$  bins and a combined fit of the multiclassifier discriminant is performed. This was used to calculate expected uncertainties of the signal strengths in the STXS bins, as well as expected upper limits on possible excesses above the SM predictions.

The sensitivity to an exemplary beyond the Standard Model signal was studied. It was found that the analysis presented here is sensitive to an excess of about 8 events above the Standard Model prediction of about 13 signal events in the  $Z$  depleted signal region.

Before the analysis can be unblinded and an STXS measurement with the real data can be performed, systematic uncertainties will need to be studied and included in

## 8. Summary and Outlook

the fits. Furthermore, changes may need to be made to the binning to accommodate combinations with different measurement channels.

A study of one of the most important statistical uncertainties showed that the STXS analysis is dominated by statistical uncertainties. Therefore, its sensitivity will improve when additional data with an expected integrated luminosity of  $300 \text{ fb}^{-1}$  is taken in the upcoming Run 3 of the LHC [58]. Furthermore, it is expected that the sensitivity reported here is close to that of the full treatment.

Once published as part of the  $VH H \rightarrow WW$  analysis currently being developed [29], the analysis presented in this thesis will be the first STXS measurement of  $pp \rightarrow WH \rightarrow WWW \rightarrow l\nu l\nu l\nu$  in the  $Z$  depleted channel at the ATLAS experiment.

# Acknowledgments

I would like to thank Prof. Volker Büscher for the opportunity to join an ATLAS analysis and become part of the ongoing hunt for new physics. I am grateful for his advice and guidance over the past year.

I would like to thank Dr. Christian Schmitt and Dr. Ralf Gugel for their support in my day-to-day work. I never felt left alone with any problems, be they physics questions or technical trouble.

I am thankful to the VH-HWW analysis team and the other analyzers in Mainz for the helpful discussions and advice during our weekly meetings. I would also like to thank my colleagues in Mainz for the welcoming office atmosphere, even though due to the pandemic I hardly got to experience it.

I am grateful to my family and friends for their enthusiastic support during my thesis work and my studies before that.

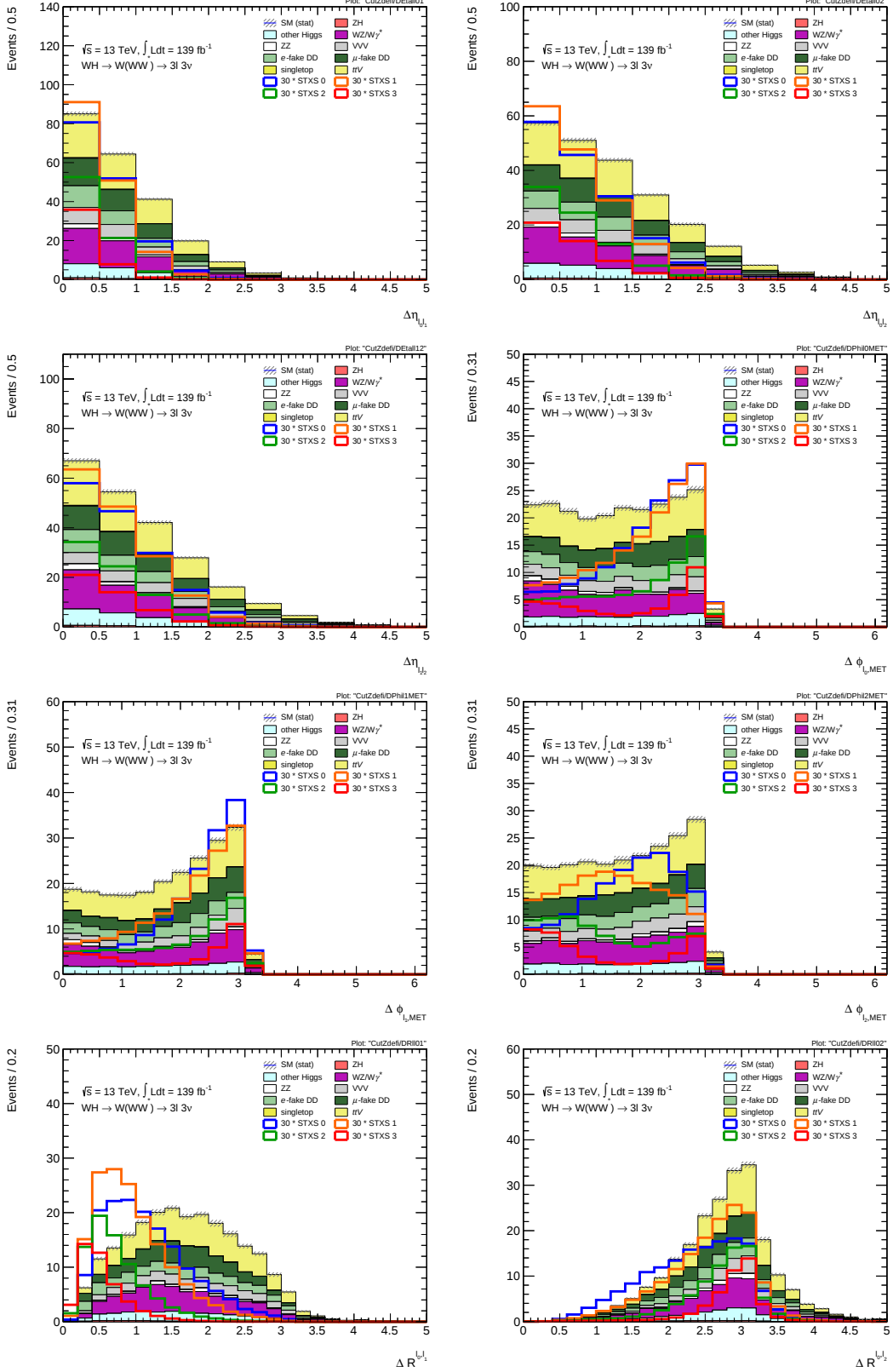
Last but not least I would like to thank everyone who proofread and spell-checked my thesis and helped me add the finishing touches.



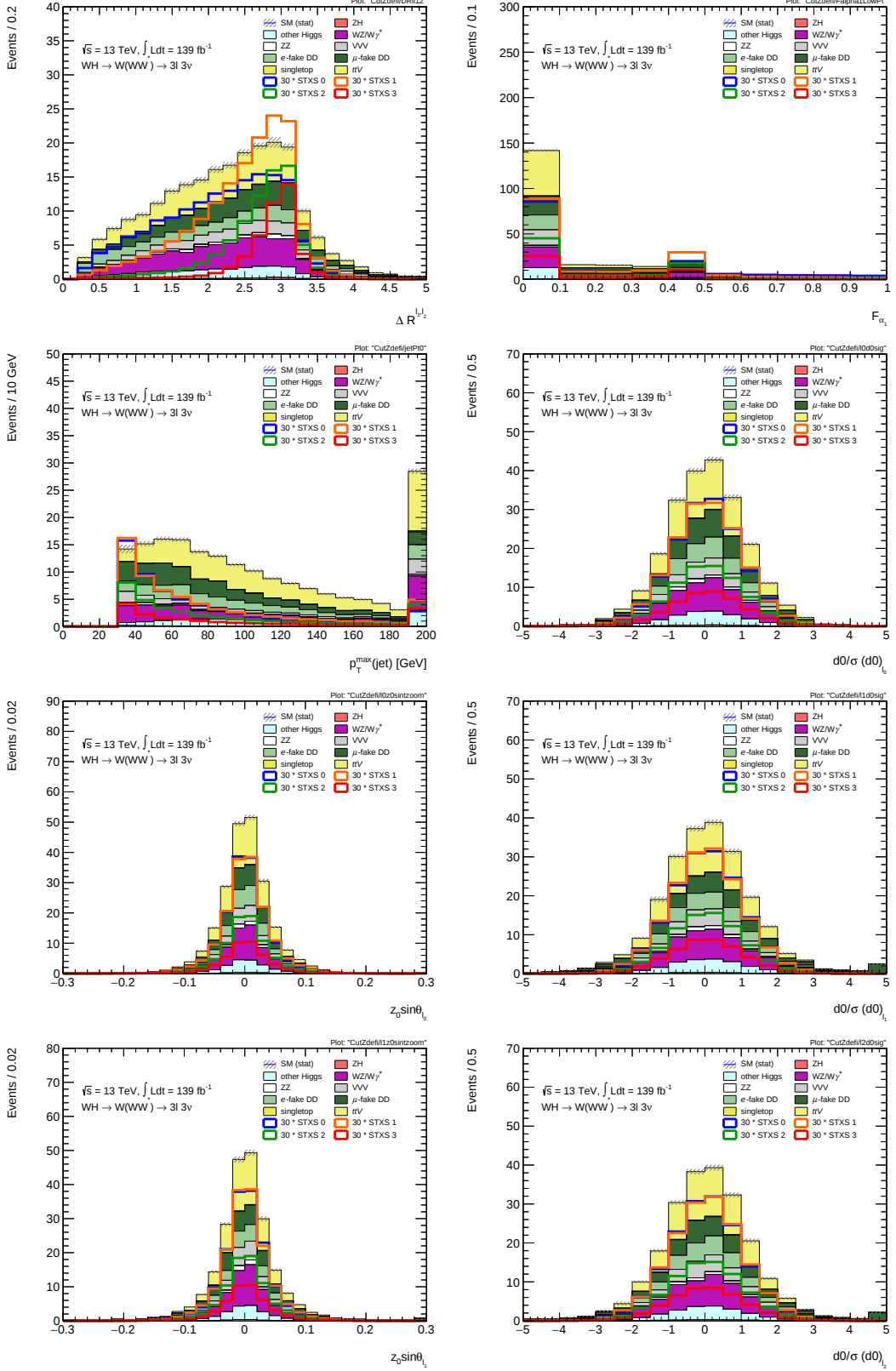
## A. Neural Network Input Variables

Shown here are the distributions of all variables that were used as inputs to neural networks. The 0 SFOS selection of the  $Z$  depleted signal region is applied.

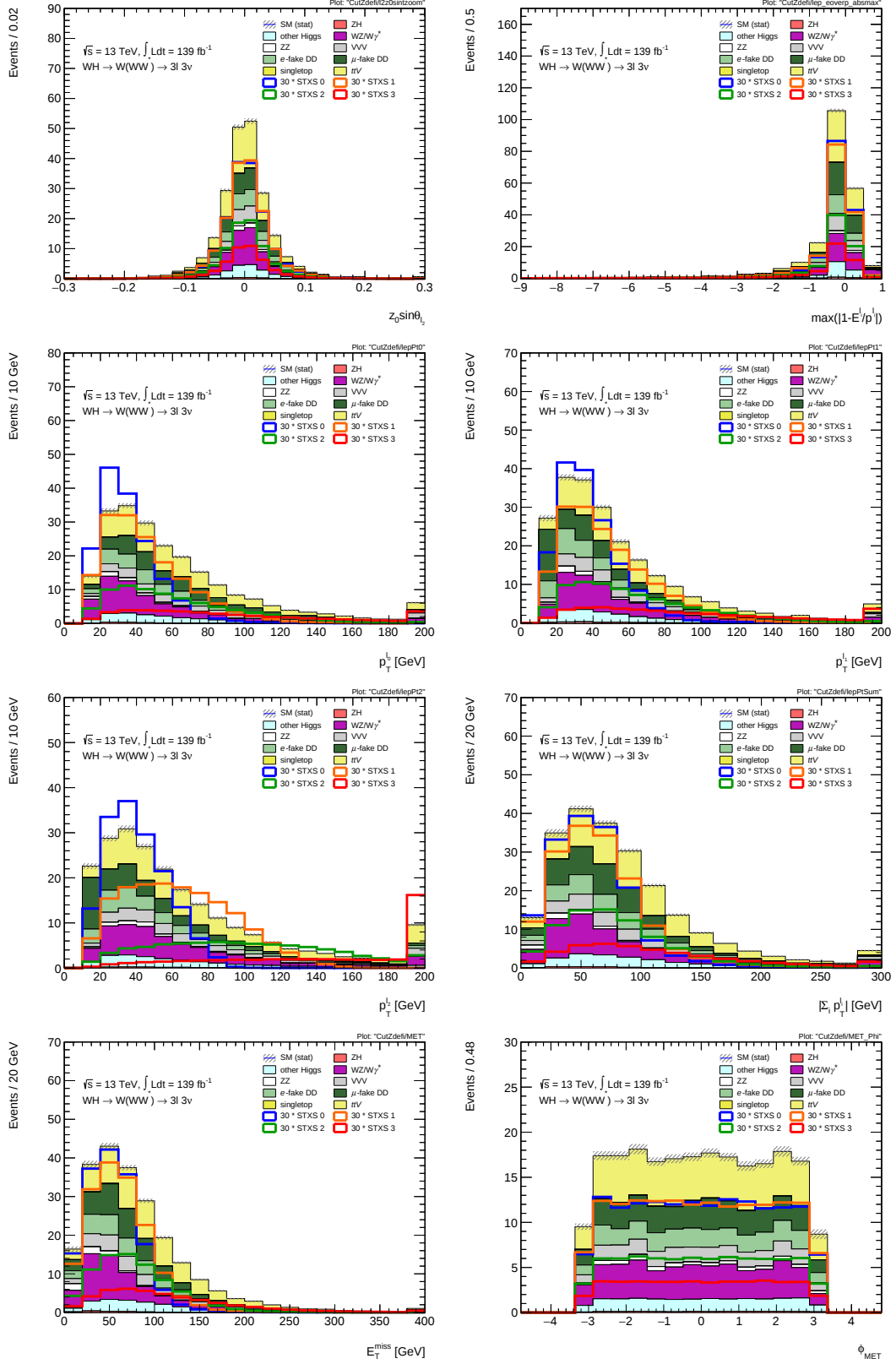
## A. Neural Network Input Variables



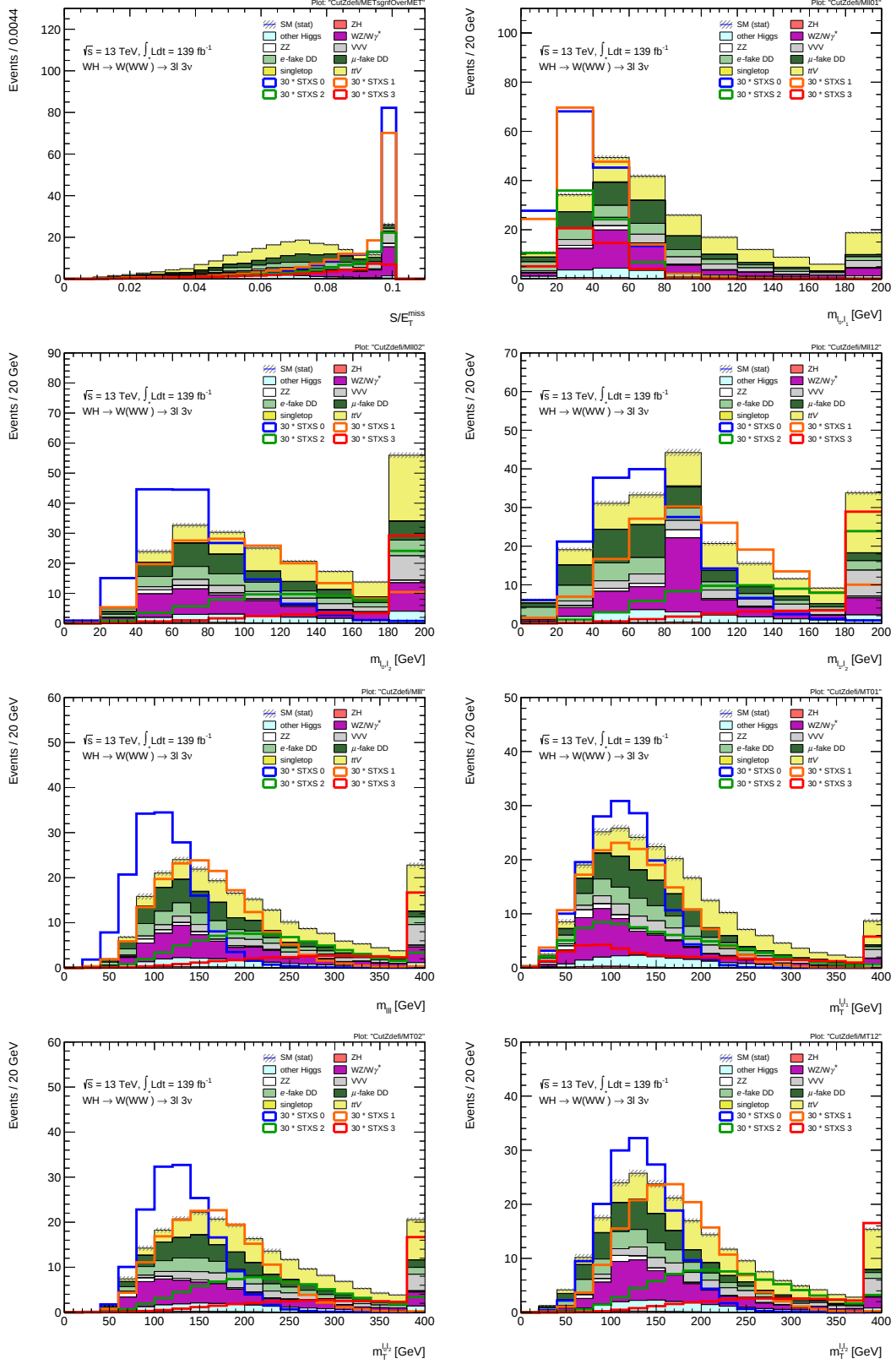
## A. Neural Network Input Variables



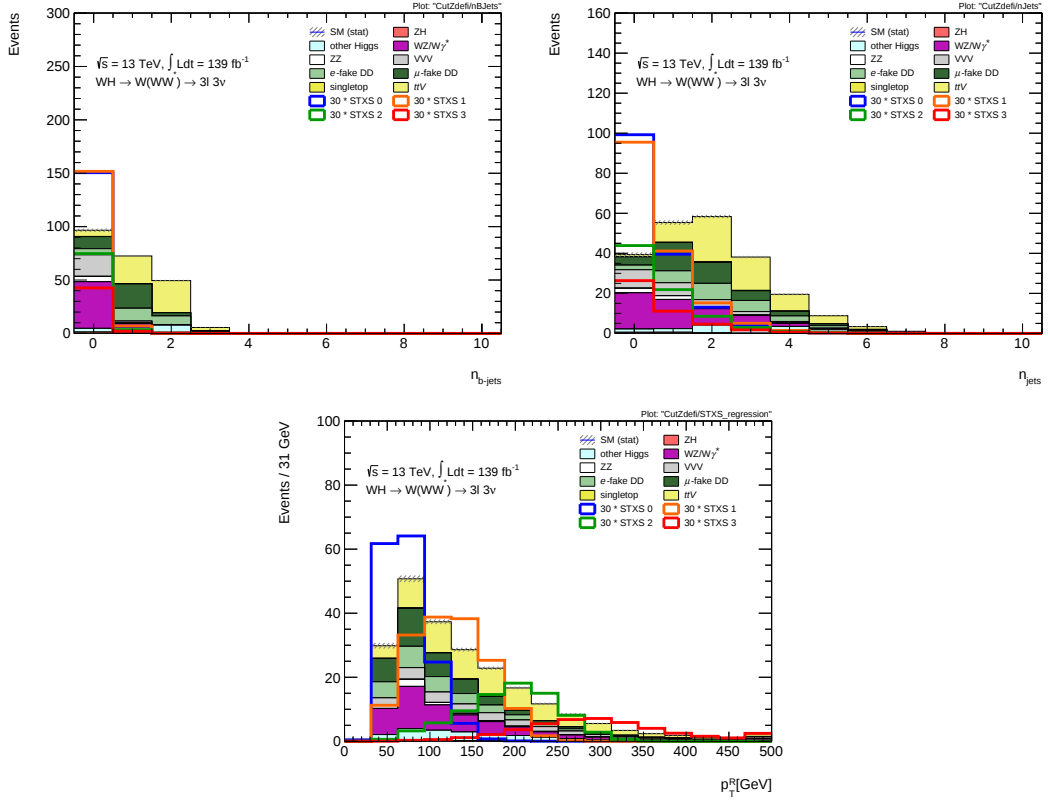
## A. Neural Network Input Variables



## A. Neural Network Input Variables



## A. Neural Network Input Variables



## B. Cutflows

In the following, cutflows are shown for the different signal, control and validation regions. Each row shows the yields of the different processes after a selection and the selections above it are applied.

## B. Cutflows

Table B.1.:  $Z$  depleted signal region including  $p_T^R$  bins.

|                                                | STXS 0           | STXS 1           | STXS 2           | STXS 3          | total WH         | ZH               | other Higgs       |
|------------------------------------------------|------------------|------------------|------------------|-----------------|------------------|------------------|-------------------|
| Preselection                                   | $21.74 \pm 0.07$ | $22.02 \pm 0.06$ | $10.89 \pm 0.02$ | $4.76 \pm 0.01$ | $59.41 \pm 0.10$ | $57.91 \pm 0.32$ | $103.28 \pm 0.52$ |
| $Z$ depleted SR: 0 SFOS                        | $5.25 \pm 0.03$  | $5.30 \pm 0.03$  | $2.62 \pm 0.01$  | $1.16 \pm 0.01$ | $14.34 \pm 0.05$ | $1.46 \pm 0.06$  | $18.45 \pm 0.17$  |
| All electrons pass ECIDS                       | $5.13 \pm 0.03$  | $5.17 \pm 0.03$  | $2.56 \pm 0.01$  | $1.14 \pm 0.01$ | $14.00 \pm 0.05$ | $1.37 \pm 0.06$  | $17.99 \pm 0.16$  |
| $n_{b-jets} = 0$                               | $4.90 \pm 0.03$  | $4.93 \pm 0.03$  | $2.43 \pm 0.01$  | $1.07 \pm 0.01$ | $13.32 \pm 0.05$ | $1.14 \pm 0.05$  | $3.35 \pm 0.14$   |
| ttbar node $\leq 0.3$                          | $4.77 \pm 0.03$  | $4.84 \pm 0.03$  | $2.41 \pm 0.01$  | $1.07 \pm 0.01$ | $13.08 \pm 0.05$ | $1.09 \pm 0.05$  | $3.09 \pm 0.14$   |
| $p_T^R < 90$ GeV                               | $3.70 \pm 0.03$  | $1.22 \pm 0.02$  | $0.10 \pm 0.00$  | $0.01 \pm 0.00$ | $5.02 \pm 0.03$  | $0.45 \pm 0.04$  | $1.20 \pm 0.09$   |
| $90 \text{ GeV} \leq p_T^R < 180 \text{ GeV}$  | $1.07 \pm 0.02$  | $3.10 \pm 0.02$  | $0.80 \pm 0.01$  | $0.07 \pm 0.00$ | $5.04 \pm 0.03$  | $0.48 \pm 0.03$  | $1.03 \pm 0.08$   |
| $180 \text{ GeV} \leq p_T^R < 270 \text{ GeV}$ | $0.01 \pm 0.00$  | $0.52 \pm 0.01$  | $1.33 \pm 0.01$  | $0.32 \pm 0.00$ | $2.17 \pm 0.01$  | $0.12 \pm 0.02$  | $0.64 \pm 0.06$   |
| $p_T^R \geq 270 \text{ GeV}$                   | $0.00 \pm 0.00$  | $0.00 \pm 0.00$  | $0.18 \pm 0.00$  | $0.67 \pm 0.00$ | $0.85 \pm 0.00$  | $0.03 \pm 0.01$  | $0.22 \pm 0.03$   |

|                                                | WZ/ $W\gamma^*$      | $t\bar{t}V$        | Single Top        | $ZZ^*$              | VVV               | fake electrons      | fake muons        | Total Bkg            |
|------------------------------------------------|----------------------|--------------------|-------------------|---------------------|-------------------|---------------------|-------------------|----------------------|
| Preselection                                   | $22277.60 \pm 33.85$ | $1045.47 \pm 2.48$ | $299.52 \pm 0.52$ | $3971.57 \pm 16.38$ | $125.16 \pm 0.26$ | $2288.36 \pm 17.73$ | $458.33 \pm 3.26$ | $30893.09 \pm 41.79$ |
| $Z$ depleted SR: 0 SFOS                        | $47.42 \pm 1.32$     | $64.77 \pm 0.69$   | $0.70 \pm 0.02$   | $5.43 \pm 0.33$     | $21.89 \pm 0.09$  | $26.63 \pm 0.57$    | $37.42 \pm 0.85$  | $224.55 \pm 1.85$    |
| All electrons pass ECIDS                       | $32.21 \pm 1.10$     | $62.82 \pm 0.68$   | $0.49 \pm 0.02$   | $3.27 \pm 0.16$     | $21.32 \pm 0.09$  | $18.72 \pm 0.37$    | $36.98 \pm 0.85$  | $195.43 \pm 1.61$    |
| $n_{b-jets} = 0$                               | $29.52 \pm 1.08$     | $5.69 \pm 0.20$    | $0.09 \pm 0.01$   | $2.73 \pm 0.15$     | $19.55 \pm 0.08$  | $3.90 \pm 0.26$     | $11.18 \pm 0.47$  | $77.38 \pm 1.24$     |
| ttbar node $\leq 0.3$                          | $28.12 \pm 1.06$     | $4.93 \pm 0.19$    | $0.07 \pm 0.01$   | $2.62 \pm 0.15$     | $18.64 \pm 0.08$  | $3.14 \pm 0.25$     | $7.33 \pm 0.38$   | $69.25 \pm 1.19$     |
| $p_T^R < 90$ GeV                               | $13.48 \pm 0.74$     | $0.81 \pm 0.07$    | $0.02 \pm 0.00$   | $1.54 \pm 0.11$     | $4.61 \pm 0.04$   | $1.38 \pm 0.20$     | $3.42 \pm 0.26$   | $27.07 \pm 0.83$     |
| $90 \text{ GeV} \leq p_T^R < 180 \text{ GeV}$  | $8.99 \pm 0.67$      | $2.05 \pm 0.13$    | $0.04 \pm 0.01$   | $0.62 \pm 0.08$     | $7.20 \pm 0.05$   | $1.08 \pm 0.10$     | $2.88 \pm 0.23$   | $24.43 \pm 0.73$     |
| $180 \text{ GeV} \leq p_T^R < 270 \text{ GeV}$ | $3.73 \pm 0.28$      | $1.42 \pm 0.10$    | $0.01 \pm 0.00$   | $0.30 \pm 0.05$     | $4.04 \pm 0.04$   | $0.46 \pm 0.11$     | $0.69 \pm 0.11$   | $11.43 \pm 0.35$     |
| $p_T^R \geq 270 \text{ GeV}$                   | $1.91 \pm 0.23$      | $0.65 \pm 0.07$    | $0.00 \pm 0.00$   | $0.15 \pm 0.03$     | $2.78 \pm 0.03$   | $0.22 \pm 0.05$     | $0.34 \pm 0.08$   | $6.32 \pm 0.27$      |

## B. Cutflows

Table B.3.: Top validation region.

| $\sqrt{s} = 13\text{TeV}, \mathcal{L} = 139\text{fb}^{-1}$ (Full Run 2) | STXS 0       |  | STXS 1           |                  | STXS 2           |                 | STXS 3           |                  | total WH          |  | ZH              |                  | other Higgs |  |
|-------------------------------------------------------------------------|--------------|--|------------------|------------------|------------------|-----------------|------------------|------------------|-------------------|--|-----------------|------------------|-------------|--|
|                                                                         | Preselection |  | 21.74 $\pm$ 0.07 | 22.02 $\pm$ 0.06 | 10.89 $\pm$ 0.02 | 4.76 $\pm$ 0.01 | 59.41 $\pm$ 0.10 | 57.91 $\pm$ 0.32 | 103.28 $\pm$ 0.52 |  | 1.46 $\pm$ 0.06 | 18.45 $\pm$ 0.17 |             |  |
| 0 SFOS                                                                  |              |  | 5.25 $\pm$ 0.03  | 5.30 $\pm$ 0.03  | 2.62 $\pm$ 0.01  | 1.16 $\pm$ 0.01 | 14.34 $\pm$ 0.05 | 0.23 $\pm$ 0.02  | 5.60 $\pm$ 0.06   |  |                 |                  |             |  |
| $n_{b-jets} = 1$                                                        |              |  | 0.23 $\pm$ 0.01  | 0.24 $\pm$ 0.01  | 0.12 $\pm$ 0.00  | 0.07 $\pm$ 0.00 | 0.65 $\pm$ 0.01  |                  |                   |  |                 |                  |             |  |

| $\sqrt{s} = 13\text{TeV}, \mathcal{L} = 139\text{fb}^{-1}$ (Full Run 2) | WZ/ $W\gamma^*$ |  | $t\bar{t}$           |                    | Single Top        |                     | $ZZ^*$            |                     | VVV               |                      | fake electrons |  | fake muons |  | Total Bkg |  | Data |  |
|-------------------------------------------------------------------------|-----------------|--|----------------------|--------------------|-------------------|---------------------|-------------------|---------------------|-------------------|----------------------|----------------|--|------------|--|-----------|--|------|--|
|                                                                         | Preselection    |  | 22277.60 $\pm$ 33.85 | 1045.47 $\pm$ 2.48 | 299.52 $\pm$ 0.52 | 3971.57 $\pm$ 16.38 | 125.16 $\pm$ 0.26 | 2288.36 $\pm$ 17.73 | 458.33 $\pm$ 3.26 | 30893.09 $\pm$ 41.79 | 30497          |  |            |  |           |  |      |  |
| 0 SFOS                                                                  |                 |  | 47.42 $\pm$ 1.32     | 64.77 $\pm$ 0.69   | 0.70 $\pm$ 0.02   | 5.43 $\pm$ 0.33     | 21.89 $\pm$ 0.09  | 26.63 $\pm$ 0.57    | 37.42 $\pm$ 0.85  | 224.55 $\pm$ 1.85    | 285            |  |            |  |           |  |      |  |
| $n_{b-jets} = 1$                                                        |                 |  | 3.47 $\pm$ 0.25      | 25.90 $\pm$ 0.44   | 0.42 $\pm$ 0.02   | 0.71 $\pm$ 0.06     | 1.71 $\pm$ 0.03   | 12.03 $\pm$ 0.24    | 22.72 $\pm$ 0.66  | 72.83 $\pm$ 0.87     | 79             |  |            |  |           |  |      |  |

## B. Cutflows

Table B.5.:  $WZ$  control region.

| $\sqrt{s} = 13TeV, \mathcal{L} = 139fb^{-1}$ (Full Run 2)                     | STXS 0                |                      | STXS 1            |                     | STXS 2            |                     | STXS 3            |                      | total WH         |                   | ZH           |  | other Higgs |
|-------------------------------------------------------------------------------|-----------------------|----------------------|-------------------|---------------------|-------------------|---------------------|-------------------|----------------------|------------------|-------------------|--------------|--|-------------|
|                                                                               | Preselection          |                      | Preselection      |                     | Preselection      |                     | Preselection      |                      | Preselection     |                   | Preselection |  |             |
| SFOS pair with $ m_{ll} - m_Z  < 25$ GeV<br>SFOS pairs have $m_{ll} > 12$ GeV | $i, 0$ SFOS           | $21.74 \pm 0.07$     | $22.02 \pm 0.06$  | $10.89 \pm 0.02$    | $8.27 \pm 0.02$   | $3.60 \pm 0.01$     | $4.76 \pm 0.01$   | $59.41 \pm 0.10$     | $57.91 \pm 0.32$ | $103.28 \pm 0.52$ |              |  |             |
|                                                                               | $n_{jets} = 0$        | $16.49 \pm 0.06$     | $16.72 \pm 0.06$  | $8.27 \pm 0.02$     | $8.27 \pm 0.02$   | $3.60 \pm 0.01$     | $3.60 \pm 0.01$   | $45.07 \pm 0.08$     | $56.45 \pm 0.31$ | $84.83 \pm 0.49$  |              |  |             |
|                                                                               | $n_{b-jets} = 0$      | $14.52 \pm 0.06$     | $14.40 \pm 0.05$  | $6.89 \pm 0.01$     | $6.89 \pm 0.01$   | $2.84 \pm 0.01$     | $2.84 \pm 0.01$   | $38.65 \pm 0.08$     | $39.93 \pm 0.27$ | $29.19 \pm 0.41$  |              |  |             |
|                                                                               | $E_T^{miss} > 30$ GeV | $14.09 \pm 0.06$     | $13.95 \pm 0.05$  | $6.66 \pm 0.01$     | $6.66 \pm 0.01$   | $2.74 \pm 0.01$     | $2.74 \pm 0.01$   | $37.44 \pm 0.08$     | $35.27 \pm 0.26$ | $22.05 \pm 0.38$  |              |  |             |
|                                                                               |                       | $11.14 \pm 0.05$     | $11.50 \pm 0.05$  | $5.84 \pm 0.01$     | $5.84 \pm 0.01$   | $2.52 \pm 0.01$     | $2.52 \pm 0.01$   | $31.01 \pm 0.07$     | $25.86 \pm 0.22$ | $16.05 \pm 0.33$  |              |  |             |
| SFOS pair with $ m_{ll} - m_Z  < 25$ GeV<br>SFOS pairs have $m_{ll} > 12$ GeV |                       | $3.38 \pm 0.03$      | $3.81 \pm 0.03$   | $1.18 \pm 0.01$     | $1.18 \pm 0.01$   | $0.24 \pm 0.00$     | $0.24 \pm 0.00$   | $8.60 \pm 0.04$      | $23.94 \pm 0.21$ | $13.02 \pm 0.29$  |              |  |             |
|                                                                               |                       | $3.26 \pm 0.03$      | $3.68 \pm 0.03$   | $1.14 \pm 0.01$     | $1.14 \pm 0.01$   | $0.23 \pm 0.00$     | $0.23 \pm 0.00$   | $8.32 \pm 0.04$      | $23.80 \pm 0.21$ | $12.97 \pm 0.29$  |              |  |             |
| $\sqrt{s} = 13TeV, \mathcal{L} = 139fb^{-1}$ (Full Run 2)                     | Preselection          | WZ/ $W\gamma^*$      | Single Top        | $ZZ^*$              | VVV               | fake electrons      | fake muons        | Total Bkg            | Data             |                   |              |  |             |
|                                                                               | $i, 0$ SFOS           | $22277.60 \pm 33.85$ | $299.52 \pm 0.52$ | $3971.57 \pm 16.38$ | $125.16 \pm 0.26$ | $2288.36 \pm 17.73$ | $458.33 \pm 3.26$ | $30893.09 \pm 41.79$ | $30497$          |                   |              |  |             |
|                                                                               | $n_{jets} = 0$        | $22230.18 \pm 33.83$ | $298.82 \pm 0.51$ | $3966.14 \pm 16.38$ | $103.27 \pm 0.25$ | $2261.74 \pm 17.72$ | $420.91 \pm 3.14$ | $30668.55 \pm 41.75$ | $30212$          |                   |              |  |             |
|                                                                               | $n_{b-jets} = 0$      | $16308.69 \pm 32.79$ | $51.69 \pm 0.21$  | $3204.02 \pm 15.31$ | $69.05 \pm 0.19$  | $1937.84 \pm 15.76$ | $312.66 \pm 2.72$ | $22210.95 \pm 39.58$ | $22278$          |                   |              |  |             |
|                                                                               | $E_T^{miss} > 30$ GeV | $15665.55 \pm 32.32$ | $24.49 \pm 0.15$  | $2940.91 \pm 14.96$ | $64.17 \pm 0.18$  | $1849.28 \pm 15.15$ | $252.31 \pm 2.48$ | $21038.76 \pm 38.80$ | $20938$          |                   |              |  |             |
| SFOS pair with $ m_{ll} - m_Z  < 25$ GeV<br>SFOS pairs have $m_{ll} > 12$ GeV |                       | $11319.45 \pm 27.84$ | $20.84 \pm 0.14$  | $777.25 \pm 6.12$   | $55.52 \pm 0.17$  | $250.37 \pm 6.54$   | $87.56 \pm 1.51$  | $12625.39 \pm 29.29$ | $12369$          |                   |              |  |             |
|                                                                               |                       | $10567.36 \pm 27.03$ | $6.54 \pm 0.19$   | $652.08 \pm 4.96$   | $31.10 \pm 0.14$  | $176.59 \pm 4.97$   | $70.90 \pm 1.37$  | $11619.77 \pm 27.96$ | $11343$          |                   |              |  |             |
|                                                                               |                       | $10501.47 \pm 26.97$ | $6.51 \pm 0.19$   | $635.43 \pm 4.76$   | $30.99 \pm 0.14$  | $165.88 \pm 4.83$   | $69.58 \pm 1.35$  | $11524.22 \pm 27.85$ | $11233$          |                   |              |  |             |

## B. Cutflows

Table B.7.:  $WWW$  validation region.

| $\sqrt{s} = 13TeV, \mathcal{L} = 139fb^{-1}$ (Full Run 2) |              |              |              |             |              |              |               |  |  |  |
|-----------------------------------------------------------|--------------|--------------|--------------|-------------|--------------|--------------|---------------|--|--|--|
|                                                           | STXS 0       | STXS 1       | STXS 2       | STXS 3      | total WH     | ZH           | other Higgs   |  |  |  |
| Preselection                                              | 21.74 ± 0.07 | 22.02 ± 0.06 | 10.89 ± 0.02 | 4.76 ± 0.01 | 59.41 ± 0.10 | 57.91 ± 0.32 | 103.28 ± 0.52 |  |  |  |
| $i, 0$ SFOS                                               | 16.49 ± 0.06 | 16.72 ± 0.06 | 8.27 ± 0.02  | 3.60 ± 0.01 | 45.07 ± 0.08 | 56.45 ± 0.31 | 84.83 ± 0.49  |  |  |  |
| $n_{jets} = 0$                                            | 14.52 ± 0.06 | 14.40 ± 0.05 | 6.89 ± 0.01  | 2.84 ± 0.01 | 38.65 ± 0.08 | 39.93 ± 0.27 | 29.19 ± 0.41  |  |  |  |
| $n_{b-jets} = 0$                                          | 14.09 ± 0.06 | 13.95 ± 0.05 | 6.66 ± 0.01  | 2.74 ± 0.01 | 37.44 ± 0.08 | 35.27 ± 0.26 | 22.05 ± 0.38  |  |  |  |
| $E_T^{miss} > 30$ GeV                                     | 11.14 ± 0.05 | 11.50 ± 0.05 | 5.84 ± 0.01  | 2.52 ± 0.01 | 31.01 ± 0.07 | 25.86 ± 0.22 | 16.05 ± 0.33  |  |  |  |
| SFOS pair with $ m_l - m_Z  < 25$ GeV                     | 3.38 ± 0.03  | 3.81 ± 0.03  | 1.18 ± 0.01  | 0.24 ± 0.00 | 8.60 ± 0.04  | 23.94 ± 0.21 | 13.02 ± 0.29  |  |  |  |
| SFOS pairs have $m_l > 12$ GeV                            | 3.26 ± 0.03  | 3.68 ± 0.03  | 1.14 ± 0.01  | 0.23 ± 0.00 | 8.32 ± 0.04  | 23.80 ± 0.21 | 12.97 ± 0.29  |  |  |  |

| $\sqrt{s} = 13TeV, \mathcal{L} = 139fb^{-1}$ (Full Run 2) |                  |                |               |                 |               |                 |               |                  |       |  |
|-----------------------------------------------------------|------------------|----------------|---------------|-----------------|---------------|-----------------|---------------|------------------|-------|--|
|                                                           | WZ/ $W\gamma^*$  | $t\bar{t}V$    | Single Top    | $ZZ^*$          | VVV           | fake electrons  | fake muons    | Total Bkg        | Data  |  |
| Preselection                                              | 22277.60 ± 33.85 | 1045.47 ± 2.48 | 299.52 ± 0.52 | 3971.57 ± 16.38 | 125.16 ± 0.26 | 2288.36 ± 17.73 | 458.33 ± 3.26 | 30893.00 ± 41.79 | 30497 |  |
| $WWW$ node $\geq 0.5$                                     | 8577.94 ± 22.23  | 126.12 ± 0.88  | 38.26 ± 0.18  | 582.93 ± 4.37   | 70.82 ± 0.19  | 160.56 ± 4.27   | 26.23 ± 0.83  | 9646.72 ± 23.09  | 9340  |  |
| $0$ SFOS                                                  | 7.89 ± 0.72      | 9.70 ± 0.27    | 0.05 ± 0.01   | 0.61 ± 0.07     | 12.63 ± 0.07  | 1.60 ± 0.10     | 2.15 ± 0.20   | 35.96 ± 0.81     | 53    |  |
| All electrons pass ECIDS                                  | 3.53 ± 0.53      | 9.38 ± 0.26    | 0.02 ± 0.00   | 0.19 ± 0.04     | 12.30 ± 0.06  | 1.13 ± 0.08     | 2.15 ± 0.20   | 29.93 ± 0.64     | 48    |  |
| $n_{b-jets} = 0$                                          | 3.40 ± 0.53      | 2.97 ± 0.15    | 0.01 ± 0.00   | 0.18 ± 0.04     | 11.77 ± 0.06  | 0.79 ± 0.07     | 1.94 ± 0.18   | 21.72 ± 0.59     | 41    |  |
| ttbar node $\leq 0.3$                                     | 3.35 ± 0.53      | 2.85 ± 0.15    | 0.01 ± 0.00   | 0.17 ± 0.04     | 11.61 ± 0.06  | 0.72 ± 0.07     | 1.77 ± 0.18   | 21.11 ± 0.59     | 39    |  |

## References

- [1] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”.  
In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509.  
DOI: 10.1103/PhysRevLett.13.508.
- [2] F. Englert and R. Brout.  
“Broken Symmetry and the Mass of Gauge Vector Mesons”.  
In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323.  
DOI: 10.1103/PhysRevLett.13.321.
- [3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble.  
“Global Conservation Laws and Massless Particles”.  
In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587.  
DOI: 10.1103/PhysRevLett.13.585.
- [4] The ATLAS Collaboration.  
“The ATLAS Experiment at the CERN Large Hadron Collider”.  
In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003.  
DOI: 10.1088/1748-0221/3/08/s08003.
- [5] The CMS Collaboration. “The CMS experiment at the CERN LHC”.  
In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004.  
DOI: 10.1088/1748-0221/3/08/s08004.
- [6] O. Brüning, H. Burkhardt, and S. Myers. “The Large Hadron Collider”.  
In: *Progress in Particle and Nuclear Physics* 67.3 (2012), pp. 705–734.  
ISSN: 0146-6410. DOI: <https://doi.org/10.1016/j.pnnp.2012.03.001>.
- [7] The ATLAS collaboration. “A Particle Consistent with the Higgs Boson Observed with the ATLAS Detector at the Large Hadron Collider”.  
In: *Science* 338.6114 (Dec. 2012), pp. 1576–1582. ISSN: 0036-8075.  
DOI: 10.1126/science.1232005.
- [8] The CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”.  
In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693.  
DOI: <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [9] P.A. Zyla et al. “Review of Particle Physics”.  
In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

## References

- [10] Nicolas Berger, Claudia Bertella, Thomas P. Calvet, Milene Calvetti, Valerio Dao, Marco Delmastro, Michael Duehrssen-Debling, Paolo Francavilla, Yacine Haddad, Oleh Kivernyk, Jonathon M. Langford, Changqiao Li, Giovanni Marchiori, Predrag Milenovic, Carlo E. Pandini, Edward Scott, Frank J. Tackmann, Kerstin Tackmann, Lorenzo Viliani, and Meng Xiao. *Simplified Template Cross Sections - Stage 1.1*. 2019. arXiv: 1906.02754 [hep-ph].
- [11] David J. Griffiths. *Introduction to elementary particles*. Weinheim, 2008.
- [12] Mark Thomson. *Modern particle physics*. Cambridge, 2013.
- [13] A. Einstein.  
*Die Grundlage der allgemeinen Relativitätstheorie [AdP 49, 769 (1916)]*.  
Vol. 14. S1. 2005, pp. 517–571.  
DOI: <https://doi.org/10.1002/andp.200590044>.
- [14] Carsten Burgard and David Galbraith.  
*A standard diagram of the current standard model of physics*.  
URL: <https://texample.net/tikz/examples/model-physics/>.
- [15] Gerardus 't Hooft. “Renormalizable lagrangians for massive Yang-Mills fields”. In: *Nuclear physics: B* 35.1 (1971), pp. 167–188.
- [16] John Ellis. *Higgs Physics*. 2013. arXiv: 1312.5672 [hep-ph].
- [17] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. In: *CERN-2017-002-M* (2017). arXiv: 1610.07922 [hep-ph].
- [18] Adam Kaluza. “The Effective Lagrangian Morphing and the Measurement of Effective Higgs Boson Couplings in VBF  $H \rightarrow W^\pm W^\mp \rightarrow l^+ \bar{\nu}_l l'^+ \nu_{l'}$  at  $\sqrt{s} = 13$  TeV with the ATLAS Experiment”.  
PhD thesis. Johannes Gutenberg-Universität Mainz.
- [19] Rende Steerenberg, Markus Albert, Reyes Alemany-Fernández, Theodoros Argyropoulos, Enrico Bravin, Guy Crockford, Jean-Charles Dumont, Kajetan Fuchsberger, Rossano Giachino, Massimo Giovannozzi, et al. “Operation and performance of the CERN Large Hadron Collider during proton Run 2”. In: *10th Int. Particle Accelerator Conf. (IPAC'19), Melbourne, Australia, 19-24 May 2019*. JACOW Publishing, Geneva, Switzerland. 2019, pp. 504–507.
- [20] S. Myers and E. Picasso. “The design, construction and commissioning of the CERN Large Electron-Positron collider”. In: *Contemporary Physics* 31.6 (1990), pp. 387–403.  
DOI: 10.1080/00107519008213789.
- [21] The LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005.  
DOI: 10.1088/1748-0221/3/08/S08005.

## References

- [22] The ALICE Collaboration. In: 3.08 (Aug. 2008), S08002–S08002.  
DOI: 10.1088/1748-0221/3/08/s08002.
- [23] Boris Dolgoshein. “Transition radiation detectors”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 326.3 (1993), pp. 434–469.
- [24] The ATLAS collaboration.  
“Operation of the ATLAS trigger system in Run 2”.  
In: 15.10 (Oct. 2020), P10004–P10004.  
DOI: 10.1088/1748-0221/15/10/p10004.
- [25] The ATLAS Collaboration. *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*. Tech. rep. Geneva: CERN, Mar. 2012. URL: <https://cds.cern.ch/record/1435196>.
- [26] The ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”.  
In: *The European Physical Journal C* 77.5 (2017), pp. 1–35.
- [27] The ATLAS Collaboration.  
“Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data”.  
In: 14.12 (Dec. 2019), P12006–P12006.  
DOI: 10.1088/1748-0221/14/12/p12006.
- [28] The ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”.  
In: *The European Physical Journal C* 77.7 (2017), pp. 1–73.
- [29] Toni Baroncelli, Matthew Joseph Basso, Michela Biglietti, Volker Buescher, Carsten Daniel Burgard, Maria Teresa Camerlingo, Biagio Di Micco, Roberto Di Nardo, Ada Farilla, Ralf Gugel, Fudong He, Yesenia Hernandez Jimenez, Moritz Elias Hespig, Thomas Gerhard Honig, Jianbei Liu, Yanwen Liu, Kristin Lohwasser, Sana Ketabchi, Karsten Koeneke, Peter Krieger, Heng Li, Yun-Ju Lu, Robert Orr, Rustem Ospanov, Christian Schmitt, David Richard Shope, Jonas Strandberg, Natalie Wieseotte, Benedict Tobias Winter, and Xiao Yang. *Measurement of the Higgs boson production cross section in association with a vector boson and decaying into  $WW^*$  with the ATLAS detector at  $\sqrt{s} = 13$  TeV*. Tech. rep. Geneva: CERN, May 2021. URL: <https://cds.cern.ch/record/2765898>.
- [30] The ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV”.  
In: *Eur. Phys. J. C* 76.5 (2016), p. 292.  
DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

## References

- [31] The ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”.  
In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052.  
DOI: 10.1140/epjc/s10052-017-5031-2.  
URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5031-2>.
- [32] Matteo Cacciari, Gavin P Salam, and Gregory Soyez.  
“The anti-ktjet clustering algorithm”. In: 2008.04 (Apr. 2008), pp. 063–063.  
DOI: 10.1088/1126-6708/2008/04/063.
- [33] The ATLAS Collaboration.  
“Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at  $\sqrt{s} = 13$  TeV”.  
In: *The European Physical Journal C* 78.11 (Nov. 2018). ISSN: 1434-6052.  
DOI: 10.1140/epjc/s10052-018-6288-9.
- [34] Olaf Behnke, Kevin Kröninger, Grégory Schott, and Thomas Schörner-Sadenius.  
*Data analysis in high energy physics: a practical guide to statistical methods*. John Wiley & Sons, 2013.
- [35] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells.  
“Asymptotic formulae for likelihood-based tests of new physics”.  
In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052.  
DOI: 10.1140/epjc/s10052-011-1554-0.
- [36] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*.  
<http://www.deeplearningbook.org>. MIT Press, 2016.
- [37] Kurt Hornik, Maxwell Stinchcombe, and Halbert White.  
“Multilayer feedforward networks are universal approximators”.  
In: *Neural Networks* 2.5 (1989), pp. 359–366. ISSN: 0893-6080.  
DOI: [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- [38] Prajit Ramachandran, Barret Zoph, and Quoc V. Le.  
*Searching for Activation Functions*. 2017. arXiv: 1710.05941 [cs.NE].
- [39] Diederik P. Kingma and Jimmy Ba.  
*Adam: A Method for Stochastic Optimization*. 2017.  
arXiv: 1412.6980 [cs.LG].
- [40] Natalie Wieseotte. “Higgs Boson Cross Section Measurement in the  $pp \rightarrow WH \rightarrow WWW \rightarrow l\nu l\nu l\nu$  Channel at  $\sqrt{s}=13$  TeV with the ATLAS Experiment”. PhD thesis. Johannes Gutenberg-Universität Mainz.
- [41] *Simplified Template Cross Sections*. URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGFiducialAndSTXS>.
- [42] “Measurement of the  $t\bar{t}$  production cross-section in the lepton+jets channel at  $\sqrt{s} = 13$  TeV with the ATLAS experiment”.  
In: *Physics Letters B* 810 (2020), p. 135797. ISSN: 0370-2693.  
DOI: <https://doi.org/10.1016/j.physletb.2020.135797>.

## References

- [43] The ATLAS Collaboration. “Measurement of the production cross section for a Higgs boson in association with a vector boson in the  $H \rightarrow WW^* \rightarrow l\nu l\nu$  channel in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *Physics Letters B* 798 (2019), p. 134949. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2019.134949>.
- [44] The ATLAS Collaboration. “Observation and measurement of Higgs boson decays to  $WW^*$  with the ATLAS detector”. In: *Physical Review D* 92.1 (July 2015). ISSN: 1550-2368. DOI: [10.1103/physrevd.92.012006](https://doi.org/10.1103/physrevd.92.012006).
- [45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [46] Warren Sarle. *Should I normalize/standardize/rescale the data?* URL: <http://www.faqs.org/faqs/ai-faq/neural-nets/part2/section-16.html>.
- [47] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. URL: <https://www.tensorflow.org/>.
- [48] Thomas Gerhard Honig. “Higgs Boson Cross Section Measurement in the W-Associated Production and Fully Leptonic Final Stat at  $\sqrt{s} = 13$  TeV with the ATLAS Experiment using Deep Neural Networks”. 2020. URL: <https://cds.cern.ch/record/2752498>.
- [49] Paolo Nason and Carlo Oleari. “NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG”. In: *Journal of High Energy Physics* 2010.2 (Feb. 2010). ISSN: 1029-8479. DOI: [10.1007/jhep02\(2010\)037](https://doi.org/10.1007/jhep02(2010)037).
- [50] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: [10.1007/jhep07\(2014\)079](https://doi.org/10.1007/jhep07(2014)079).

## References

- [51] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands.  
“A brief introduction to PYTHIA 8.1”.  
In: *Computer Physics Communications* 178.11 (2008), pp. 852–867.  
ISSN: 0010-4655. DOI: <https://doi.org/10.1016/j.cpc.2008.01.036>.
- [52] T Gleisberg, S Höche, F Krauss, M Schönherr, S Schumann, F Siegert, and J Winter. “Event generation with SHERPA 1.1”.  
In: *Journal of High Energy Physics* 2009.02 (Feb. 2009), pp. 007–007.  
ISSN: 1029-8479. DOI: [10.1088/1126-6708/2009/02/007](https://doi.org/10.1088/1126-6708/2009/02/007).
- [53] *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Tech. rep. Geneva: CERN, July 2017.  
URL: <http://cds.cern.ch/record/2273281>.
- [54] *IsolationSelectionTool*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/IsolationSelectionTool#Leptons>.
- [55] Lukas Heinrich, Matthew Feickert, and Giordon Stark. *pyhf: v0.6.3*.  
Version 0.6.3. DOI: [10.5281/zenodo.1169739](https://doi.org/10.5281/zenodo.1169739).
- [56] *Observation of WWW production in pp collisions at  $\sqrt{s}=13$  TeV with the ATLAS detector*. Tech. rep. Geneva: CERN, Aug. 2021.  
URL: <https://cds.cern.ch/record/2779173>.
- [57] *Measurement of Higgs boson production in association with a W or Z boson in the  $H \rightarrow WW$  decay channel*. Tech. rep. Geneva: CERN, 2021.  
URL: <https://cds.cern.ch/record/2758367>.
- [58] Magdalena Slawinska.  
“High-luminosity LHC prospects with the upgraded ATLAS detector”.  
In: *arXiv preprint arXiv:1609.08434* (2016).