Optimization of a Regression-Based Deep Neural Network for Di-Tau Mass Reconstruction in ATLAS

Lena Maria Herrmann

Masterarbeit in Physik
angefertigt im Physikalischen Institut

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

Dezember 2020

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...

Date Signature

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

Contents

CHAPTER 1

Introduction

In Particle Physics, the fundamental principles, orders and relations are described by the Standard Model (SM). According to this theory, the Higgs mechanism describes the mass acquirement for the gauge bosons and also guarantees the consistency of the model. The Higgs boson, which is associated to this mechanism, was discovered in 2012 at the Large Hadron Collider (LHC) at CERN, where beams of protons are made to collide. The resulting reaction products are measured by detectors like ATLAS. The Higgs boson is an element that had been missing for a long time and now serves as a huge verification of the theory. Those days' news, such as the following *ScienceNews* quote from July 2012, express the excitement about the discovery [1]:

> "For more than three decades, the Higgs has been physicists' version of King Arthur's Holy Grail, Ponce de Leon's Fountain of Youth, Captain Ahab's Moby Dick. It's been an obsession, a fixation, an addiction to an idea that almost every expert believed just had to be true."

With the discovery, new questions arose mainly probing the consistency of the SM. Quantities like production and decay rates of the Higgs as well as the coupling strengths to the elementary particles had to be determined experimentally and compared to the expectations. One sector of interest is the coupling of the Higgs to the leptons. With a branching ratio of roughly 6% [2], the $H \rightarrow \tau^+\tau^-$ decay is the most suitable channel for this investigation. However, the desired signal events are dominated by Z decays with the same final state. They represent the predominant background of the analysis. In order to provide an adequate separation of the two resonances, reliable di-tau mass reconstruction techniques are crucial. However, neutrinos are involved in the decay kinematics of the tau leptons. They are exclusively weakly interacting particles and traverse the detector unmeasured, which complicates the event reconstruction. Presently, the task of di-tau mass estimation is performed by the Missing Mass Calculator (MMC) [3], which requires many computational resources. Its concept is based on the maximization of likelihood products in order to estimate the most probable solution for the reconstruction. In this thesis, alternative approaches using artificial neural networks (ANNs) will be investigated. They comprise computing systems inspired by the human brain which in turn implies the ability to learn. ANNs infer their knowledge from a set of input data by recognizing patterns and structures. In order to do so, depending on the task, thousands of internal parameters are adapted in the learning process. Nowadays, ANNs are at the core of everyday applications like image recognition or spam filters. They also have distinguished themselves in science. However, their main field of

application covers classification tasks. In this thesis, though, a contribution to the ascending sector of regression models predicting continuous variables will be provided.
Initial studies for the application of regression based neural networks in the di-tau mass reconstruction were already performed in reference [4]. In the presented approach, the di-tau mass was directly used as ANN output. This thesis, though, focuses on an estimation of the neutrino momenta and determines the di-tau mass based on their prediction. The advantage of the multidimensional ANN output is the possibility to incorporate physical constraints in the learning process, which will be the central aspect of investigation.

In chapter 2, an overview over the SM is given, focusing on the physical aspects relevant for the understanding of the thesis. Subsequently, in chapter 3, the ATLAS detector is introduced, whose simulated responses provide the data used in the following investigations. Additionally, the reconstruction techniques of different physical objects as well as the Monte Carlo (MC) based simulation tools of the particle reactions are presented. In chapter 4, the concepts of ANNs are introduced and methods for constraint realizations within the training process are discussed.
In the following chapter 5, the data samples used for the analysis as well as their preprocessing are addressed, so that a suitable training of ANNs can be performed. As basis of the following investigations, a fundamental setup, comprising a default set of hyperparameters and a certain architecture of the network, is presented in chapter 6. Furthermore, different performance measures are addressed which enable the rating and the comparability of different models. In order to realize this task, a calibration method is also developed which corrects biases in the di-tau mass estimation. Following these chapters of introductory content, chapter 7 focuses on target and setup related effects. Their proper understanding is crucial for a suitable di-tau mass reconstruction, which is addressed in chapter 8. Here, different setups are developed and compared regarding their performance on the training sample. Generally, the models can be assigned to two different categories, namely beyond or on par with the MMC. The latter requires the incorporation of physical constraints in the learning process, which is discussed in detail. In the following chapter 9, the developed models are evaluated, using a Z and a Higgs sample. In this context, the ability to separate signal and background is investigated by means of the overlap of the reconstructed mass peaks. Furthermore, the influence of physical properties of the data samples on the di-tau mass reconstruction is analyzed. In conclusion, in chapter 10, a summary of the final status of the investigation is given and an outlook for further studies is presented.

CHAPTER 2

The Standard Model

The foundation of modern particle physics is the Standard Model, describing the elementary particles and their interactions. Its mathematical description is based on a $SU(3)_c \otimes SU(2)_C \otimes U(1)_Y$ gauge theory comprising the strong, weak and electromagnetic force. This chapter gives a short introduction into the fundamental concepts of the theory. Furthermore, the focus lies on the physical aspects needed for the understanding of the following analysis. Reference [5] serves as source for this chapter, if not stated otherwise. It can also be used for further reading.

2.1 Overview

The Standard Model of particle physics is the theory of the elementary particles and forces, describing their interactions. For a certain time, it was assumed that the atom is the smallest building stone of matter, but with the realization of higher energy scales, further substructure became observable. In the eV regime it became apparent that electrons orbit around a nucleus, which in turn consists of protons and neutrons (MeV scale). However, further constituents of matter can be observed at energies in the GeV regime. Here, the substructure of the nucleons, the quarks, can be resolved.
According to the current understanding, the visible matter of the universe consists of the SM fermions (spin-$1/2$ particles), the quarks and the leptons (see figure 2.1). Each of these 12 particles has an antiparticle of opposite charge. Quarks and leptons can be differentiated by characterizing properties which enables their participation in certain interactions as described in the following paragraph.

The Fundamental Forces

The whole theory comprises three forces which differ in their relative strengths. In rising order, the *weak*, the *electromagnetic* and the *strong force* can be differentiated. They are mediated by gauge bosons which are spin-1 particles. In order to participate in an interaction, a particle has to couple to the force-carrying boson. For this, it needs to carry the corresponding charge.
The charge of the weak force is the weak isospin, which can be assigned to every SM particle. Consequently, all fermions can be involved in weak interactions, which are mediated by $W^{\pm}$ and Z^0 bosons. The "weakness" of the force is attributed to the high masses of the force carrying bosons. As an example for weak interactions, the β-decay of radioactive material can be considered.
The electromagnetic force, though, is reserved for the electrically charged particles and is mediated

by the photon. Within an atom, the addressed force binds the negatively charged electrons to the oppositely charged nucleus.
The third force is the strong interaction ensuring the binding of protons and neutrons within the nucleus. In order to interact strongly, colour charge is necessary, which is carried by the quarks. The exchange particles are the massless gluons.

The Quarks

The class of the quarks is comprised of the up (u), down (d), charm (c), strange (s), top (t) and bottom (b) quark. They take part in all three particle interactions, but their outstanding characteristic is the color they carry. Thus, they are able to interact strongly. In nature, no colored objects can be observed but their confined singlet states. Consequently, quarks are confined by the strong force to so called *hadrons*. A differentiation is made between *mesons*, *baryons* and *antibaryons* which consist of quark-antiquark ($q\bar{q}$) pairs, three quarks (qqq) or three antiquarks ($\bar{q}\bar{q}\bar{q}$) which all form colour neutral states.

The Leptons

The electron e, muon μ and tau τ as well as their corresponding neutrinos belong to the class of the leptons. In contrast to the quarks, they carry no color charge. Hence, they can be observed as free particles. Whereas the e, the μ and the τ are electrically charged, the neutrinos are not. Furthermore, in this thesis, the neutrinos are assumed to be massless. However, an upper mass bound can be deduced from neutrino oscillations [5]. Neutrinos are exclusively weakly interacting particles. Thus, they are not detectable in non-specialized detector designs.

Within the quarks and leptons a further categorization into so called *generations* can be performed. Each generation corresponds to one of the columns of figure 2.1. The different generations copy the particle properties of the previous one except for the mass.
The remaining element of the overview is the Higgs boson which is associated to the Higgs field. It is the central part of the Higgs mechanism which describes the mass acquirement of the gauge bosons. A more detailed explanation is given in section 2.5.

2.2 Particle and Interaction Related Properties

In this section, the concepts of parity and helicity are introduced. They are fundamental properties and influence the decay behavior and polarisation state of particles, as described in section 2.7.

2.2.1 Parity

The parity of a physical object can be used as classifying property. The operation describes a spatial inversion which can be expressed quantum mechanically as [5]

$$\Psi(x,t) \rightarrow \Psi^{'}(x,t) = \hat{P}\Psi(x,t) = \Psi(-x,t).$$

While vectors like the momentum change sign under parity inversion, axial vectors, which are the cross product of two vectors, do not. Experiments have shown that the weak charged-current

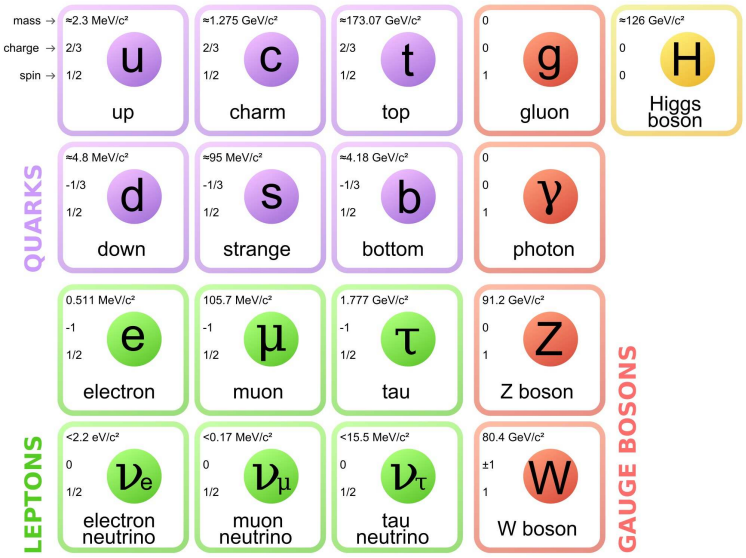


Figure 2.1: The Standard Model particles. Taken from [6].

interaction is maximally parity violating and mediated between fermions, which differ in their electrical charge by one unit. It is the only flavor changing interaction and therefore an important part of particle decays.
In order to guarantee a maximal parity violation, the coupling to vectors (g_V) **and** axialvectors (g_A) has to exist in equal strength. It was shown that

$$|g_V| = -|g_A|$$

for weak interactions. Therefore, it is called the *V-A* structure.
The electromagnetic interaction is solely vectorial, so that parity is conserved.

2.2.2 Helicity & Chirality

The helicity is a particle property defined as the normalized spin projection in the direction of flight [7]

$$h = \frac{\vec{s}\vec{p}}{|\vec{s}||\vec{p}|}. \tag{2.1}$$

For spin-1/2 fermions, the two possible helicity states are $\pm\frac{1}{2}$ which assign the so called handedness. While particles with the spin parallel to the direction of flight are called *right-handed* (RH), the ones with opposite orientation are *left-handed* (LH). The V-A structure of the weak interaction implies only couplings to left-handed fermions and right-handed antifermions. In contrast, the electromagnetic interaction couples equally to both of them.
Since helicity is no Lorentz invariant concept, chirality is introduced for massive particles. It can be described by the eigenstates of the γ^5 matrix which result from Dirac matrix multiplications. In high-energy interactions (E $\gg$ m) both quantities are equivalent.

2.3 Electroweak Unification

In the theoretical elaboration of a theory of the particles and their interactions, one of the main aims has always been the idea of an overall unifying picture. One big achievement in this domain was the development of the electroweak unification from Glashow, Salam and Weinberg (GSW) which was awarded with the Nobel Prize in 1979 [8]. According to the theory, weak charged-current interactions refer to the $SU(2)_L$ symmetry group where the index L denotes the exclusive coupling to left-handed chiral particles or right-handed chiral antiparticles. In a consistent theory, a local gauge invariance is required since it ensures renormalisability. In order to satisfy this aspect, three gauge fields with associated gauge bosons $W^{(1)}, W^{(2)}$ and $W^{(3)}$ are introduced for the weak interaction. While $W^{(1)}$ and $W^{(2)}$ are associated with the W^+ and the W^- boson,

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \tag{2.2}$$

the correspondence of $W^{(3)}$ and the Z boson as mediator of neutral-currents has to be excluded since experiments proved the coupling of the Z boson to left **and** right handed chiral states. From Quantum Electro Dynamics (QED) another neutral boson, namely the photon is known. In the context of electroweak unification, a local $U(1)_Y$ gauge symmetry is introduced which takes the role of the $U(1)$ gauge symmetry of electromagnetism. The index Y denotes the coupling of the new gauge field B_μ to the weak hypercharge Y. In the GSW theory, the Z field Z_μ as well as the photon field A_μ consist of a mixture of the new field and $W^{(3)}$. This mixture is defined by the weak mixing angle θ_W [5].

$$A_\mu = +B_\mu \cos(\theta_W) + W_\mu^{(3)} \sin(\theta_W) \tag{2.3}$$

$$Z_\mu = -B_\mu \sin(\theta_W) + W_\mu^{(3)} \cos(\theta_W) \tag{2.4}$$

The measurement of this angle can be used as a precision test of the SM. It also defines the coupling of the Z boson to left and right-handed chiral states [5]

$$c_L = I_W^{(3)} - Q \sin^2\theta_W \qquad \text{and} \qquad c_R = -Q \sin^2\theta_W \tag{2.5}$$

with the electromagnetic charge Q and the third component of the weak isospin $I_W^{(3)}$. Due to the inequality of c_L and c_R polarisations occur in Z decays (see section 2.7.1).

2.4 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is described by a non-abelian SU(3) local gauge symmetry and represents the field theory of the strong interaction. The eight generators of the symmetry can be associated to the eight gluons forming an octet of colored states. The color combinations of red, blue, green and the corresponding anti-colors read [5]

$$r\bar{g}, \quad g\bar{r}, \quad r\bar{b}, \quad b\bar{r}, \quad g\bar{b}, \quad b\bar{g}, \quad \frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g}) \quad \text{and} \quad \frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b}).$$

An important property of colored objects is their confinement to color neutral objects at low energies. However, the coupling constant α_s of QCD is not constant, but rather a running variable which becomes sufficiently small at high energies. This leads to the concept of asymptotic freedom in this regime. In hadron hadron collisions, the event signatures are characterized by hadronic jets, which directly result from the peculiarities of strong interactions and the concept of *hadronisation*. The fact that gluons carry color themselves leads to attractive interactions which squeezes the color field between two quarks to a tube. When an initially produced quark and antiquark separate, the energy of this field increases up to the point where it is energetically favorable to create a new $q\bar{q}$ pair and to separate the field into smaller strings. This procedure continues and with decreasing energy, the produced quarks and antiquarks will combine to colorneutral hadrons.

2.5 Higgs Mechanism

The Higgs mechanism is one of the fundamental concepts of the SM theory. It preserves the gauge invariance of the theory which would otherwise be violated by the boson masses associated to the electroweak interaction [5]. Coupling to the Higgs field produces the masses of the SM particles. The larger the coupling strength, the larger the mass of the particle. The Higgs boson itself can be understood as excitation of the Higgs field. The scalar particle was discovered in 2012 ([9], [10]), which was a milestone for particle physics and the SM theory. The following overview describes the theory in rough strokes but might not be sufficient to understand it fully. For further reading see [5] and [11].

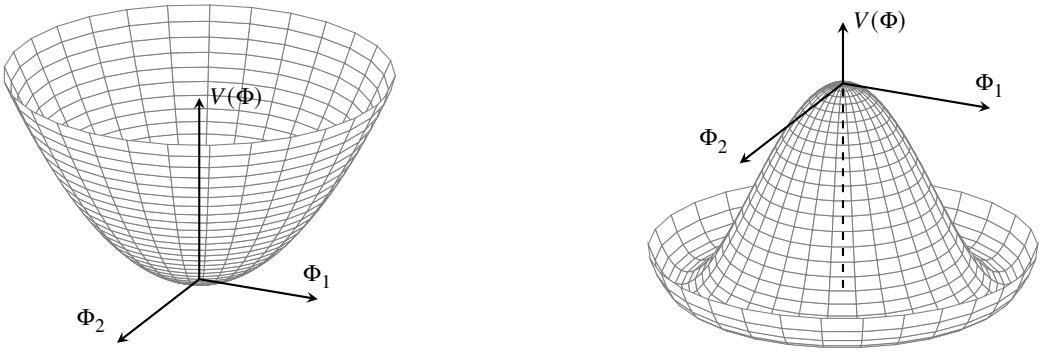
The Higgs mechanism is developed on the basis of the SM Lagrangian densities which can be used in order to describe the dynamics of a quantum field. Excitations of these fields represent particles. As described in section 2.3, the electroweak sector of particle physics follows a $SU(2)_L \times U(1)_Y$ local gauge symmetry. Breaking this symmetry generates the observed masses of the gauge bosons [5]. The assumed field is a complex scalar field with a potential $V(\Phi)$ of the form

$$V(\Phi) = \mu^2(\Phi^*\Phi) + \lambda(\Phi^*\Phi)^2 \quad \text{with} \quad \Phi = \frac{1}{\sqrt{2}}(\Phi_1 + i\Phi_2). \tag{2.6}$$

The visualization of the potential for $\lambda > 0$ is shown in figure 2.2. Whereas figure 2.2(a) represents a positive choice of μ^2, figure 2.2(b) assumes $\mu^2 < 0$. These two cases lead to different field minima, and therefore to different vacuum states (state of lowest energy). While for 2.2(a), the minimum is centered at zero ($\Phi_1 = \Phi_2 = 0$), the potential in 2.2(b) becomes minimal for all points around the hollow. As soon as one specific vacuum state is picked out of the possible solutions of $\Phi_1^2 + \Phi_2^2 = \frac{-\mu^2}{\lambda}$, the global symmetry is broken. The field excitation in this specific direction is associated with a massive scalar particle. For a weak isospin doublet, containing two scalar fields and thus having four degrees of freedom,

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi_1 + i\Phi_2 \\ \Phi_3 + i\Phi_4 \end{pmatrix} \tag{2.7}$$

this particle can be associated to the Higgs boson [5].



(a) $\mu^2 > 0$ (b) $\mu^2 < 0$

Figure 2.2: Potential of the form $V(\Phi) = \mu^2(\Phi^*\Phi) + \lambda(\Phi^*\Phi)^2$ with a complex scalar field $\Phi = {}^1/\sqrt{2}(\Phi_1 + i\Phi_2)$ for $\mu^2 > 0$ (a) and $\mu^2 < 0$ (b). Inspiration from [5], produced with [12].

In order to understand the mass generation of the gauge bosons, the derivatives of the Lagrangian

$$\mathcal{L} = (\partial_\mu \Phi)^\dagger (\partial^\mu \Phi) - V(\Phi) \tag{2.8}$$

have to be considered in the unitary gauge [5]. From the terms which are quadratic in the gauge boson fields, the masses of the associated particles can be determined. The term responsible for the W-boson reads [5]

$$\frac{1}{8} v^2 g_W^2 \left(W_\mu^{(1)} W^{(1)\mu} + W_\mu^{(2)} W^{(2)\mu} \right) \tag{2.9}$$

with the vacuum expectation value v and the coupling constant of the $SU(2)_L$ interaction g_w. This implies a W boson mass of

$$m_W = \frac{1}{2} g_W v. \tag{2.10}$$

Analogously, the masses of the photon and the Z boson can be received by considering the Lagrangian terms, which are quadratic in $W^{(3)}$ and B_μ [5]

$$\frac{v^2}{8} \begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix} \underbrace{\begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix}}_{\mathbf{M}} \begin{pmatrix} W_\mu^{(3)} \\ B_\mu \end{pmatrix}. \tag{2.11}$$

However, in order to receive the physical boson fields, a diagonalised mass matrix $\mathbf{M}$ has to be ensured. After the diagonalisation, a massless photon with $m_A = 0$ and a Z boson mass of [5]

$$m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2} \tag{2.12}$$

can be received with the coupling g' to the $U(1)_Y$ gauge symmetry.

Fermions acquire their mass through the coupling to the Higgs field. Whereas right-handed fermions are represented by SU(2) singlets (R), left-handed fermions are described in SU(2) doublets (L). The expressions

$$-g_f\,(\bar{L}\phi R + \bar{R}\phi^\dagger L) \qquad \text{and} \qquad g_f\,(\bar{L}\phi_C R + \bar{R}\phi_C^\dagger L) \tag{2.13}$$

are invariant under $SU(2)_L$ and $U(1)_Y$ gauge transformations and can therefore be considered in the SM theory. The factor g_f is the so called Yukawa coupling which describes the coupling between a scalar field ϕ and a Dirac field, associated to the fermions. The fermion masses are related to the Yukawa coupling by the following expression [5]

$$g_f = \sqrt{2}\frac{m_f}{v}. \tag{2.14}$$

2.6 Conservation Laws

In colliding beam experiments based on hadron beams, the final state of a collision is characterized by an immense amount of particles resulting from underlying events, parton showers, hadronization etc. The processes taking place follow the conservation laws of particle physics. The following quantities have to be conserved, independently on the underlying interaction:

- energy
- charge
- spin
- baryon number
- lepton (family) number

There are also other conservation laws and aspects that are solely violated by specific interactions. For further information see [5]. The focus of this section lies on the lepton number conservation, since it strongly affects the event signature of the tau decays which are of particular interest for this thesis. While a lepton receives the lepton number 1, the corresponding antiparticle is assigned with a -1. Apart from the total lepton number, the lepton family number has to be conserved as well. This means that the lepton number within one generation has to stay the same in a reaction.
In the example of a single decaying tau lepton, the initial family lepton number, which also corresponds to the total lepton number, is $L = L_\tau = +1$. Hence, in order to ensure $L_\tau = +1$ in the final state as well, tau neutrinos ν_τ have to be involved since no other tau leptons can be produced in a tau decay due to energy conservation.

2.7 The Tau-Lepton

With a mass of $m_\tau = (1\,776.82 \pm 0.16)$ MeV [13], the tau lepton is the heaviest known lepton, which involves some unique characteristics like a strong coupling to the Higgs field. Therefore, it is suitable in order to probe the theory behind this mechanism. The pie chart of figure 2.3 shows the decay

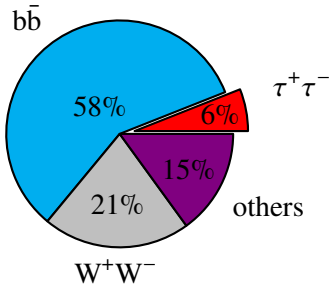


Figure 2.3: Higgs decay branching fractions [12]. With 6%, the di-tau channel is suitable in order to probe the Higgs coupling to leptons. Numbers taken from [2].

Decay mode		Branching ratio [%]
hadronic		**64.76**
	1 prong	**46.7**
	`1p0n`	10.83
	`1p1n`	25.52
	`1pXn`	10.35
	3 prong	**11.69**
	`3p0n`	8.99
	`3pXn`	2.7
	others	**6.37**
leptonic		**35.24**
	$e^-\bar{\nu}_e\nu_\tau$	17.83
	$\mu^-\bar{\nu}_\mu\nu_\tau$	17.41

Table 2.1: Tau decay branching ratios, data taken from [14]. The leptonic final states of the τ^+ can be realized by a charge conjugation. In the decay mode description, p and n denote $\pi^\pm$ and π^0 respectively. X is the number of neutral pions. For 1-prong events $X > 1$ and for 3-prong events $X > 0$. In hadronic decays, kaons are involved as well, but with a low percentage (summarized in others).

branching fractions of the Higgs boson. The di-tau channel accounts for 6% of the decays, which is a non-negligible contribution.
A second consequence of the high mass is the short lifetime of (290.3 ± 0.5) fs [13], which causes decay lengths of a few micrometres. Thus, only the decay products of the tau can be observed in a detector. A distinction can be made between leptonic (35%) and hadronic decays (65%) [14] which are dominated by pions and a further evolution to hadronic jets. In the following, the decay characteristic of the tau will be denoted by τ_{lep} and τ_{had} respectively. Generally, "1-prong" and "3-prong" decays are differentiated, which denote one or three charged pions. As mentioned in section 2.6, all tau decays involve neutrinos. An overview of the possible decay modes and their relative occurrences are listed in table 2.1.

2.7.1 Polarisation

The polarisation asymmetry of tau leptons influences their decay kinematics significantly (see section 2.7.2). It is defined as [15]

$$\mathcal{A}_{\text{pol}} = \frac{\sigma_L - \sigma_R}{\sigma_L + \sigma_R} \tag{2.15}$$

with σ_L and σ_R being the production cross sections of left-handed and right-handed leptons respectively. The cross section is a probability measure for a process to occur. In this analysis, di-tau states from three different sources will be considered, namely from a virtual photon γ^*, an interference of Z and γ^* and a Higgs boson. Since the photon couples equally to left and right-handed fermions, no polarisation is expected in this case. Additionally, no polarisation occurs for the taus originating from the scalar Higgs boson. However, this does not apply for the Z boson with its different coupling strengths for left- and right handed fermions as described in section 2.3.

2.7.2 Decay Kinematics

In a particle decay, the conservation laws of physics have to be fulfilled. One constraining aspect is the invariant mass of the mother particle which needs to be reproduced by the squared sum of the four-momenta of the decay products. In a tau lepton decay, the visible contributions (P_{vis}) which comprise the charged (P_c) and neutral (P_n) decay products are distinguished from the unmeasured neutrino four-momentum (P_ν).

$$m_\tau = 1.77\,\text{GeV} = \sqrt{(P_{\text{vis}} + P_\nu)^2} = \sqrt{(P_c + P_n + P_\nu)^2}. \tag{2.16}$$

However, certain relations hold between the decay products. One of them is induced by the spin of the tau lepton which will be exemplary discussed for a 1-prong decay like $\tau^- \rightarrow \pi^- \nu_\tau$. Since the pion does not carry any spin, the neutrino copies the spin characteristics of the tau lepton. In nature, only left-handed neutrinos and right-handed antineutrinos occur. Therefore, they have to be emitted opposite to the tau spin direction. The angular decay distribution is described in the center of mass frame by [16]

$$\frac{1}{\Gamma_\pi}\frac{d\Gamma_\pi}{d\cos\theta} = \frac{1}{2}(1 + \mathcal{A}_{\text{pol}}\cos\theta), \tag{2.17}$$

and is relevant in many respects. First of all, the polarisation influences the flight direction of the undetected neutrinos, which might be a useful information in their four-momenta estimation. Secondly, the information can also directly be used in the separation of Higgs from Z events due to their different spins. Furthermore, the pion energy correlates with the tau polarisation [17]. If the meson is emitted in the direction of flight of its mother particle, high energies can be expected because of the tau lepton boost.

CHAPTER 3

LHC & ATLAS: The Experiment

In order to probe the theory of the SM, particle reactions in the TeV regime need to be performed. For the realization of this immense task, thousands of people from a variety of countries work together at the largest particle accelerator in the world, the LHC. It accommodates four large detectors, as well as smaller experiments, all with the aim to increase the understanding of particle interactions. The analysis of the reactions constitutes a large challenge since, especially in hadron hadron collisions, lots of processes occur simultaneously. In order to disentangle the huge amount of information and to reconstruct the full event from the detector responses, suitable reconstruction algorithms are required. Furthermore, simulations of the expectations have to be produced for cross-checking with the actual measurement.
The di-tau mass reconstruction technique, developed in this thesis, is based on MC simulations of proton-proton collisions (pp collisions) recorded by the ATLAS detector at the LHC. For the comprehension of the analysis, it is crucial to understand the underlying concepts of the experiment, the detector designs of ATLAS, the generation of the data samples as well as the common event reconstruction procedures which transform the detector responses in physical analysable events. In order to provide the necessary knowledge, these aspects will be introduced in the following.

3.1 Concepts of Colliding Beam Experiments

In colliding beam experiments, two counter-running beams of particles are made to collide. Because of the opposite beam momenta ($\vec{p}_1 = -\vec{p}_2$), center-of-mass energies $\sqrt{s}$ which are twice as large as the respective beam energies ($E_1 = E_2$ assumed) are reached [5]

$$s = \left(\sum_i E_i\right)^2 - \left(\sum_i p_i\right)^2 . \tag{3.1}$$

This is especially important in the search for new particles, since the sum of the particle masses being produced is limited by the center-of-mass energy in the collision. Furthermore, the event production rates of specific processes are of interest in order to reach sufficient statistics for significant results. This rate is controllable by the beam parameters which are combined in an accelerator specific quantity, the instantaneous luminosity $\mathcal{L}$. It comprises the number of particles per bunch n, the frequency of

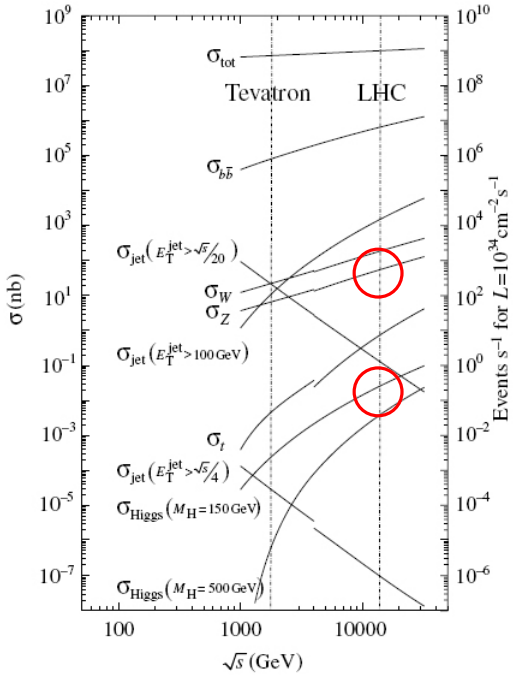


Figure 3.1: Cross sections of different processes in dependence on the center-of-mass energy [18]. In this thesis, the discrimination of Higgs and Z events is of specific interest. Therefore, their cross-sections at the design center-of-mass energy of the LHC are marked by red circles.

bunch crossings f and the root-mean-square beam measures σ_x and σ_y [5]

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}. \tag{3.2}$$

The integration over the operation time of the accelerator, combined with the cross section σ of the desired process which can be considered as the reaction probability, leads to the expected number of events N [5]

$$N = \sigma \int \mathcal{L}(t)dt. \tag{3.3}$$

Figure 3.1 shows the cross sections of different processes as a function of the center-of-mass energy for a fixed luminosity of $10 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Since the Higgs and the Z bosons are of particular interest in this thesis, their cross-sections are marked for the initially aimed LHC center-of-mass energy of 14 TeV. Their cross-sections differ by three orders of magnitude. Considering the branching fractions of 3.3% [13] and 6% [2] for the $Z \to \tau\tau$ and $H \to \tau\tau$ decay results in roughly 500 background events for 1 signal event in the ditau final state. However, it has to be considered that the figure assumes a Higgs mass of 150 GeV and a higher center-of-mass energy than is actually reached. Nevertheless, the discrimination of Higgs and Z events has to be efficient, in order to separate the signal from the predominant background.

3.2 CERN Accelerator Complex

At the LHC at CERN (**C**onseil **e**uropéen pour la **r**echerche **n**ucléaire), beams of protons or heavy nuclei with energies of several TeV are made to collide in order to learn more about the open questions of particle physics. The design luminosity for proton collisions of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [19] was reached in 2016 and is planned to reach $\mathcal{L} = 10 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ in future runs [20]. The beam energy was also increased from 3.5 TeV in 2010 [21] to 6.5 TeV in 2015 [22].
The beam particles reach their immense velocities near the speed of light by passing different acceleration steps, as depicted in figure 3.2. A more detailed description of the single stations can be found in reference [23]. The maximum energy is finally reached in the 27 km long tunnel of the LHC where, during Run 2, bunches of 10^{11} particles were made to collide every 25 ns [24] at four locations, where particle detectors, developed for different purposes, are placed.
ATLAS and CMS (**C**ompact **M**uon **S**olenoid) are multi-purpose detectors including the data taking for the general Higgs analysis and the search for dark matter particles. They differ in their design and especially in their respective magnet systems. The other two detectors are designed for more specialized tasks. While ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment) aims to investigate quark-gluon plasma, the LHCb (**L**arge **H**adron **C**ollider **b**eauty experiment) deals with the topic of matter and anti-matter on the basis of particles containing b-quarks. Apart from these huge detectors, several smaller experiments are placed on the CERN site [25].

3.3 ATLAS

This thesis is based on simulations of events detected by ATLAS. Therefore, an introduction of the detector design, its functioning and basic event reconstruction techniques will be given.
Although it is a detector for general purposes, its design fulfills the Higgs analysis demands. Since all possible decay products of particle reactions need to be considered, the detector has to offer good electromagnetic as well as hadronic calorimeters in order to be sensitive to electrons and photons as well as jets. Because of the huge amount of particles originating from pp collisions, the trigger system needs to be efficient and the electronics should be fast and radiation hard. Furthermore, a good particle-identification is crucial because of the large QCD jet production rates which dominate the rare events of interest. The immense detector has a length of 44 m and a height of 25 m. Its weight is about 7 000 t [27]. An illustration of the ATLAS design and the arrangement of the individual constituents is depicted in figure 3.3.

3.3.1 The Spatial Description

In this section, the coordinate system and spatial definitions of ATLAS are introduced in order to enable a proper description of different event signatures.
The coordinate system is depicted in figure 3.4 and has its origin in the center of the detector. While the x-direction points towards the center of the LHC ring, the z-direction follows the beam orientation and the positive y-direction points upwards [27]. The polar orientation is described in terms of the pseudorapidity η because differences in η are not affected by longitudinal boosts. They are Lorentz invariant. While the pseudorapidity of particles along the beamline diverges, it is small for events going upwards (see figure 3.5). The mathematical dependence on the angle θ is expressed in equation 3.4 [5].

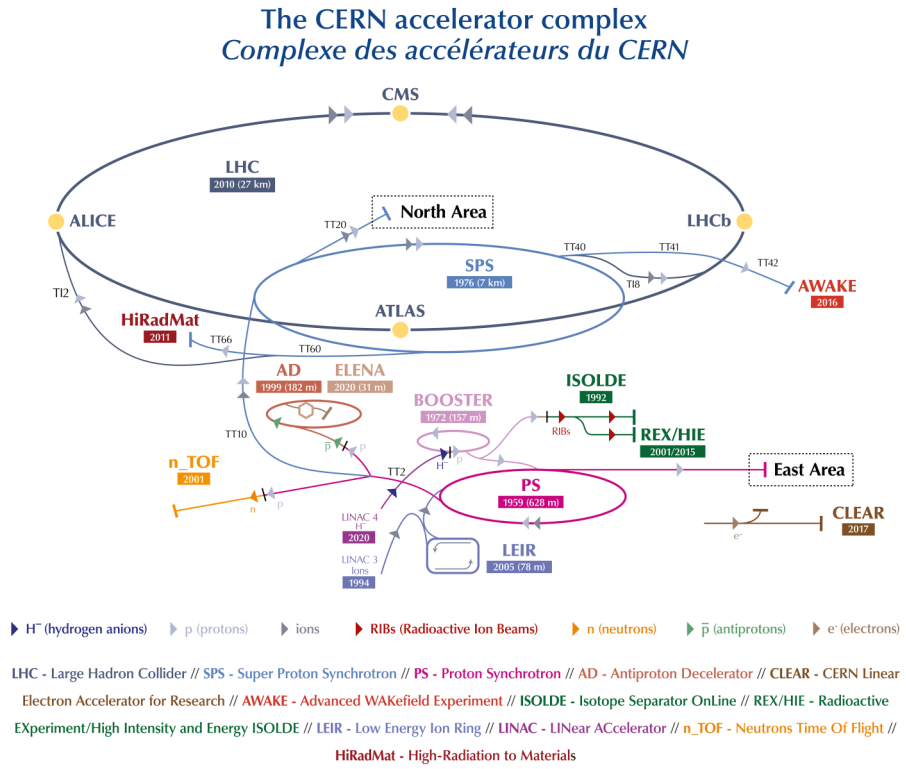


Figure 3.2: The CERN complex with its different stations for particle acceleration. Taken from [26]. The path of the protons colliding in ATLAS goes from Linac 4 (50 MeV reached) over the PSB (Booster) (1.4 GeV) to the PS (25 GeV) and SPS (450 GeV) until they are injected into the LHC (pushed to maximum energy of 6.5 TeV) [23].

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{3.4}$$

In combination with the azimuthal angle, the pseudorapidity forms the distance measure [27]

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{3.5}$$

Another important particle property is the transverse momentum p_T which corresponds to the momentum magnitude in the transverse plane

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{3.6}$$

Since mass terms can often be neglected in high energy physics, the sum of all reconstructed transverse momenta offers a valid expression for the total transverse energy E_T

$$E_T = \sum_i p_{T,i}. \tag{3.7}$$

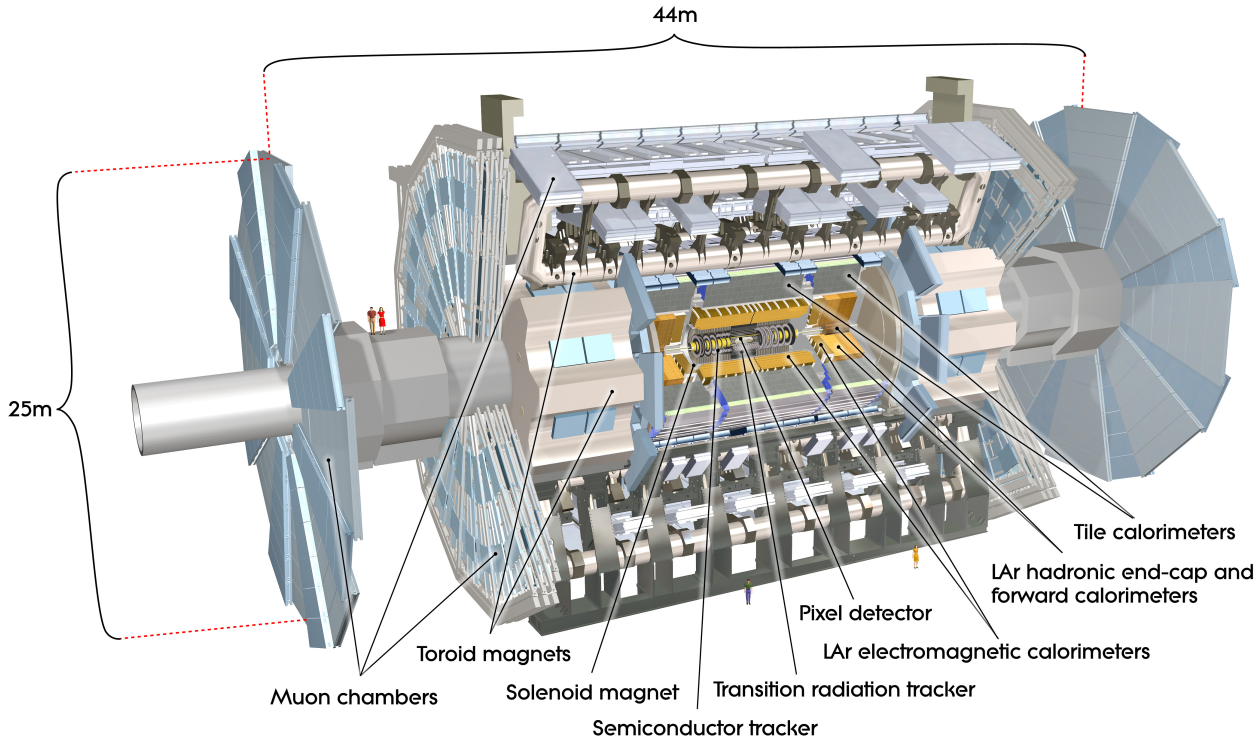


Figure 3.3: The ATLAS detector and its constituents.The whole setup is forward-backward symmetric. Taken from [28].

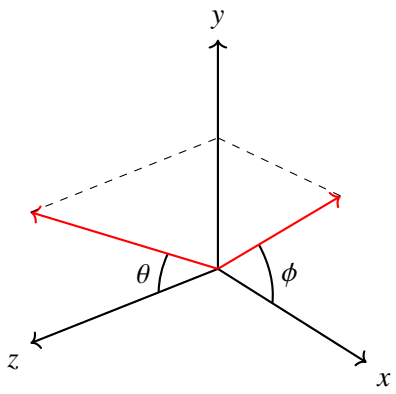


Figure 3.4: ATLAS coordinate system [12]. The x-direction points towards the LHC center while the z-direction is parallel to the beam orientation [27].

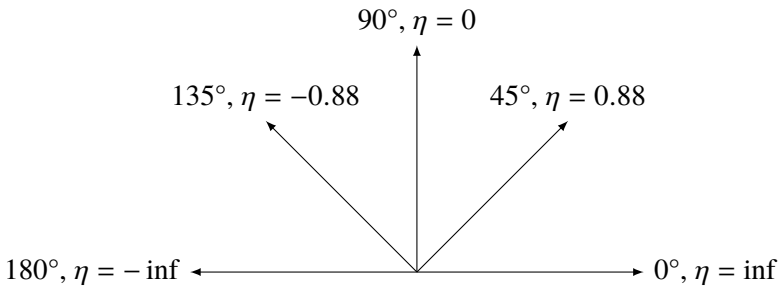


Figure 3.5: Spatial orientation of the pseudorapidity. The absolute pseudorapidity increases from 0 in y-direction to infinity along the beam axis. [12]

Corresponding to the coordinate system, the momentum can be expressed in terms of (p_T, η, ϕ) which is an alternative choice to the cartesian description (p_x, p_y, p_z)

$$\vec{p} = \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} p_t \cos(\phi) \\ p_t \sin(\phi) \\ p_t \sinh(\eta) \end{pmatrix}. \tag{3.8}$$

The ATLAS detector design is symmetrical in the azimuthal angle ϕ which is measured from the x-axis around the beam line.

3.3.2 The Design

The ATLAS detector comprises different constituents which are specialized for the detection of different physical properties like position or energy as well as different particles and their individual interaction types. The balanced design is necessary in order to enable a good particle identification. In this section, the main tasks and features of the main components, namely the tracking, the calorimeter and the muon system, will be shortly described. Furthermore, the trigger system, which controls the event processing, will be introduced.

The Tracking System

The tracking system is responsible for recording the trajectory of charged particles within the detector. They ionize the detector material and cause an electrical signal which is read out by the electronics. A magnetic field provokes a track curvature due to the Lorentz force. The following applies: The larger the radius of the circular movement, the larger the momentum of the particle. Because of the huge particle flux, a high detector granularity is necessary which is also needed for the vertex reconstruction.
The ATLAS tracking system is placed around the interaction point and consists of three parts.

- **Pixel Detector**
- **Semiconductor Tracker**
- **Transition Radiation Tracker**

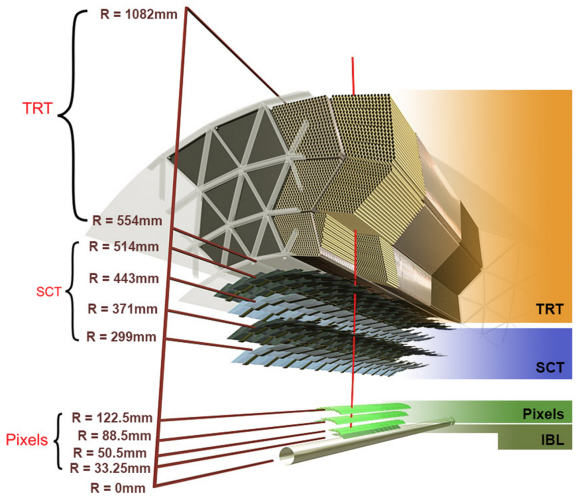


Figure 3.6: Transverse profile of the inner detector. It comprises pixel detectors around the beam line, semiconductor trackers as well as transition radiation trackers. Taken from [29].

While the Pixels and the Semiconductor Tracker (SCT) are responsible for precision tracking, the Transition Radiation Tracker (TRT) is mainly involved in the momentum measurement and the e^- – identification by the detection of transition radiation photons in the straw tubes. The unit is called the inner detector and follows the scheme of figure 3.6 which illustrates the transverse profile of the barrel region. The pixel detectors are arranged in three concentric cylinders around the beam axis and can also be found as transverse disks in the end-cap region. Together with the SCT, they cover a pseudorapidity range of $|\eta| < 2.5$ and are responsible for precision tracking with an accuracy of 10 µm (R–ϕ) and 15 µm (z) for the pixels and 17 µm (R–ϕ) and 580 µm (z) for the SCT in the barrel region respectively [27]. The latter is made up of strip layers with eight of them crossed by each track. The outer part of the inner detector forms the TRT which covers tracks with $|\eta| < 2.0$. Its straw tubes are oriented parallelly to the beam line in the barrel region and radially in the end-caps. The whole setup is exposed to a magnetic field of 2 T which originates from the central superconducting solenoid and enables the reconstruction of momentum and charge.

The Calorimeters

The calorimeter system is designed for electromagnetic and hadronic measurements in a region of $|\eta| < 4.9$. In the spatial region of the inner detector ($|\eta| < 3.2$), a liquid-argon (LAr) electromagnetic sampling calorimeter of fine granularity is placed which enables precision measurements of electrons and photons. For the rest of the system, a coarser granularity was chosen which, however, is sufficient for the detection of jets and the missing transverse energy reconstruction. The electromagnetic calorimeter constituents can be separated into a barrel and two end-cap parts. They are sourrounded by hadronic components, so that showers are contained and a punch through into the outer muon system can be prevented. In the barrel region, a sampling calorimeter with steel as absorber and scintillating tiles as active material is used. Furthermore, LAr calorimeters are placed in the end-cap region for the hadronic reactions which matches the preceding electromagnetic components. A third

constituent is the LAr forward calorimeter which is designed for hadronic as well as electromagnetic measurements [27].

The Muon System

The muon system represents the outer detector structure. It consists of a long barrel and two end-cap magnets which induce a strong curvature of the muon tracks. The muon momentum is measured in three layers of tracking chambers which are placed in the barrel as well as in the end-cap region. Their high precision alignment is guaranteed by precision mechanical-assembly techniques and optical alignment systems. For the tracking, Monitored Drift Tubes (MDTs) are used in the principle bending direction while Cathode Strip Chambers (CSCs) cover large η regions. Furthermore, the muon system is equipped with trigger chambers, namely Resistive Plate Chambers (RPCs) in the barrel and Thin Gap Chambers (TGCs) in the end-cap regions. One of their tasks is the position measurement orthogonal to the one determined by the tracking chambers [27].

The Trigger System

The proton beam collisions at the LHC produce interaction rates about 1 GHz. Due to resource limitations and the provided technology not all events can be recorded. Therefore, a trigger system is used which decides about the processed events. In a first step, the Level-1 (L1) trigger reduces the data rate to 75 kHz by considering a subset of the overall detector information. Afterwards, a Level-2 trigger as well as an event filter are applied which reduce the data taking rate to processable 200 Hz [27].

3.4 Event Reconstruction

The event reconstruction is another important step in the data processing of the experiment. The detector responses need to be mapped to particle reactions, so that physical phenomena can be probed. However, the event reconstruction is a complex field which requires sophisticated algorithms. Physical particles like electrons or muons leave different signatures in the detector. In some cases, however, the signatures of different objects are hard to distinguish as is the case for tau decays and QCD jets. For the event reconstruction, sensitive reconstruction techniques were developed. Some of their basic concepts will be presented, whereas the focus will be set on the tau lepton reconstruction which is a fundamental requirement for the analysis presented in this thesis.

3.4.1 Electrons & Photons

Electrons as well as photons deposit their energy in the electromagnetic calorimeters via electromagnetic showers. The particles differ by their signature in the tracking system. While the charged electron leaves a track, the uncharged photon cannot be detected in the innermost detector layers unless it converts to an e^+e^--pair. In the reconstruction process, electromagnetic calorimeter clusters above a threshold of 2.5 GeV [30] are identified and specialized algorithms match them to electrons, photons or converted photons depending on the overall detector signature. For a detailed description see [30] and [31].

3.4.2 Muons

Muons are charged, minimally-ionizing particles, detectable in every single detector structure. Different combined reconstruction processes using the subdetector information can be distinguished.
One of them is the Combined Muon (CB). In this case, a track was reconstructed in the Inner Detector (ID) as well as in the Muon Spectrometer (MS) and a combined track was formed in a refit procedure. A Segment-Tagged Muon (ST) results from a track in the ID which can be extrapolated and associated to a muon chamber track segment. Another combined muon object is the Calorimeter-Tagged Muon (CT). It combines an ID track with the energy deposition of a minimum ionizing particle. The fourth possibility is the Extrapolated Muon (ME). In this case, only a track in the MS, which is loosely compatible with the interaction point, was reconstructed. Further details can be taken from [32].

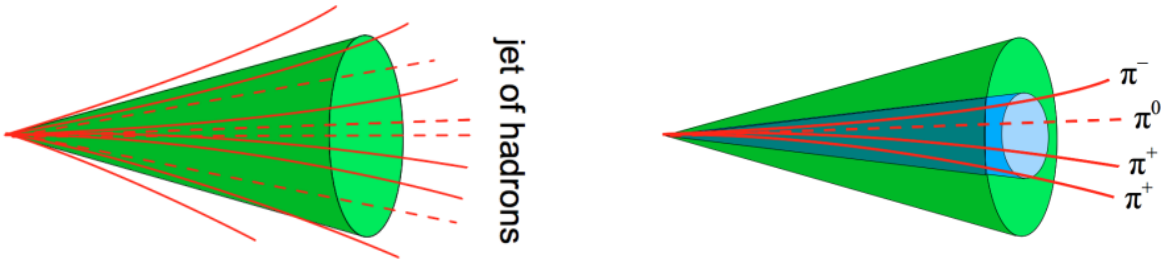
3.4.3 Jets

Jets are collimated hadrons which result from the hadronisation of colour-charged objects (see section 2.4). Their reconstruction is based on the anti-k_t algorithm [33] using calorimeter clusters above noise level. The reconstruction is limited to the radius $\Delta R < 0.4$. For further information see [34].

3.4.4 Tau Leptons

The signature of the tau lepton can take a variety of forms as listed in table 2.1. In the case of leptonic decays, only a light lepton (electron or muon) is visible in the detector since the two remaining neutrinos interact with the detector material with a small probability. Therefore, there is no dedicated leptonic tau reconstruction. However, both cases can be differentiated by the reconstructed missing transverse energy induced by the neutrinos, see section 3.4.5.
For hadronic decays, though, the signature comprises predominantly charged and neutral pions. Their appearance in the detector is quite similar to the one of jets initiated by quarks and gluons. Therefore, already reconstructed jets are used as starting point of the tau lepton reconstruction. Significant differences between a QCD jet and the hadrons of the tau decay are the number of involved hadrons and the collimation of the tracks. Whereas in a QCD jet several tracks spread within a large cone, the few pions originating from the tau are typically collimated as depicted in figure 3.7. The differentiation of both cases is performed by Boosted Decision Trees (BDTs) [35]. Since the neutrino information is missing in the reconstruction, the provided tau four momentum only considers the visible contributions. Depending on the efficiency of the tau candidate identification (ID), three different classes are distinguished, namely *loose, medium* and *tight*.

Apart from the reconstruction of the visible tau, the reconstruction of its visible decay products is provided by the *Tau Particle Flow* technique, which also provides a decay mode classification as well as a four momentum reconstruction.
In 90% of the cases, a hadronic tau decay can be assigned to one of the five decay modes listed in figure 3.8 [37]. Because of charge conservation, there are one or three charged hadrons and mainly up to two neutral pions involved which almost immediately decay to two photons. For the charged hadrons, only pions are considered which constitute most of the decays. While their reconstruction is based on the track system and the hadronic calorimeter, the π^0 reconstruction depends on the energy deposition in the electromagnetic calorimeters. Since the charged pions deposit energy in the electromagnetic calorimeters as well, their contribution $E^{\mathrm{EM}}_{\pi^\pm}$ has to be estimated, so that a correction



(a) Typical shape of a QCD jet.

(b) Typical shape of a hadronic tau decay.

Figure 3.7: Comparison between a QCD jet and the signature of a hadronic tau decay. Taken from [36].

for the neutral components can be applied. This measure is necessary since, due to the collimation of the pion tracks, energy clusters are likely to overlap. The estimation falls back on the energy information provided by the track system $E^{\text{trk}}_{\pi^\pm}$ and the hadronic calorimeter $E^{\text{HAD}}_{\pi^\pm}$. In an ideal setup

$$E^{\text{EM}}_{\pi^\pm} = E^{\text{trk}}_{\pi^\pm} - E^{\text{HAD}}_{\pi^\pm} \tag{3.9}$$

would be valid and can be used for correction [37]. Since a lot of neutral pion candidates also originate from $\pi^\pm$ remnants or pile-up, which describes multiple pp interactions per bunch crossing, a BDT decides about the direct assignment to a tau decay. Theoretically, a decay mode can be determined at this point by counting the number of reconstructed charged and neutral pions. Simultaneously applying another BDT, though, which considers among other aspects the tau decay kinematics as well as the number of previously reconstructed photons, can remarkably improve the performance [37]. Figure 3.8 shows the efficiency matrix of the decay mode classification for simulated $Z/\gamma^* \rightarrow \tau\tau$ events. In total, 74.7% of the events are correctly classified whereas the largest contribution of misclassification arises from events with multiple neutral pions involved. Their reconstruction often suffers from the assignment of multiple objects into one electromagnetic calorimeter cluster.
Apart from the decay mode classification, a four momentum reconstruction is provided as well. For this, the four momenta of the decay constituents are summed and a pion mass hypothesis is used for π^0 candidates. Exceptions cover the decay modes $h^\pm\pi^0$ with 2 π^0 candidates and $h^\pm \geq 2\pi^0$ with a single π^0 candidate but three or more assigned photons. For the former, each π^0 mass is set to zero, whereas for the latter, a mass of $2m_\pi$ is assumed for the single candidate. Furthermore, an energy calibration is applied [37].
The technique provides a suitable energy resolution for hadrons with transverse momenta less than 100 GeV, because of the high four momentum resolution of the tracking system. This aspect is exploited in an energy calibration applied in the reconstruction of the visible tau. The energy calibration is based on a Boosted Regression Tree (BRT) and leads to a superior energy resolution of the visible tau, especially for high momenta [38]. Since in the tau reconstruction, the mass is assumed to be zero, energy and transverse momentum are identical.

3.4.5 Missing Transverse Energy

The reconstruction of the missing transverse energy is based on on the principle of momentum conservation. Since the proton beams are aligned in z-direction, there is no initial momentum

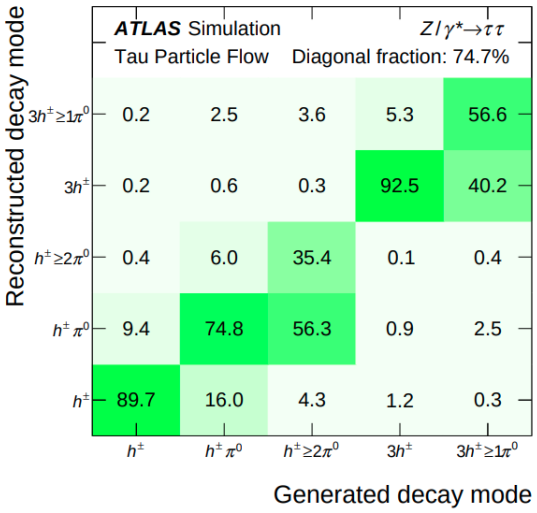


Figure 3.8: Efficiency matrix of the hadronic tau decay mode classification. In the notation, $h^\pm$ represents the charged hadrons of the decay which are predominantly pions. Matrix taken from [37].

contribution in the transverse plane. Consequently, the sum of the transverse momenta in the final state has to add up to zero. If this is not the case, it is assumed that this *missing energy* originates from weakly interacting particles in the event. Such particles are for example SM neutrinos.
The total missing energy in x and y direction comprises the negative sum of the summed momenta resulting from different objects [39].

$$E_{x,y}^{\text{miss}} = \underbrace{-\sum_{e_i} p_{x,y}^{e}}_{E_{x,y}^{\text{miss,e}}} \underbrace{-\sum_{\tau_i} p_{x,y}^{\tau}}_{E_{x,y}^{\text{miss},\tau}} \underbrace{-\sum_{\mu_i} p_{x,y}^{\mu}}_{E_{x,y}^{\text{miss},\mu}} \underbrace{-\sum_{\text{jet}_i} p_{x,y}^{\text{jet}}}_{E_{x,y}^{\text{miss,jet}}} \underbrace{-\sum_{\substack{\text{unassigned} \\ \text{track}_i}} p_{x,y}^{\text{track}}}_{E_{x,y}^{\text{miss,soft}}} \tag{3.10}$$

Hard and soft contributions are distinguished. The former include the reconstructed and calibrated e, τ, μ and jet objects. The applied calibration maps the detector signal to a physical four momentum. The soft contributions include all remaining tracks not assigned to a reconstructed hard object. They originate from underlying event activity and soft radiation from hard events [40]. The missing transverse energy vector $\vec{E}_T^{\text{miss}}$ can be described by its amplitude $|\vec{E}_T^{\text{miss}}|$ and its azimuthal orientation Φ^{miss} [39]

$$|\vec{E}_T^{\text{miss}}| = \sqrt{E_{\text{miss},x}^2 + E_{\text{miss},y}^2} \quad , \tag{3.11}$$

$$\Phi^{\text{miss}} = \arctan\left(\frac{E_y^{\text{miss}}}{E_x^{\text{miss}}}\right). \tag{3.12}$$

The missing transverse energy is an important but also uncertain quantity. On the one hand, a poor reconstruction as well as particles outside the detector acceptance can contribute besides the weakly

interacting particles. On the other hand, the multitude of involved objects increases the systematic uncertainty propagation to E_T^{miss}.

3.5 Di-tau Mass Reconstruction

Because of the short lifetime of the tau lepton (see section 2.7), the reconstruction is based on the decay products. The signature comprises the visible decay products and the unmeasured neutrinos. The physical relation between the di-tau mass and the four-momenta of the daughter particles reads

$$m_{\tau\tau} = \sqrt{\left(P_{\text{vis}_0} + P_{\nu_0} + P_{\text{vis}_1} + P_{\nu_1}\right)^2}, \tag{3.13}$$

where the indices 0 and 1 represent the belonging to the leading (more energetic) or the subleading (less energetic) tau respectively. Di-tau mass reconstruction techniques have to deal with the unknown momenta of the neutrinos involved in the event. In this context, the missing transverse energy, introduced in section 3.4.5 provides insight in the sum of the missing momenta. The individual neutrino properties, though, stay unknown.
Suitable di-tau mass reconstruction techniques are necessary in the Higgs analysis investigating the coupling to leptons. However, large background contributions arise from the Z boson, whose mass is with 91.19 GeV close to the Higgs mass of 125.1 GeV [13]. Therefore, a good di-tau mass resolution is crucial.

3.5.1 The Missing Mass Calculator

The MMC [3] is the current method used for di-tau mass reconstruction at ATLAS. It enables the handling of the missing event information caused by the neutrinos by considering the probability of possible configurations between the neutrinos and the visible decay products of the taus.
The whole reconstruction technique is based on a system of four under-constrained equations which include the missing polar and azimuthal angle θ_{miss_i} and ϕ_{miss_i}, the missing momentum p_{miss_i} as well as the missing mass m_{miss_i} of the neutrino systems. The latter is non-zero in leptonic decays. Again, the index i $\in [0, 1]$ denotes the belonging to the leading or subleading tau respectively. It is assumed that the total missing information originates solely from the neutrinos, whose numbers increase with the occurance of leptonic decays. In these cases, the combined information of the whole neutrino system is considered.

$$\not{E}_{T_x} = \underbrace{p_{\text{miss}_0} \sin\theta_{\text{miss}_0} \cos\phi_{\text{miss}_0}}_{p_{\text{x miss}_0}} + \underbrace{p_{\text{miss}_1} \sin\theta_{\text{miss}_1} \cos\phi_{\text{miss}_1}}_{p_{\text{x miss}_1}} \tag{3.14}$$

$$\not{E}_{T_y} = \underbrace{p_{\text{miss}_0} \sin\theta_{\text{miss}_0} \sin\phi_{\text{miss}_0}}_{p_{\text{y miss}_0}} + \underbrace{p_{\text{miss}_1} \sin\theta_{\text{miss}_1} \sin\phi_{\text{miss}_1}}_{p_{\text{y miss}_1}} \tag{3.15}$$

$$M_{\tau_0}^2 = m_{\text{miss}_0}^2 + m_{\text{vis}_0}^2 + 2\sqrt{p_{\text{vis}_0}^2 + m_{\text{vis}_0}^2}\sqrt{p_{\text{miss}_0}^2 + m_{\text{miss}_0}^2} - 2p_{\text{miss}_0}^2 p_{\text{vis}_0}^2 \cos\Delta\theta_{\text{vm}_0} \tag{3.16}$$

$$M_{\tau_1}^2 = m_{\text{miss}_1}^2 + m_{\text{vis}_1}^2 + 2\sqrt{p_{\text{vis}_1}^2 + m_{\text{vis}_1}^2}\sqrt{p_{\text{miss}_1}^2 + m_{\text{miss}_1}^2} - 2p_{\text{miss}_1}^2 p_{\text{vis}_1}^2 \cos\Delta\theta_{\text{vm}_1} \tag{3.17}$$

The equations 3.14 and 3.15 represent the missing energy in x and y direction, whereas 3.16 and 3.17 are expressions for the invariant masses of the tau leptons [3]. The angle $\Delta\theta_{\text{vm}}$ describes the polar separation between the visible tau and the missing neutrino system. In the following, the indices *vm* denote the comparison of these two systems. Since the ensemble of equations is underconstrained, possible solutions can be found by performing scans in the parameter space. They are applied for the azimuthal angular difference $\Delta\phi_{\text{vm}}$ as well as for the missing energy in x and y direction $\not{E}_x^{\text{miss}}$ and $\not{E}_y^{\text{miss}}$ [41]. The scans are performed for each tau using Markov Chains. In order to save computing power, the procedure is constrained on physically valid solutions. However, not each of them is equally likely. Therefore, an event likelihood $\mathcal{L}$ is formulated which considers the probability density functions (PDFs) of the spatial separation $\mathcal{P}(\Delta R_{\text{vm}})$, the resolution of the missing transverse energy $\mathcal{P}(\Delta\not{E}_T^{\text{miss}})$, and the ratio of the visible and the missing momentum $\mathcal{P}\left(|\vec{p}_{\text{vis}_i}|/|\vec{p}_{\text{miss}_i}|\right)$ [42]. The total event likelihood serves as weight for the single solutions.

$$\mathcal{L} = \mathcal{P}(\Delta R_0) \times \mathcal{P}(\Delta R_1) \times \mathcal{P}\left(|\vec{p}_{\text{vis}_0}|/|\vec{p}_{\text{miss}_0}|\right) \times \mathcal{P}\left(|\vec{p}_{\text{vis}_1}|/|\vec{p}_{\text{miss}_1}|\right) \times \mathcal{P}(\Delta\not{E}_T^{\text{miss}}) \tag{3.18}$$

The PDFs themselves vary with specific event signatures. Therefore, they are parametrized as functions of the variables and quantities which evoke dependencies. Whereas $\mathcal{P}(\Delta R)$ is parametrized in $p_{T,\text{vis}}$, $\mathcal{P}(\Delta\not{E}_T^{\text{miss}})$ takes the dependence of the transverse energy contributions $\sqrt{\sum_i E_{T,i}}$ as well as the polar angular difference of the two visible taus ($\Delta\phi_{\text{vis}}^{\tau_0,\tau_1}$) into account. Furthermore, different PDFs are considered, depending on the tau decay type ($\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{lep}}$). Additionally, the different decay modes of hadronic decays are distinguished for all likelihood contributions except $\mathcal{P}(\Delta\not{E}_T^{\text{miss}})$ [42].
In order to define the final reconstructed di-tau mass, the most suitable solution can be chosen from different possibilities. Either the configuration with the largest total likelihood (maxw) or the most probable neutrino three momentum (mlnu3p) is selected. In these cases, the di-tau mass can be calculated from the four momenta of the visible taus and the neutrinos. Alternatively, the most probable di-tau mass (mlm) can be chosen, but in this case no neutrino information is available due to a scaling process for different tau decay types ([41],[42]).

3.6 Event Generation & Simulation

Scientific models remain theoretical constructs until they are confirmed by experimental outcomes. The same also applies for the SM of particle physics whose assumptions are, by now, in accordance with the observations. Because of the complexity of the model and the huge amount of processes involved, especially in hadron hadron collisions, precise simulations of the physical expectations are necessary in order to compare them to the actual measurements.
The different processes involved in pp collisions at the LHC are modeled by MC based techniques which generate random events with frequency distributions following the physical laws of nature. An overview over the processes that need to be considered by the event generators is depicted in figure 3.9. The hard processes are depicted in red. Their perturbative calculation is performed on matrix element level [43] considering the probability of a parton carrying a certain momentum fraction of the proton. The hard scattering process is accompanied by initial state radiation (ISR) and final state radiation (FSR) represented by the blue objects. Because of the fact that the emitted gluons carry colour themselves, further radiation processes occur forming parton showers. Furthermore,

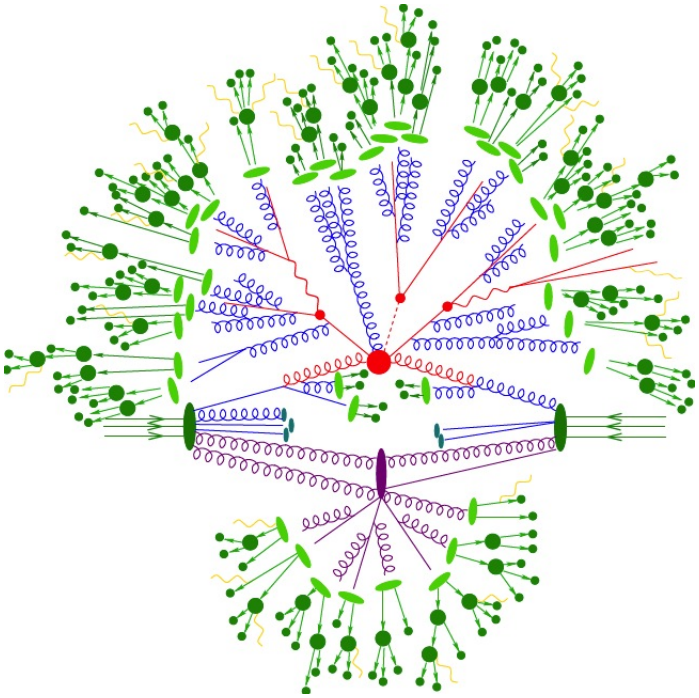


Figure 3.9: Exemplary sequence of processes in pp collisions considered by MC event generators (taken from [44]). Red colored objects refer to hard parton parton scattering. ISR and FSR are shown in blue whereas the fragmentation and hadronisation process is represented by green objects. Furthermore, QED Bremsstrahlung is shown in yellow and a soft underlying event is represented in purple.

fragmentation and hadronisation result in particle jets (green). The QED Bremsstrahlung is depicted in yellow and a further underlying soft event is shown in purple. Because of the low energy transfer, they cannot be described in the perturbative sector.

For the modelling of the whole sequence of processes a multitude of generators, specialized in specific steps of the chain, can be used. There are also integrated frameworks available like *Sherpa* [45] or *Pythia* [46] which already comprise the different simulation processes like matrix-element calculations, parton shower and underlying event modeling, hadronisation, particle decays etc.. The following list describes the generators relevant for this thesis.

Sherpa models the leading order (LO) processes using further tools like *Apacia* [47] for parton showering or *Photos* for QED radiation [48]. The hadronization process is based on an internal clustering algorithm and the tau leptons are also generated by the framework itself [49].

Pythia is also a LO multi purpose event generator using the Lund string model [50] for hadronization. It generates tau leptons internally and is often used in combination with other generators.

Powheg [51] is specialized in the matrix element generation up to next-to-leading order (NLO) . For parton showering and hadronization, Pythia is used. In order to consider overlaps between the NLO

matrix element calculations and the showering algorithm, p_T based vetos are applied in the latter [51].

In a next step, the generated events need to be processed by the the ATLAS detector simulation [52]. For this task, a copy of the detector geometry is built in *Geant4* [53] in order to imitate the interaction of the generated particles with the detector material. The corresponding electrical signals of the detector components are determined, so that the outcome of the simulation is comparable to the real detector response. Therefore, the event reconstruction algorithms of section 3.4 can be equally applied to the simulation and the real data. In order to guarantee consistency, every set of simulated events has to be cross-checked with the data. For this task, well understood data of so called control regions can be used.

CHAPTER 4

Artificial Neural Networks

In the past, the concept of artificial intelligence was part of wild imaginations and futuristic visions and covered a wide array of domains and forms. The general idea, which is still valid, referred to machines or algorithms that acted cognitive and intelligent. Apart from manlike robots, often mimicking the human body, self driving cars or other machines which undertook ordinary tasks were considered. Even though not every ambitious idea has been realized by today, artificial intelligence is no longer pure fiction but rather part of our daily life. Apart from speech recognition [54] and translation tools [55], spam filters [56] and gaming engines [57], artificial intelligence is also applied in economy [58] and science where, again, a multitude of applications can be observed. In medicine, image recognition is used for cancer pathology [59], in meteorology climate models can be parametrized [60] and this thesis itself is just another example for the application of artificial intelligence in particle physics.
The term *artificial intelligence* is quite extensive and includes many subfields and categories. One of the most prominent ones is the term *machine learning* which comprises algorithms that learn without being specifically programmed for the task. The learning, however, is not totally independent since a human intervention is still necessary in order to initiate the learning process, for example in form of a preselection of features. This aspect differentiates machine learning from the field of *deep learning* (see section 4.2) which manages this task independently [61]. Nowadays, it is one of the most successful applications of artificial intelligence. Its basic concepts will be presented in this chapter while the selection of listed aspects focuses on regression tasks (see section 4.1). For further reading see [62] or [63].

4.1 General Concepts

Machine learning algorithms are supposed to act intelligent requiring the capability to learn. A proper definition of the term is given by Mitchell [64]:

> "A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

This statement includes the most important principles. First of all, each algorithm has a task T to fulfill. Two frequently used tasks are classification and regression. While in the former, the algorithm tries to assign a specific label to the input data, the latter predicts a continuous variable [63]. In order

to estimate in what sense a task is fulfilled, a suitable performance measure needs to be chosen. For regression models, the *mean-squared-error* (see section 4.3.3) is commonly used which compares the output of the algorithm to the desired value. The performance has to improve with the experience E which is gained from multiple processing of the provided data (see section 4.4).
In general, three different learning strategies can be differentiated.

Supervised Learning algorithms work with labeled data sets as is the case for classification and regressions tasks. The input data is provided as a vector containing different features. Each feature set receives a label, i.e. a target which represents the desired/ true solution of the learning task [62].

Unsupervised Learning , in contrast, has no labels. It processes the data and tries to identify hidden properties and structures. A common application is the clustering of data which can, for example, be used to identify groups of customers with common interests [63].

Reinforcement Learning is a concept that is based on the interplay of the actions of an agent with an environment. A chess engine can be used as illustrative example in which the arrangement of the chess pieces represents the environment and different moves of the engine the actions. These moves are rewarded by winning or losing [63].

4.2 Deep Learning

Deep learning is based on artificial neural networks (ANNs) which are computing systems that are inspired by the human brain. They consist of multiple connected neurons that process the data. For the data processing, a multitude of free parameters θ are available, namely one bias term b per neuron and one weight w per connection. Depending on the model's architecture, this results in several thousands of parameters which have to be adapted in the learning process. The final goal of the network is to approximate a function $g(x;\theta)$ which represents a mapping between the input data $\mathbf{x}$ and the desired output $\mathbf{y}$, so that

$$y \approx \hat{y} = g(\mathbf{x};\theta) \tag{4.1}$$

is nearly valid. The neurons are arranged in different layers as depicted in figure 4.1. The depicted model is a fully connected feed-forward neural network without any feedback loops like it would be the case in recurrent neural networks. The initial layer, the so called *input layer* receives the training data. Thus, its number of neurons corresponds to the feature size of the input data set. In the following hidden layers, the data will be processed, so that a final prediction can be made in the output layer whose number of nodes is fixed by the task. From another point of view, the model can be interpreted as a chain of L layers which are passed by the data. Each layer performs some operations f^i, so that the final output is the result of chained operations [62]

$$g(\mathbf{x};\theta) := f^L(f^{L-1}(f^{L-2}(..f^1(\mathbf{x};\theta)))). \tag{4.2}$$

The number of links defines the *depth* of the model which coined the term of *deep learning*.

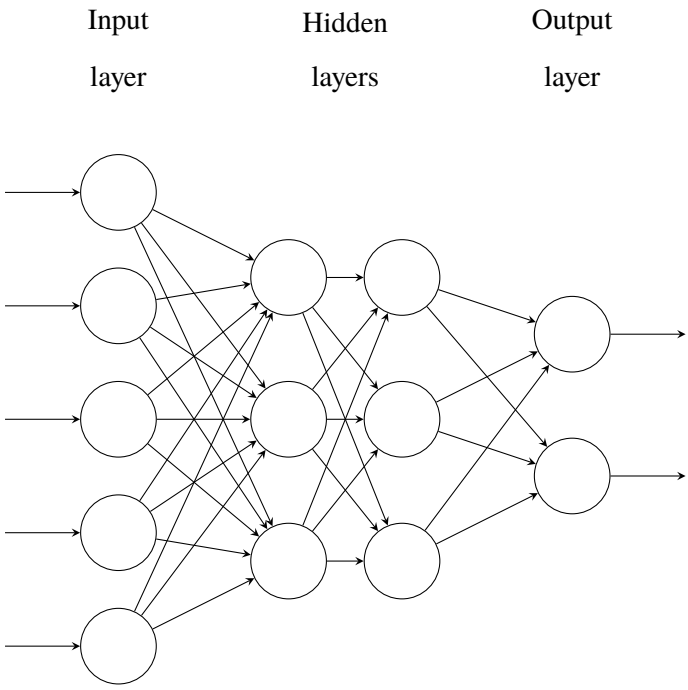


Figure 4.1: Basic architecture of a fully connected feed-forward artificial neural network. [12]

4.2.1 The Neuron

Since artificial neural networks try to mimick the human brain, the substructure, namely the neurons perform the same tasks. Whereas in the biological case, chemical and electrical signals are fed into the neuron, the artificial one receives input values which will be further processed. If the signal/ final digit exceeds a threshold value, the neuron "fires" and generates an output which will be passed on inside the the network [63].
Figure 4.2 illustrates the described data processing in detail. The single neuron can receive several inputs x_i which correspond to the outputs of all connected previous neurons. Each of these inputs receives an individual weight w_i. The first mathematical operation within the neuron corresponds to a matrix multiplication between the input vector $\vec{x}$ and the weight vector $\vec{w}$. Adding an additional bias term leads to the net input z, which is handed to a non-linear activation function f_{act} (see section 4.3.2) which produces the final output of the neuron [63].

$$\text{output} = f_{\text{act}}(z) = f_{\text{act}}\left(\underbrace{\mathbf{w}^{\mathbf{T}}\mathbf{x} + b}_{z}\right) = f_{\text{act}}\left(\underbrace{\sum_i x_i w_i + b}_{z}\right) \tag{4.3}$$

The non-linearity is essential in order to reflect complex interactions of several inputs [62].

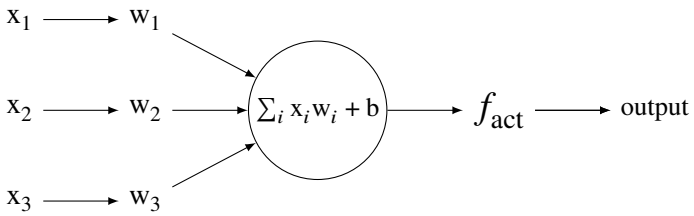


Figure 4.2: Data processing in a single neuron. Inspiration from [63], created with [12].

4.2.2 The Learning Process: Loss, Backpropagation & Gradient Descent

The free parameters of the network, the weights and biases have to be iteratively adapted, so that the produced output comes as close as possible to the desired one. This is done by minimizing a loss function which considers the deviation of the target value y and the prediction $\hat{y}$. The simplest loss for regression models is the *mean-squared-error* which determines the mean of the squared deviations of all m outputs $\hat{y}_i$ w.r.t the true target value y_i

$$L(\mathbf{y};\theta) = \frac{1}{m}\sum_{i}^{m}\left(\hat{y}_i - y_i\right)^2 . \tag{4.4}$$

The minimization is often performed by gradient based optimization algorithms like stochastic gradient descent. Since neural networks benefit from large statistics of the training sets, it is necessary to provide an efficient gradient determination. Therefore backpropagation ([65], [66]) is used which performs the determination layer-wise and propagates the information backwards through the network. This is a very efficient application of the chain rule. In the minimization process, the contribution of each weight w_{ij} to the final loss L is of interest in order to apply a corresponding adaption. In this notation, the first index i denotes the neuron of the previous and j the one of the current layer. The parameter z corresponds to the weighted inputs and o represents the output of the neuron as defined in equation 4.3. In the following, the bias term b will be incorporated in the weight matrix as w_0 with a corresponding input value of $x_0 = 1$. Applying the chain rule twice leads to equation 4.5. The second term corresponds to the derivative of the activation function and the final expression represents the i^{th} input of neuron j, i.e. the output of the neuron i in the previous layer.

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial o_j}\underbrace{\frac{\partial o_j}{\partial z_j}}_{f'_{\text{act}}(z_j)}\underbrace{\frac{\partial z_j}{\partial w_{ij}}}_{o_i} \tag{4.5}$$

In order to determine the derivative of the loss with respect to the output of neuron j, all neurons N, which receive input from neuron j, have to be considered .

$$\frac{\partial L}{\partial o_j} = \sum_{l\in N}\left(\frac{\partial L}{\partial o_l}\frac{\partial o_l}{\partial z_l}\underbrace{\frac{\partial z_l}{\partial o_j}}_{w_{jl}}\right) \tag{4.6}$$

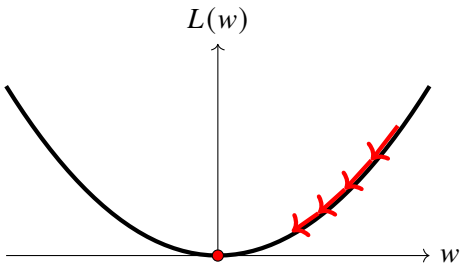


Figure 4.3: Visualization of the weight update using gradient descent. The aim is to reach the global minimum, represented by the red spot. Inspiration from [63], created with [12].

The recursive expression and the second term of 4.5 form a variable δ_j which can be efficiently computed as long as the determination is performed backwards starting at the output layer.

$$\delta_j = \frac{\partial L}{\partial o_j} \frac{\partial o_j}{\partial z_j} \tag{4.7}$$

Now, that the gradient of the loss function with respect to the variable parameters can be determined, a rule for the weight update has to be defined. The simplest optimization algorithm uses gradient descent in order to approach the minimum of the loss function. Therefore, the initial weight is changed by a term proportional to the negative gradient [63].

$$w_{ij} = w_{ij} - \alpha \frac{\partial L}{\partial w_{ij}} \tag{4.8}$$

This term is called the *learning rate* α and determines the step size towards the minimum, represented by the arrow length in figure 4.3. Its appropriate choice is decisive in the optimization process. Whereas small learning rates require lots of weight updates, large values might lead to pass over the minimum.

4.3 Common Algorithms & Methods

4.3.1 Generalization & Regularization

A useful neural network needs to generalize the knowledge provided in the training data set so that the model can also be applied to unknown data. In this context, the aspects of over- and underfitting become relevant. Overfitting occurs if the model insists on the training sample's peculiarities and does not recognize general features and properties. In this case, the error on the training data set might be small but the generalization error from unseen data is not. This aspect differentiates simple optimization tasks from machine learning which needs to provide general functionality. In the case of underfitting, the optimization process fails to reduce the error on the training set. Ideally, a balance between both loss types can be established but regarding the applicability of a model, the focus is set on the generalization. In order to reduce its error, regularization techniques can be applied which range from soft constraints to strict restrictions on the parameter values. These measures for the reduction of the generalization error might be at the expense of an increased training loss. Common approaches are the L^1 and L^2 parameter regularization which base on a weight decay [62].

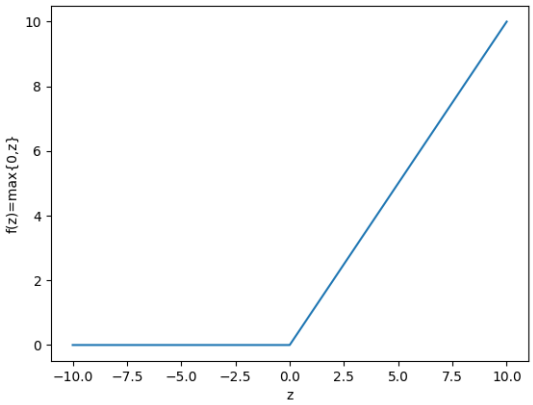


Figure 4.4: Rectified linear activation function (ReLU).

4.3.2 Activation Function

The activation function defines the final output of a single neuron and has to be non-linear, which is necessary in order to describe the correlation of multiple features. Furthermore, differentiability is required for the application of backpropagation. A common choice is the rectified linear activation function *ReLU*, depicted in figure 4.4,

$$f_{\text{act}}(z) = max\{0, z\} \tag{4.9}$$

which is a piecewise linear function. Because of the maintenance of linear characteristics, it is well suited for gradient-based optimization algorithms [62]. Nevertheless, the overall output is non-linear. Since the function is not continuous at zero, the derivative is defined as

$$f'_{\text{act}}(z) = \begin{cases} 0, & \text{if} \quad z \leq 0, \\ 1, & \text{otherwise} \end{cases} . \tag{4.10}$$

4.3.3 Loss Function

The loss function measures the error of the network's prediction. It decides about the learning process and therefore has to be chosen carefully. Since many optimizers rely on gradient based methods, the loss function has to be differentiable with well defined gradients in order to ensure a good guidance towards the minimum. The most common choice in regression tasks is the *mean-squared-error* (MSE) whose general form follows equation 4.4. Its quadratic weighting increases the sensitivity for outlier predictions. Alternatively, the *mean-absolute-error* (MAE) can be chosen which replaces the square by the absolute value and provides a linear weighting instead. Another choice, similar to the mean-squared-error but not as strict for outlier predictions, is the *Log-Cosh-loss*, which is defined as [67]

$$L(\mathbf{y}; \theta) = \frac{1}{m} \sum_{i=0}^{m} \log\left(\cosh(\hat{y}_i - y_i)\right) . \tag{4.11}$$

4.3.4 Optimizer

Nowadays, one of the most common optimization algorithms is *stochastic gradient descent* (SGD). Instead of using the full training sample at once, it randomly picks small *minibatches* in order to estimate the gradients. This approach is computationally more efficient [62]. Furthermore, stochastic algorithms introduce noise which might help to escape shallow minima of the loss function [62].
One common variation of SGD uses momentum [68]. Its weight update is based on a velocity term v, whose update is characterized by a moving average of the previous gradients which decay exponentially [62]. The parameter $\gamma \in [0, 1]$ decides about the extent of the decay.

$$v_t \leftarrow \gamma v_{t-1} - \alpha \nabla_\Theta \left(\frac{1}{m} \sum_i^m L(\hat{y}_i(x_i, \theta), y_i) \right) \quad , \tag{4.12}$$

$$\theta_t \leftarrow \theta_{t-1} + v_t \quad . \tag{4.13}$$

The concept of momentum benefits from a fast convergence compared to the common SGD [62].

Another class of common optimizers are those with adaptive learning rates like *AdaGrad* [69], *RMSProp* [70] or *Adam* [71]. They introduce individual learning rates for each parameter and vary them during the training process [62]. Since the loss sensitivity depends on the direction in parameter space, this measure might positively effect convergence. For the task of this thesis, the Adam (adaptive moments) optimizer turned out to be useful. Apart from an exponentially decaying moving average of the gradient, which is comparable to the momentum, it also considers an exponentially decaying moving average of the squared gradient as well.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{4.14}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{4.15}$$

The variables m_t, and v_t, can be understood as the first and second momentum of the gradient (mean and uncentered variance). The variables β_1, and β_2, proved to lead to satisfactory performances with default values of 0.9 and 0.999 respectively [71]. In order to avoid biases of the momenta, a correction of the form

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad , \tag{4.16}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad , \tag{4.17}$$

is applied [62]. The final parameter update follows [71]

$$\theta_{t+1} = \theta - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad , \tag{4.18}$$

with the learning rate α and ϵ being small (default: 10^{-8}).

4.4 Hyperparameters

The hyperparameters of a neural network are the characteristic settings that need to be specified before the training process and which are not adapted by the learning algorithm itself [62]. The final performance depends strongly on their choice, which is why they should be chosen carefully. This section will expand the scope of already mentioned hyperparameters.
In general, two types can be distinguished, namely model and algorithm based settings. The former comprises the topological architecture of the ANN which is defined by the number of layers, their specific number of nodes and the connection among each other. Most of the hyperparamters affect the speed and the quality of the optimization process. Algorithm based hyperparameters cover, for example, the learning rate, which sets the step size of the weight updates or the mini-batch size determining how many input/target pairs are processed at once. A typical choice is a batchsize of 32 [72]. Generally, the full data set has to pass the network several times in order to achieve satisfactory results. One passage of the full set is called an *epoch* [63]. In order to ensure a proper learning environment, all parameters have to be initialized carefully, also the weights and biases. Their starting values can prevent exploding or vanishing gradients in the training process. A common choice are, for example, random samples from a truncated normal distribution.

4.5 Constrained Neural Networks

The fundamental idea of artificial neural networks is the independent learning from data. In general, they are not specifically programmed for a task, but rather left to their own. In some applications, it is essential, though, that certain constraints are fulfilled. In common neural network implementations, parameter constraints like non-negativity or unit norm are already available. Task specific constraints, however, which concern the output of the network, have to be realized individually. This applies for regression models which are subjected to the conservation laws of physics. However, enforcing fundamental requirements does not automatically improve the performance. In this section, two approaches of constraint implementations will be presented in the context of di-tau mass estimation and their influence on the learning process will be addressed.

4.5.1 The Concept

The basic approach for di-tau mass reconstruction, using regression based neural networks, is the training on the true invariant mass of the di-tau system. This strategy was already pursued in reference [4]. A model based on this target makes not obligatorily use of the full knowledge which can be provided. Even though, a variety of inputs can be used for the training, correlations between them have to be revealed by the network itself. Depending on the learning achievement, this task is fulfilled more or less.
One of these correlations is the invariant tau mass which is imposed on the decay products of the lepton. In order to incorporate this requirement in the training of the network, the regression task has to be adapted. Instead of the di-tau mass as output, a multidimensional output predicting the neutrino momenta is necessary. Since the four momentum of the visible tau is known (see section 3.4.4), the estimation of the neutrino momenta can be controlled, requiring the correlation of equation 2.16. As target, the triplet (p_T, η, ϕ) was chosen, like motivated in section 7.1. Although the multidimensional

output increases the complexity of the neural network, the expected profit is based on the additional knowledge provided by the constraint.

4.5.2 The Soft Constraint

One possibility to incorporate the requirement of the invariant tau mass into the training process is the *soft constraint*. It penalizes the optimizer depending on the extent of the constraint violation by an additional term in the loss function. Thus, minimizing the loss also means fulfilling the imposed restrictions.

The predictions of the neutrino momenta need to be controlled in favour of the invariant tau mass requirement. Therefore, the hypothetical tau mass resulting from the network's outputs is determined as described by expression 4.19

$$m^2_{\text{calc}} = \left(\begin{array}{c} \underbrace{\sqrt{m_c^2 + p_{T,c}^2 \left(1 + \sinh(\phi_c)^2\right)}}_{E_c} + \underbrace{\sqrt{m_n^2 + p_{T,n}^2 \left(1 + \sinh(\phi_n)^2\right)}}_{E_n} + \underbrace{\sqrt{\boldsymbol{p_T}^2 \left(1 + \sinh(\boldsymbol{\eta})^2\right)}}_{E_\nu} \\ \underbrace{p_{T,c} \cos(\phi_c)}_{p_{x,c}} + \underbrace{p_{T,n} \cos(\phi_n)}_{p_{x,n}} + \underbrace{\boldsymbol{p_T} \cos(\boldsymbol{\phi})}_{p_{x,\nu}} \\ \underbrace{p_{T,c} \sin(\phi_c)}_{p_{y,c}} + \underbrace{p_{T,n} \sin(\phi_n)}_{p_{y,n}} + \underbrace{\boldsymbol{p_T} \sin(\boldsymbol{\phi})}_{p_{y,\nu}} \\ \underbrace{p_{T,c} \sinh(\phi_c)}_{p_{z,c}} + \underbrace{p_{T,n} \sinh(\phi_n)}_{p_{z,n}} + \underbrace{\boldsymbol{p_T} \sinh(\boldsymbol{\eta})}_{p_{z,\nu}} \end{array} \right)^2 \tag{4.19}$$

The calculation is performed within the loss function in each prediction step of the network for the leading and subleading tau respectively. The red variables are related to the neutrinos and estimated by the ANN. The remaining quantities, though, are known from the reconstruction. They are fed into the network by additional input nodes, which are not connected to the hidden layer structure. Through this measure, the necessary information can be accessed in its original form in the loss function. Furthermore, all input data passes the same shuffling procedure so that the information of all nodes always corresponds to the same event. For the tau mass determination the charged and neutral decay products of the tau decay, denoted by the indices c and n, are considered. Their reconstructed mass is, different to the visible tau mass, unequal zero (see section 3.4.4). This fact is in favour of the sensitive determination of the invariant tau mass. It reduces the number of invalid calculations, which may result from negative expressions under the square-root. Such calculations need to be intercepted in order to prevent the abortion of the training process.
The deviation between the determined tau mass and the desired 1.77 GeV is considered in an additional penalty term in the loss function. In this analysis, it is designed in the style of a further mean-squared-error term which adds up the squared mass differences per batch. Equation 4.20 describes the total loss function of the soft constraint approach. The first term represents the common mean-squared-error with the desired and predicted value y and $\hat{y}$ respectively. The second term, though, incorporates the constraint. The single contributions can be weighted by a factor $\alpha \in [0, 1]$ which resembles a

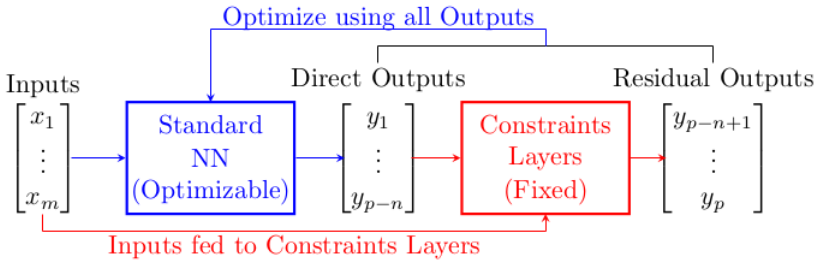


Figure 4.5: Basic architecture of a hard constrained network. The direct outputs result from the prediction of a common neural network, while the residual outputs are calculated in a following layer. This ensures the constraint realization within machine precision. All outputs are considered in the optimization process. Since $n = 2$ constraints have to be fulfilled, namely the invariant masses of the leading and subleading tau, only $p - n = 4$ direct outputs are needed. [74]

Lagrange multiplier [73]. This factor α is another parameter which needs to be optimized.

$$\mathcal{L} = (1-\alpha) \cdot \underbrace{\frac{1}{N}\sum_{i=0}^{N}\left(y_{\mathrm{i}} - \hat{y}_i\right)^2}_{\text{mean-squared-error}} + \alpha \cdot \underbrace{\frac{1}{N}\sum_{i=0}^{N}(m_\tau - m_{\mathrm{calc},i})^2}_{\text{penalty}} \tag{4.20}$$

All network outputs contribute to the penalty term so that the network has to coordinate all outputs simultaneously in order to largely fulfill the mass constraint.

4.5.3 The Hard Constraint

The hard constraint is another possibility to require the invariant tau mass. In this approach, only a reduced set of target variables will be predicted and the remaining quantity will be determined by applying the constraint. Therefore, the constraint is realized within machine precision.
However, such a technique requires a specific architecture of the model as illustrated in figure 4.5. The first part of the hard constrained model is a common neural network, consisting of an input layer, the following hidden structure and an output layer. Following the notation of figure 4.5, $n = 2$ constraints need to be fulfilled, namely the invariant masses of the leading and subleading tau. Therefore, only $p - n = 4$ direct outputs will be considered in the first network block. The predicted values are passed on to the second block of the architecture, the constraint layer. This element ensures the realization of the constraint by calculating the remaining neutrino quantity accordingly.
Constraint layers are fixed layers. They have no internal weights that will be adapted in the training process. Nevertheless, they influence the learning by their outputs. The total loss comprises the contributions from both network blocks. Since the direct outputs contribute to the calculation within the constraint layer, the optimization process needs to adapt the weights in favor of the constraint realization.

Since the task consists of 6 unknowns, theoretically each prediction can be undertaken by the fixed layers. However, the quadratic constraint can involve ambiguities for its solutions. Since the transverse

momentum is positive without exception, it is chosen as residual output. Equation 4.21 expresses the corresponding calculation within the constraint layers. It is the transformed version of equation 4.19, with an assumed tau mass of $m_\tau = 1.77\,\text{GeV}$. The calculation is performed for the neutrino belonging to the leading as well as for the subleading tau in two separate layers. The variables in red represent the direct outputs of the network. Again, the reconstructed variables of the charged and neutral decay products are provided by the additional input nodes.

$$p_T = \frac{m_\tau^2 - (E_c + E_n)^2 + \left(p_{x,c} + p_{x,n}\right)^2 + \left(p_{y,c} + p_{y,n}\right)^2 + \left(p_{z,c} + p_{z,n}\right)^2}{2\left((E_c + E_n)\sqrt{1 + \sinh(\boldsymbol{\eta})^2} - \left(p_{x,c} + p_{x,n}\right)\cos(\boldsymbol{\phi}) - \left(p_{y,c} + p_{y,n}\right)\sin(\boldsymbol{\phi}) - \left(p_{z,c} + p_{z,n}\right)\sinh(\boldsymbol{\eta})\right)} \tag{4.21}$$

4.5.4 Difficulties

The realization of constraints affects the learning process. One difficulty arises from sudden jumps of the loss, due to the described calculations. A division for example, can induce diverging values if the denominator becomes sufficiently small. Consequently, the internal weights and biases are drastically changed even if the underlying prediction was quite reasonable. Such jumps counteract a smooth learning process and might prevent the convergence towards the global minimum of the loss function. Another aspect especially concerns the hard constraint approach. The resolution of the reconstructed variables vitiates the exact constraint realization on machine precision. Depending on the setup, an enforced constraint could even decrease the performance compared to the unconstrained model. Apart from these general issues, more task specific discussions will be presented in chapter 8.

4.6 Importance in Particle Physics

In particle physics, well established algorithms are replaced increasingly by the application of ANNs such as the hadronic tau decay identification in ATLAS [75]. ANNs profit from their capability to turn data into knowledge without being specifically programmed for the final task. Thus, they might recognize correlations in data which are hard to exploit by other methods or which were even unknown. Another advantage concerns the fast evaluation time of ANNs. Once they are trained, they can be evaluated without large computational cost. Therefore, they are suitable to process huge amounts of data as it is produced in hadron hadron collisions. Furthermore, it is straightforward to expand the ANNs by additional input nodes which can provide further information for the learning process.
All the mentioned aspects motivate the di-tau mass reconstruction with ANNs. Especially the advantage of fast evaluation outperforms the MMC which requires many computational resources for the event based phase space scans (compare to section 3.5.1).

CHAPTER 5

Data Selection & Processing

The data used for the training decisively influences the model's performance since it contains the "physical knowledge". The following investigations are based on simulated Monte Carlo data whose fundamental production concepts are described in section 3.6. The usage of simulated data is advisable since real measurements might contain unknown characteristics resulting from new physics or defective measurement devices. In order to provide consistency and to understand physical phenomena a comparison of simulation and real measurements needs to be performed on a regular basis.
In this chapter, the data sets used in this thesis are introduced by describing the chosen generators and the simulation related settings. Furthermore, a motivation for their usage will be given and the cuts which are applied in order to ensure a suitable training/evaluation environment will be listed. Additionally, the features for the training are presented and their initial processing will be described.

5.1 Training Sample

The final demand on the ANN is the di-tau mass estimation based on unknown data. Therefore, the training sample is required to contain general properties of the event without emphasizing sample specific features. Otherwise, the mass estimation would be biased as soon as a sample with different characteristics would be applied.
First of all, a large di-tau mass range covering the mass region of the Higgs and Z is necessary in order

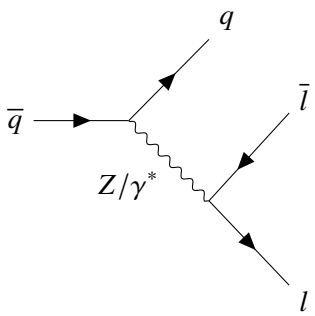


Figure 5.1: Feynman diagram of the Drell-Yan process. A quark antiquark annihilation produces an intermediate state particle which finally decays to two leptons [12].

Process	Generator	N_{total}	$N_{selected}$
$\gamma^* \rightarrow \tau_{had}\tau_{had}$	Pythia8EvtGen	5436998	2142022
$Z/\gamma^* \rightarrow \tau_{had}\tau_{had}$	Sherpa	878838	548707
$H \rightarrow \tau_{had}\tau_{had}$	PowhegPy8EG	242602	160955

Table 5.1: Overview of the simulated data samples. The simulated process, the applied generator as well as the number of events before and after the selection are listed.

to provide a proper learning environment. Furthermore, a flat mass distribution is desirable since it prevents the network from having a preferred mass. For example, if the mass distribution would be gaussian instead, the prediction of the mean would be a suitable choice inducing low losses in most of the times.
The required properties are well covered by a virtual photon decaying to two hadronically decaying tau leptons, $\gamma^* \rightarrow \tau_{had}\tau_{had}$. The simulated di-tau masses range from 60 GeV to 7 000 GeV. In order to provide a flat mass distribution, the originally falling mass distribution is reweighted [76] so that the main demands on the training sample are fulfilled. The corresponding frequency distribution is depicted in figure 5.2 in which only events passing the selection described in section 5.1.1 are considered.
However, the sample choice induces some limiting factors. First of all, the applicability of the trained network is restricted to the hadronic decay mode of the tau lepton. Leptonic tau decays would introduce further neutrinos which increase the number of unknown parameters. Furthermore, an exclusive $\gamma^* \rightarrow \tau\tau$ process is unphysical and not observable in nature. It rather interferes with $Z \rightarrow \tau\tau$ or other intermediate neutral particles that couple to the initially annihilating quark-antiquark pair of the pp collision and the final state leptons. This lepton production in hadron hadron collisions is known as Drell-Yan process [5] and is exemplary depicted in figure 5.1.
The $\gamma^* \rightarrow \tau_{had}\tau_{had}$ sample, which will be denoted as *training sample* in the following, is based on *Pythia8* and MC event generator version 16. For the pp collisions, a center of mass energy of 13 TeV was assumed. Within the integrated framework, the A14 [77] set of tuned parameters and the *NNPDF23LO* [78] set of parton distribution functions (PDFs) are used. The event generator [79] modelled the decay of bottom and charm hadrons. The full description of the sample is given in appendix A.

5.1.1 Selection

In order to ensure a proper training environment, selection cuts were applied. First of all, the fundamental requirement of two tau leptons is fulfilled by selecting events with true pdgIDs of 15 or -15 which represent the heavy lepton and its antiparticle. Another cut is applied to the true di-tau mass which limits the range from 60 GeV to 220 GeV. This interval is sufficient for the physics of this thesis since the Z and the Higgs peak are generously covered. A last requirement considers the reconstructed di-tau mass of the MMC (mlnu3p_m). Only events with a successful mass reconstruction by the MMC (mlnu3p_m > 0) are chosen in order to enable a comparison with this technique. The total number of events before and after the selection cuts can be looked up in table 5.1.

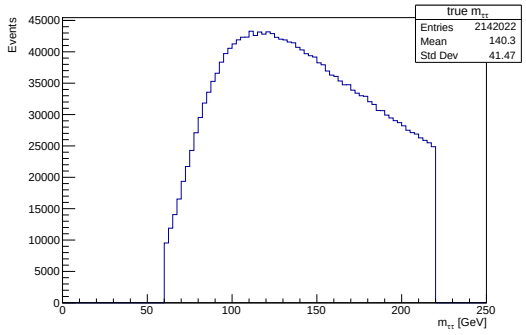


Figure 5.2: Frequency distribution of the true di-tau mass. Only events passing the selection described in section 5.1.1 are considered.

5.2 Evaluation Samples

In order to measure the network's performance, $H \rightarrow \tau\tau$ as well as $Z/\gamma^* \rightarrow \tau\tau$ samples will be evaluated on the trained models. A combination of several statistically independent samples will be used. They are simulated at center of mass energies of 13 TeV. Again, MC event generator version 16 was applied. The Z sample production is based on the *Sherpa* event generator using the PDF set *NNPDF30NNLO*, while the H samples are produced with *Powheg* and *Pythia8* using the PDF sets *NNPDF30_nnlo* as well as *NNPDF30_AZNLOCTEQ6L1*. The full description of the samples is given in appendix A. For convenience, the simulated data sets will be denoted as *Higgs* and *Z sample* in the following.

5.2.1 Selection

Since the Higgs as well as the Z sample are used for performance measurements, only the processes of interest should be considered. Therefore, events with reconstructed jets originating from energy depositions resulting from hardware problems, cosmic-ray showers etc, need to be rejected. In order to do so, cuts are applied which are based on the selection of reference [80]. The fundamental requirement demands two tau candidates as well as no light leptons. Furthermore, the following selection is applied:

- one or three charged tracks
- tau identification: medium [75]
- $0 \leq |\eta| \leq 1.37$ or $1.52 \leq |\eta| \leq 2.5$
- opposite charge of both tau leptons
- $E_T^{\text{Miss}} \geq 20\,\text{GeV}$
- $p_T^{\tau_0} \geq 40\,\text{GeV}$ and $p_T^{\tau_1} \geq 30\,\text{GeV}$
- $0.8 < \Delta R(\tau_0, \tau_1) < 2.5$

- $|\Delta\eta(\tau_0, \tau_1)| < 1.5$
- valid decay mode

As before, only events with a valid MMC reconstruction (mlnu3p_m > 0) are chosen in order to enable the comparison of different reconstruction techniques. Furthermore, true di-tau masses of 60 GeV–130 GeV for the Z and 100 GeV–140 GeV for the Higgs sample are required. The total number of events before and after the selection cuts are listed in table 5.1.

5.3 Data Processing

For training, the most suitable features need to be selected from the simulated data. Furthermore, an initial processing is advisable in order to ensure the maximum information extraction during training. The realization of these aspects will be addressed in this section.
Once the ANNs are trained, they are applied on real data. Therefore, it is crucial to provide the input samples on **reconstruction** level so that the network learns to deal with resolution effects. The target features, though, are considered on **truth** level, reflecting the ideal expectation. Consequently, an unbiased learning environment is provided that encourages a suitable mapping of the reconstructed inputs to the desired target values.

5.3.1 Input Data

The input data serves as basis for every training. Its proper choice influences the final performance decisively. Within the analysis, different sets of input variables will be investigated which either imitate the inputs of the MMC or go beyond.

On par with the MMC

One input set is chosen equivalently to the inputs of the current method for di-tau mass reconstruction, the MMC. The corresponding inputs are [42]:

- the transverse momentum p_T
- the pseudorapidity η
- the azimuthal angle ϕ
- the decay mode

} of the leading and subleading tau

- the reconstructed missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$
- the corresponding reconstructed azimuthal angle ϕ^{miss}

The last two inputs are essential because of their direct relation to the unmeasured neutrinos. The missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$, though, only returns the summed contribution of missing momenta (compare to equation 3.10) and provides no information about the individual neutrino properties. The frequency distributions of the listed variables are depicted in figure B.1, B.2,B.3, B.4 and B.5 in appendix B.

Beyond the MMC

The exchange of input variables as well as the expansion of the network by additional input nodes is straightforward and an outstanding advantage compared to the MMC. Using more input features provides more evaluable information which might improve the performance. Therefore, promising features beyond the MMC are considered. The investigated variables focus on the substructure of the tau decay and the kinematics of reconstructed jets:

- the transverse momentum p_T
- the pseudorapidity η
- the azimuthal angle ϕ

} of the charged and the neutral decay products of the tau

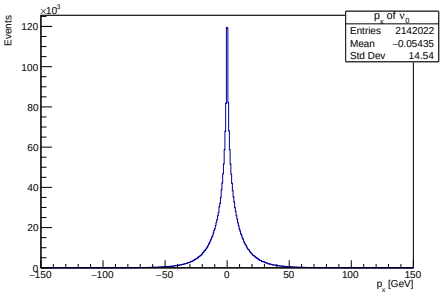
- the transverse momentum p_T
- the pseudorapidity η
- the azimuthal angle ϕ

} of the reconstructed jets

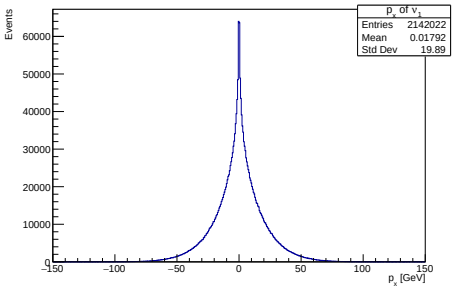
The motivation of their usage is given in section 8.2. The frequency distributions corresponding to the charged and neutral decay products are depicted in figure B.6, B.7 and B.8, while those of the jets are covered in figure B.9, B.10 and B.11 in appendix B.
Only the four momenta of the three most energetic jets are contained in the training sample although more jets might be involved in an event. Nevertheless the provided information content might be sufficient for training.

5.3.2 Target

As described in section 4.5.1, the neutrino momenta are used as target. Their frequency distributions are depicted in figure 5.3, 5.4 and 5.5 for the cartesian representation (p_x, p_y, p_z). Alternatively, the representation (p_T, η, ϕ) can be considered in figure 5.6, 5.7 and 5.8. The peculiarities of the different sets are discussed in section 7.1.
In order to receive the desired di-tau mass, relation 3.13 has to be applied to the presented neutrino momenta and the four momenta of the visible taus.

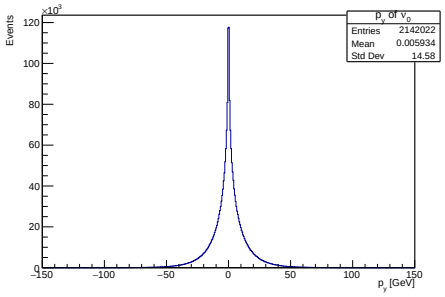


(a) True p_x of the neutrino associated to the leading tau.

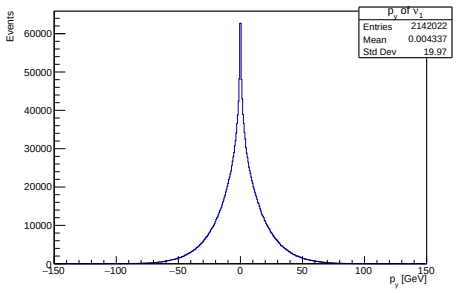


(b) True p_x of the neutrino associated to the subleading tau.

Figure 5.3: Neutrino momentum in x-direction. All events passing the selection described in section 5.1.1 are depicted.

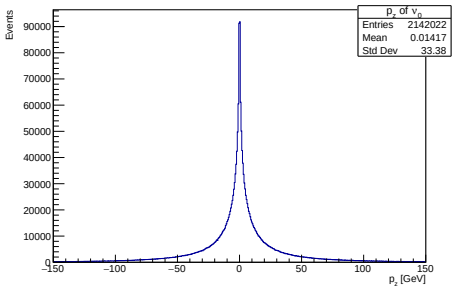


(a) True p_y of the neutrino associated to the leading tau.

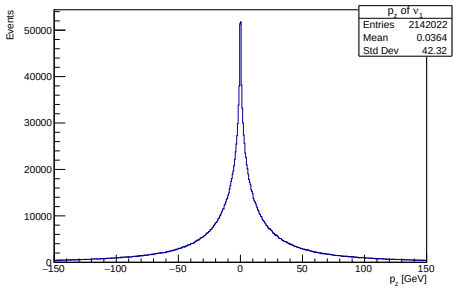


(b) True p_y of the neutrino associated to the subleading tau.

Figure 5.4: Neutrino momentum in y-direction. All events passing the selection described in section 5.1.1 are depicted.

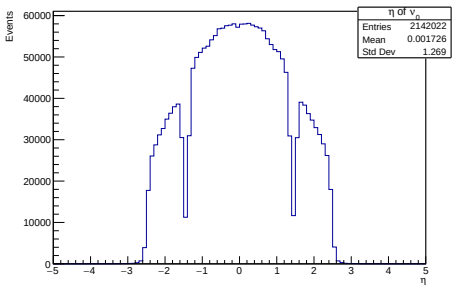


(a) True p_z of the neutrino associated to the leading tau.

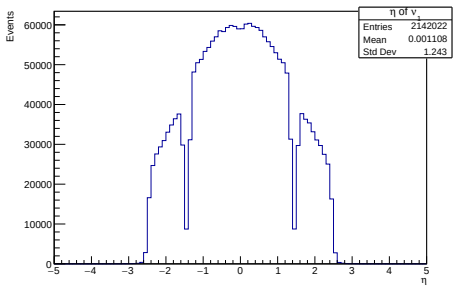


(b) True p_z of the neutrino associated to the subleading tau.

Figure 5.5: Neutrino momentum in z-direction. All events passing the selection described in section 5.1.1 are depicted.

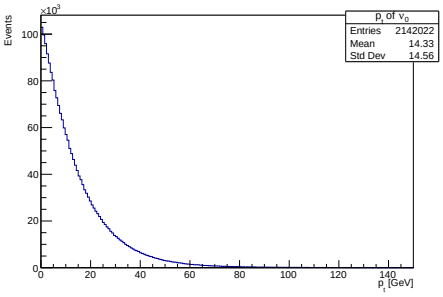


(a) True η of the neutrino associated to the leading tau.

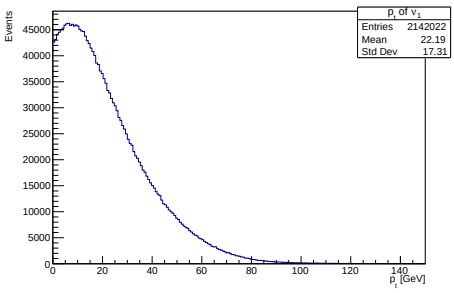


(b) True η of the neutrino associated to the subleading tau.

Figure 5.6: True pseudorapidity of the two neutrinos. All events passing the selection described in section 5.1.1 are depicted.

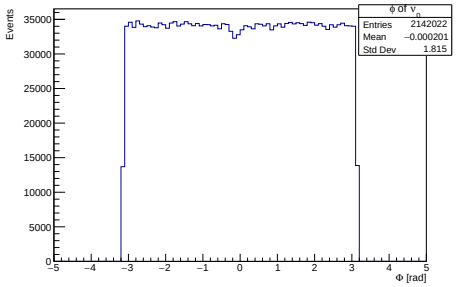


(a) True p_T of the neutrino associated to the leading tau.

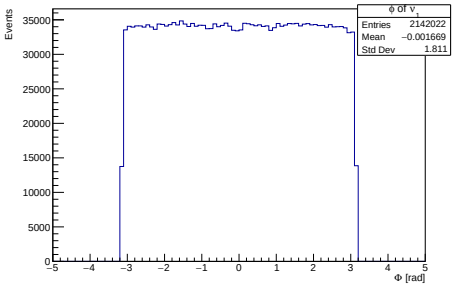


(b) True p_T of the neutrino associated to the subleading tau.

Figure 5.7: True transverse momentum of the two neutrinos. All events passing the selection described in section 5.1.1 are depicted.



(a) True ϕ of the neutrino associated to the leading tau.



(b) True ϕ of the neutrino associated to the subleading tau.

Figure 5.8: True azimuthal angle of the two neutrinos. All events passing the selection described in section 5.1.1 are depicted.

5.3.3 Preprocessing

Three different sets of data are usually used for the training of ANNs. First of all, a data set of high statistics is necessary in order to train the model and to adapt the weights and biases. This fundamental data set is called the **training sample**. In order to choose the best performing model afterwards and to recognize overfitting, a second set is applied, the **validation sample**. The last set, the **test sample**, is used as confirmation of the chosen model with unseen data. All sets have to reproduce the original probability density functions.
In order to fulfill these requirements, the events of the γ^* sample are split in a ratio 80:10:10. The selection is performed on the basis of the uniform distributed azimuthal angle of the neutrino associated to the leading tau. It is well defined in a range from $-\pi$ to π which also serves as interval for the subdivision into 60 bins. Per bin, 10% of the events are assigned to the test sample, another 10% to the validation sample and the remaining events are used for training. In order to ensure a balanced assignment, the events of each bin were shuffled beforehand. This procedure results in 214205 events for the two small sets and 1713612 for the large training set. The frequency distributions of

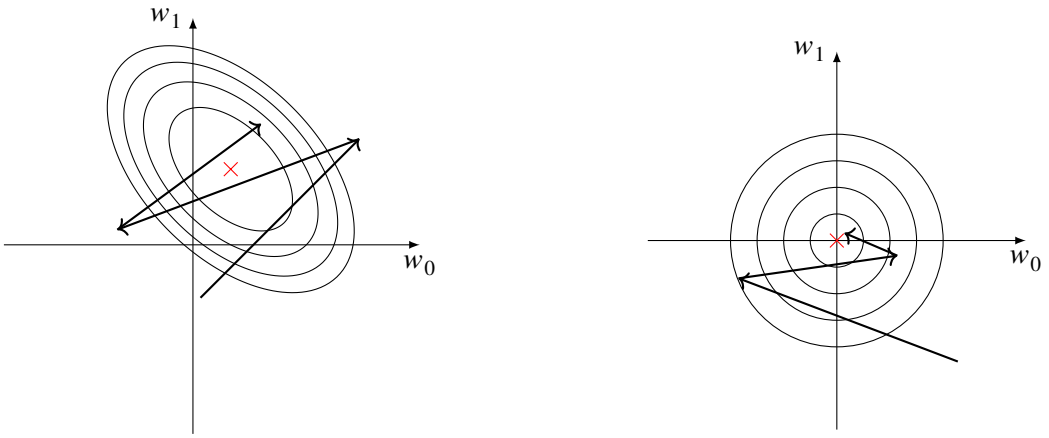
the target variables (p_T, η, Φ) are exemplary depicted for the training, validation and the test set in figure B.12–B.20. The expected frequency distributions are reproduced for every set as required.

5.3.4 Standard Scaler

Often, the input variables scale in different ranges. Although the range differences might be balanced by the weights during the training, the disproportion complicates a fast convergence for gradient based optimizers. Therefore, it is beneficial to apply a standardization beforehand. One possibility is to center the sample at zero by subtracting its mean μ and provide a standard deviation of 1 by dividing by the sample's standard deviation σ [63]

$$z = \frac{x - \mu}{\sigma}. \tag{5.1}$$

The influence of feature standardization on the optimizer's performance is illustrated in figure 5.9. It shows a two-dimensional example with the internal weights w_1 and w_2. The contours represent regions of equal losses and the arrows indicate the learning process of the optimizer. The final goal is to reach a minimal loss, indicated by the red cross. While in figure 5.9(a) no convergence can be achieved on the unprocessed data, the minimum is reached on the standardized input after a few steps as depicted in figure 5.9(b) [63].
In order to profit from this effect, the standardization is applied to all input variables. Exemplary the scaled frequency distributions of the input features on par with the MMC (section 5.3.1) are depicted in figure B.21–B.25 in appendix B.



(a) Learning for unprocessed input data.

(b) Learning for standardized input data.

Figure 5.9: Influence of feature standardization on the loss minimization. The contours show regions of equal losses for a two-dimensional example (w_1 and w_2 represent the corresponding weights of the model). Inspiration from [63] created with [12].

CHAPTER 6

Basic Setup & Performance Measurement

In the following, a basic ANN setup will be described. In this context, one architecture and a set of hyperparameters will be introduced which are used for fundamental, task related investigations. Using this setup as starting point, the model can be further developed.
In this process, it is crucial to rate different setups and to ensure comparability. Therefore, diverse figures of merit will be introduced as measure of performance.

6.1 The Setup

In this section, a basic setup will be introduced whose components will be used if not mentioned otherwise. The model is based on TensorFlow libraries [81] and the Keras [82] interface.
The fundamental architecture comprises 6 output nodes, 3 for each neutrino. The advantage of 6 nodes compared to one node returning a six-dimensional output is the possibility to apply an individual loss function for each node. The final loss is the sum of the single contributions.
The dimensionality of the input nodes corresponds to the number of features which are selected for training. Furthermore, a certain amount of unconnected input nodes are used. They receive variables of the tau decay products needed for calculations for the constraint realization as described in section 4.5.2 and 4.5.3. The following hyperparameters are chosen as default:

- hidden layers: 4
- nodes per layer: 16
- optimizer: adam
- learning rate: 0.001
- loss function: mean-squared-error
- batch size: 32
- epochs: 200

Additionally, the weights and biases have to be initialized before training. For the weights, the *LecunNormal*-initializer is applied, drawing random values from a truncated normal distribution. The biases, though, are initialized with zero. Furthermore, several callbacks are considered. They are called at different stages of the training and enable the interference in the internal learning process.

- ReduceLROnPlateau():
  This callback keeps a chosen metric under surveillance and reduces the learning rate as soon as

the learning stagnates for a certain amount of epochs.
Halving the learning rate if the mean-squared-error stays on a plateau over 15 epochs turned out to improve the performance significantly.

- TerminateOnNaN():
  Terminates the training process if the loss turns NaN (Not a Number).

- History():
  Records loss and metric values during the training.

Furthermore, a periodicity correction on the azimuthal angle as well as a loss scaling will be applied if not stated otherwise. They are introduced in section 7.3 and 7.4.

6.2 Figures of Merit & Calibration

During the development of a model, many aspects can be tuned as the general architecture, hyperparameters, realization of physical constraints etc.. In order to determine a final setup, figures of merit are introduced which provide comparability of different model's performances. Since the evaluation depends on the point of view, there is not **the** best choice. The following aspects, though, enable the differentiation of models and might emphasize setups which fulfill the demands of a proper di-tau mass reconstruction.
In order to illustrate the idea behind the different rating measures, the basic model of section 6.1 was trained with the MMC input variables on truth level. True inputs are unrealistic and a model trained on idealized data is not applicable to real measurements. The truth level inputs, though, improve the learning achievement. Consequently, the final di-tau mass estimations are less affected by diverse overlapping effects which might reduce the comprehensibility of the general concepts.
Furthermore, a calibration method will be introduced which enables the comparability of different techniques.

6.2.1 Correlation Factor

One possible measure of performance are correlation coefficients ρ which describe the degree of linear relation between two random variables X and Y. The coefficients are defined in an interval from -1 to 1. While the lower bound represents a total negative linear correlation, the upper bound describes a total positive one. A correlation factor of zero represents total independence of the two variables X and Y.
Mathematically, the correlation coefficient is described by the ratio of the covariance and the product of the standard deviations of the variables. The former is the expectation value of the product of the differences between the variables and their individual means μ_i [83]

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_x \sigma_y}. \tag{6.1}$$

In the context of di-tau mass estimation, the degree of linear relation between the reconstructed di-tau mass and the true value is a valid figure of merit. Ideally, a complete positive correlation of +1 is achieved which corresponds to an exclusive population of the diagonal in figure 6.1. This

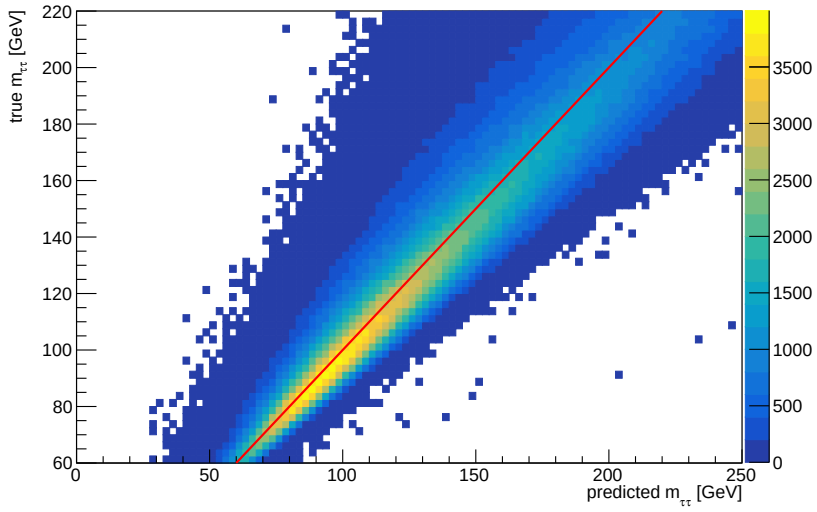


Figure 6.1: Scatter plot of the true and the determined di-tau mass. The training was performed on the basic model (see section 6.1) with truth level input. In order to provide an example of high statistics, the evaluation was performed on the events selected for training. The red line indicates the desired population of the diagonal. Correlation factor: 0.93

scatter plot illustrates the pairs of estimated and true masses on the x- and y-axis respectively. If not stated otherwise, estimated or predicted masses refer to the di-tau masses determined from the predicted neutrino momenta and the visible taus according to equation 3.13. The achieved correlation corresponds to $\rho = 0.93$. As desired, a predominant population of the diagonal, indicated by the red line, is observable.

6.2.2 Loss

The loss development during the training of ANNs needs to be considered as an indirect figure of merit since the absolute loss values of different models cannot be compared directly. The loss depends on the choice of the loss function, the value range of the target variables, the weighting of possible penalty terms (see section 4.5.2) etc.. Consequently, it is a setup dependent quantity. Nevertheless, its development provides insight into the stability of the training and may reveal effects like overtraining. Thus, the loss development rates a network on the basis of its general applicability and might reveal deficits of the setups.
In a model consisting of several output nodes, the total loss is the sum of the individual contributions. Its value is saved after each epoch but the determination of the training and validation loss differs. While the training loss is evaluated after each batch and the loss of the final batch of the epoch is saved, the validation loss is determined with the full sample at the end of an epoch. An exemplary loss development is depicted in figure 6.2. It illustrates the training achievement of the basic model using truth level inputs. Like expected, the training loss decreases with progressing epochs. The steps in the curve result from a learning rate reduction induced by the callback *ReduceLROnPlateau*. Furthermore, the validation loss is depicted in blue. Its development is more unstable but the overall

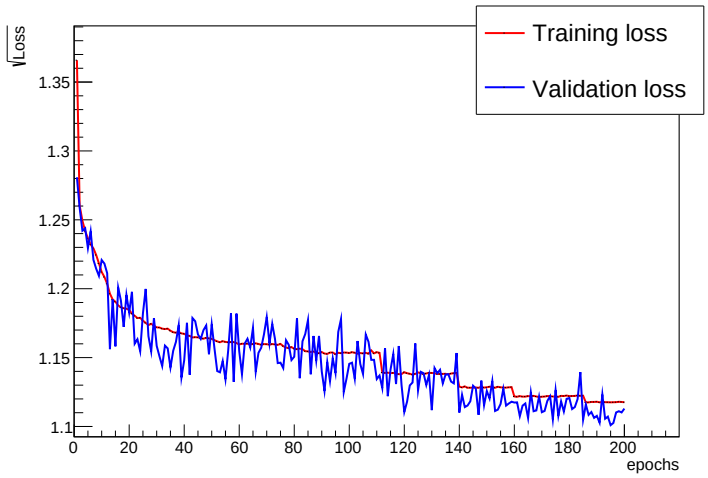


Figure 6.2: Development of the total loss over 200 epochs. The training loss as well as the validation loss are depicted. The basic model (see section 6.1) was trained with truth level inputs.

trend is similar, as desired.

6.2.3 Relative Resolution

The relative resolution is an instructive measure of performance. In the following, it is defined as the scaled deviation between estimated and true di-tau mass. Ideally a gaussian distribution is assumed and its mean represents the bias of the estimation

$$bias = \mu\left(\frac{m_{\tau\tau}^{\text{pred}} - m_{\tau\tau}^{\text{truth}}}{m_{\tau\tau}^{\text{truth}}}\right). \tag{6.2}$$

The reconstructed di-tau mass serves as discriminator between signal and background. Therefore, not the overall relative resolution but the one in the mass region of the Z and Higgs boson is of particular interest. Thus, a binned representation is chosen. The relative resolution is determined in 10 GeV bins of the true di-tau mass. The corresponding distributions are desirably centered at zero which implies an unbiased di-tau mass estimation. Apart from the bias, another figure of merit can be deduced from the widths of the distributions. Generally, small widths are preferred which imply small overlaps between the reconstructed mass peaks of Z and Higgs. In this thesis, the half 68%-quantile widths of the relative resolution are considered as one of the central measures of performance.
A quantile of a random variable is the threshold value p for which x_p% of the values are smaller [83].

$$P(x \leq p) = x_p \tag{6.3}$$

In order to define a 68%-quantile width, the threshold values are determined for which 16% of the total set are larger or smaller. Consequently, the remaining interval covers 68% of the values. The given definition imitates the 1σ interval of a normal distribution.

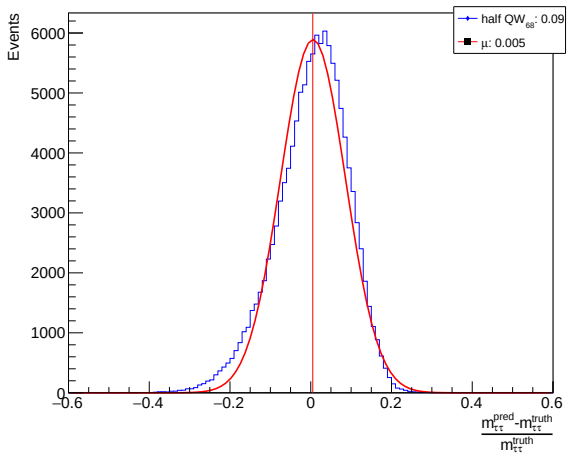


Figure 6.3: Relative resolution (blue) for true di-tau masses between 90 GeV and 100 GeV. The mean of a gaussian fit (red line) is considered as bias, while the half 68%-quantile width is based on the original distribution (blue). The example is based on the basic model trained on truth level inputs.

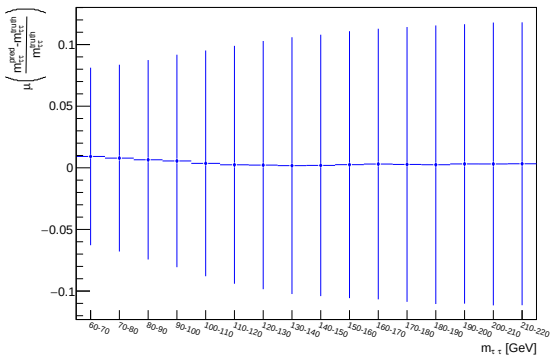
Figure 6.3 shows the relative resolution for true di-tau masses between 90 GeV and 100 GeV. The example is based on the basic model trained on truth level inputs. The mean of a gaussian fit is assumed as bias. In order to consider deviations from the gaussian fit in the determination of the width, not the standard deviation of the Gaussian but the half 68%-quantile width of the overall distribution is considered.

Determining the means and the widths in all 10 GeV bins of the provided di-tau mass range from 60 GeV to 220 GeV offers a mass dependent measure of performance. Figure 6.4(a) represents the determined means with the corresponding half 68%-quantile widths as uncertainties for the training of the basic model with truth level inputs. In the following, this kind of plot will be designated as *binned resolution plot*. The exclusive description of the half 68%-quantile widths is shown in figure 6.4(b). The presented example satisfies the demands of small widths and a bias consistent with zero.

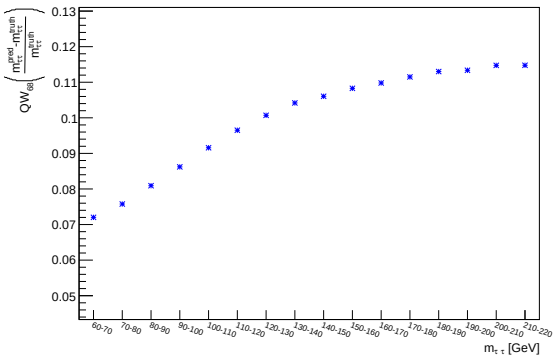
6.2.4 Calibration

Di-tau mass reconstruction algorithms dealing with resolution effects often perform biasedly. This phenomenon is observable in figure 6.5. It shows the binned resolution plot for reconstructed di-tau masses of the MMC using the mlnu3p approach. Analogously to the previous description in section 6.2.3, the gaussian means and the half 68%-quantile widths of the relative resolution plots are determined in 10 GeV bins of the true mass.
In the lower mass region, positive biases occur which express an overestimation of the di-tau masses. In the upper mass region, though, they are underestimated by the MMC. These are already instructive insights into the reconstruction performance. Different reconstruction techniques might, however, be affected by different biases which complicates the direct comparison of their 68%-quantile widths.

Therefore, a calibration needs to be applied which centers the resolution plots at zero. The desired elimination of the biases is realized as soon as the distributions of reconstructed and true masses are consistent in each bin. As a measure of compatibility, different quantities like the mean, the median or the peak can be considered, depending on the shape of the distribution. In the following, the stepwise



(a) Binned resolution plot: Per bin, the mean of a gaussian fit as well as the half 68%-quantile widths of the underlying distribution, representing the uncertainty, are considered.



(b) Half 68%-quantile widths of the relative resolution.

Figure 6.4: Binned resolution plot in 10 GeV bins of the true di-tau mass (a). The half 68%-quantile widths are exclusively depicted in (b). The results are based on the basic ANN trained on truth level inputs. The evaluation was performed on the events selected for testing.

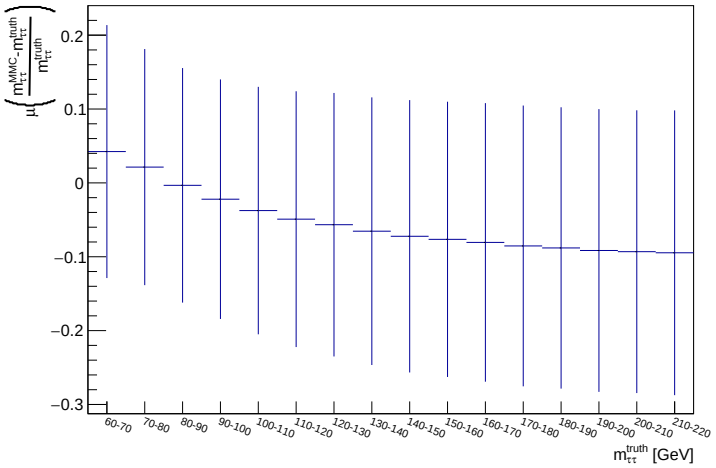


Figure 6.5: Binned resolution plot based on the reconstructed di-tau masses of the MMC using the mlnu3p approach. The single bins represent 10 GeV intervals of the true di-tau mass.

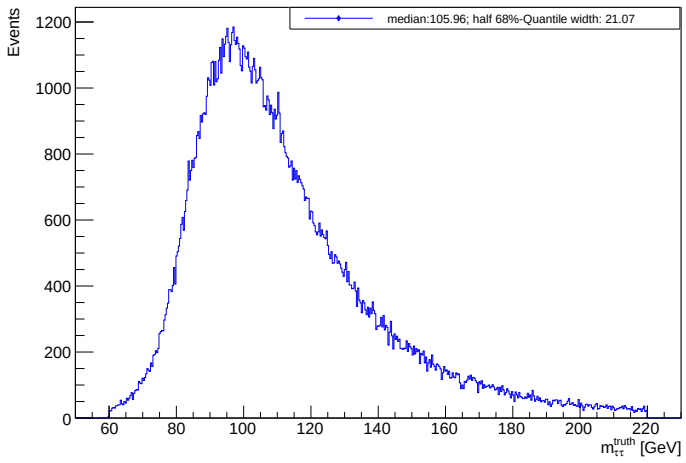


Figure 6.6: True di-tau masses for reconstructed values between 90 GeV and 100 GeV (MMC). As a measure of compatibility between truth and reconstruction, the median of 105.96 GeV as well as the half 68%-quantile width of 21.07 GeV are considered.

elimination of the bias will be presented for reconstructed di-tau masses of the MMC.
Figure 6.6 shows the frequency distribution of true di-tau masses for reconstructed values between 90 GeV and 100 GeV. The general shape is characterized by an asymmetry and a large tail towards high masses. The asymmetry becomes even more pronounced for bins close to the edges of the true mass interval (60–220 GeV). Hence, the median emerged as the best description over all bins. It provides a compromise between the peak position which does not deal with tails at all and the mean which is rather sensitive to large outliers. Nevertheless, the outer bins have to be treated cautiously since boundary effects and sharp cuts prevent smooth transitions to the adjacent bins.

Determining the medians of the true mass distributions in 10 GeV mass intervals of the reconstructed di-tau masses and plotting them against the bin center leads to the overall behavior of figure 6.7. While the uncertainty in x-direction corresponds to the bin width, the half 68%-quantile widths of the true mass distributions are considered as uncertainties in y-direction. For an unbiased di-tau mass reconstruction, a line through the origin with a slope of 1 is expected. The observable development, though, is rather polynomial. In order to achieve a proper description, a polynomial of order 4

$$m_{\tau\tau}^{\text{truth, calc}} = a \cdot \left(m_{\tau\tau}^{\text{MMC}}\right)^4 + b \cdot \left(m_{\tau\tau}^{\text{MMC}}\right)^3 + c \cdot \left(m_{\tau\tau}^{\text{MMC}}\right)^2 + d \cdot \left(m_{\tau\tau}^{\text{MMC}}\right) + e \tag{6.4}$$

was fitted to the graph in figure 6.7. In a first calibration step, the fit parameters a, b, c, d and e are used in order to determine a true mass value for every reconstructed mass by applying equation 6.4. The deviation of this value to the argument of the function will be used for correction.

$$m_{\tau\tau}^{\text{corrected}} = m_{\tau\tau}^{\text{MMC}} + \Delta m_{\tau\tau} \qquad \text{with} \qquad \Delta m_{\tau\tau} = m_{\tau\tau}^{\text{truth, calc}} - m_{\tau\tau}^{\text{MMC}} \tag{6.5}$$

Additionally, it is investigated whether the desired consistency is caused by the correction. Therefore, the medians of the true di-tau masses in bins of the corrected masses and the reversed case are considered in figure 6.8. A linear fit was applied for both cases. The former reproduces the desired

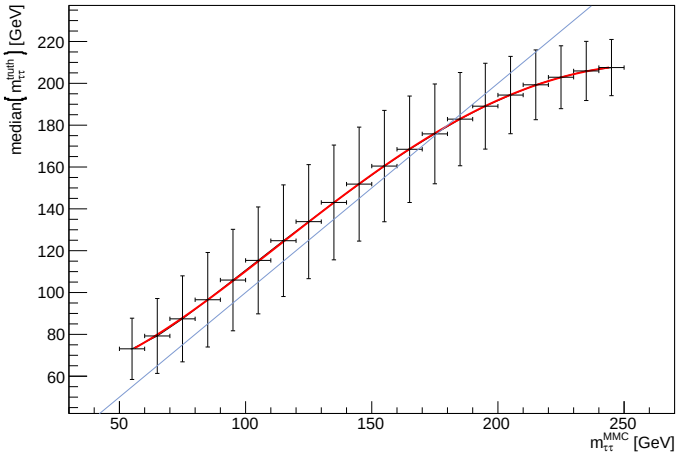
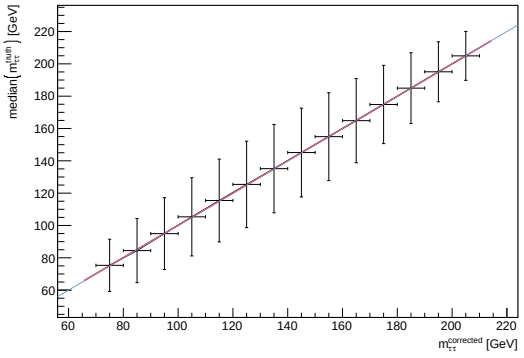
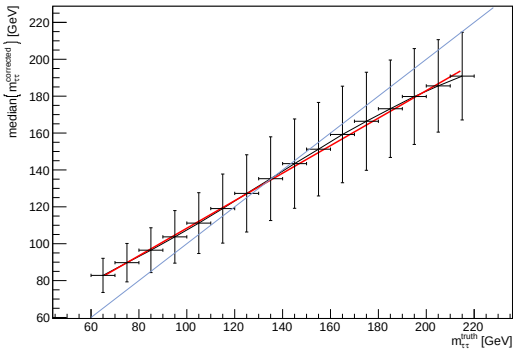


Figure 6.7: True vs. reconstructed di-tau mass (MMC). The medians of the true di-tau mass distributions are plotted against the corresponding bin center of the reconstructed mass. The red curve represents a polynomial fit of order 4. The desired linear development is shown in gray.



(a) Medians of the true masses in bins of the corrected masses.



(b) Medians of the corrected masses in bins of the true masses.

Figure 6.8: Calibration cross check. In (a), the medians of the true masses are depicted in 10 GeV bins of the corrected masses. A linear fit results in a slope of 0.999 and a Y-intercept of 0.21. In (b) the medians of the corrected masses are shown in 10 GeV bins of the true masses. The linear fit follows a slope of 0.746 and a Y-intercept of 33.79.

linear dependence with a slope of 1 and a Y-intercept compatible with zero, as depicted in figure 6.8(a). This observation is an obligatory consequence of the applied calibration. The reversed view of the corrected masses in bins of the true masses, though, deviates from the desired dependency illustrated by the gray line (figure 6.8(b)). With a slope of 0.746 and a Y-intercept incompatible with zero, no satisfactory bias correction was achieved for the corrected masses.

Therefore, a second calibration step is performed which is based on the linear fit applied to the medians of the corrected masses in bins of the true masses (figure 6.8(b)). Since corrections cannot be

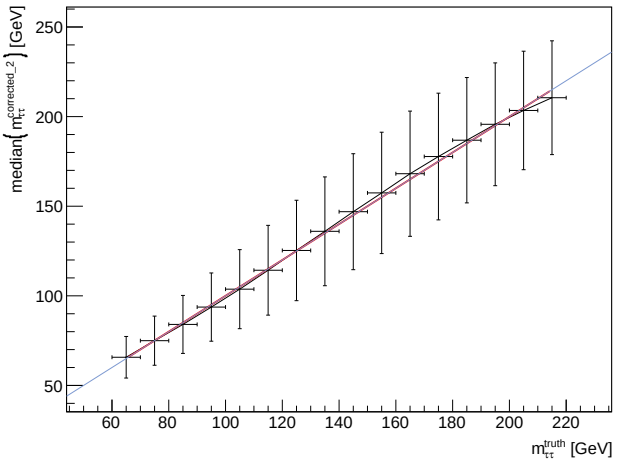


Figure 6.9: Cross check of the two-stage calibration. After the two described calibration steps, the medians of the corrected masses are determined in bins of the true mass and plotted against the bin center. The linear fit yields a slope of 1 and a Y-intercept of 0.05 which fulfills the required proportionality (gray line).

performed on true values, the reverse fit function has to be considered.

$$m_{\tau\tau}^{\text{corrected}} = p \cdot m_{\tau\tau}^{\text{truth}} + t \qquad \Rightarrow \qquad m_{\tau\tau}^{\text{truth}} = \frac{m_{\tau\tau}^{\text{corrected}} - t}{p}. \tag{6.6}$$

For each corrected di-tau mass the corresponding true value can be determined by using the fit parameters p and t. The second correction is performed analogously to equation 6.5 and causes the desired proportionality, as depicted in figure 6.9. It shows the medians of the two-stage corrected di-tau masses in bins of the true mass. A final slope 1 and Y-intercept 0.05 was achieved.

Although, a linear fit describes the data sufficiently well, slight deviations from the functional relation can be observed in the upper mass regime. They might result from bin-to-bin migrations.

Using the calibrated masses in the determination of the relative di-tau mass resolution results in the outcome of figure 6.10. The overall aim of an unbiased di-tau mass reconstruction is achieved. Thus, the developed calibration can be applied to different techniques, so that their 68%-quantile widths can be compared directly.

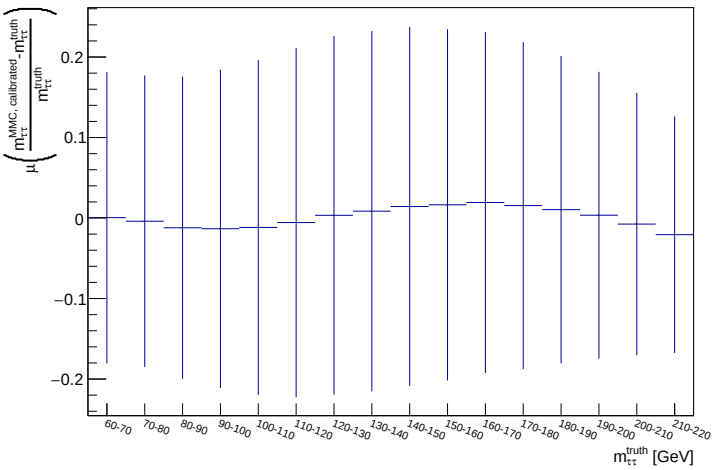


Figure 6.10: Binned resolution using the calibrated di-tau masses provided by the MMC.

CHAPTER 7

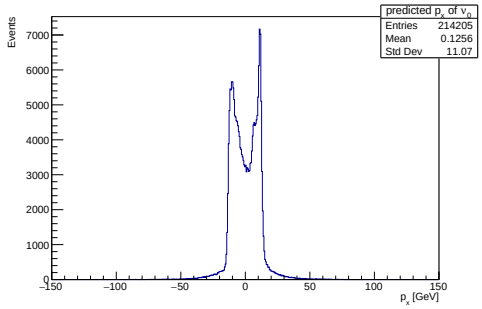
Preparatory Studies

The motivation for the application of ANNs in the di-tau mass reconstruction is already detailed in sections 4.5 and 4.6. Since this thesis builds on the initial exploration of the topic in reference [4], more extensive approaches using multiple outputs for the prediction of the neutrino momenta as well as a constrained learning are investigated. However, with the increasing complexity of the setup many overlapping effects occur which need to be at least partially disentangled in order to ensure a proper reconstruction performance. Aspects like boundary effects, periodicities of variables, ambiguities in physical relations, resolution effects etc. need to be considered. In order to approach these different effects step by step, a toy model study was performed additionally and will be introduced in section 7.5. Generally, this chapter details the investigation of different aspects influencing the learning process. Their understanding is essential for the application of ANNs in the di-tau mass reconstruction described in chapter 8. The following results are based on models trained on the input variables imitating the MMC (see section 5.3.1) and follow the basic setup described in section 6.1 if not stated otherwise. The evaluation is performed on the events selected for testing.

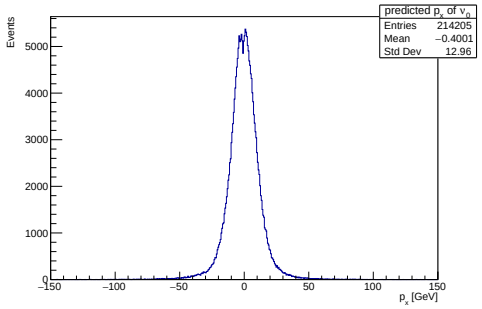
7.1 The Target: Influence of the Momentum Representation

The neutrino momenta, generally represented in a cartesian coordinate system as (p_x, p_y, p_z), were chosen as target. Considering the mathematical relation of equation 3.8, the analysis can also rely on the triplet (p_T, η, ϕ) as an alternative choice. Although, both representations have the same information content, the network may perform differently since the loss, and consequently the learning achievement depends on the general shape of the underlying frequency distribution and the value range. The investigation of the occurring effects reveals fundamental issues for the di-tau mass reconstruction.

In an initial attempt, the set (p_x, p_y, p_z) was used as target. The predictions of the network, however, are characterized by unexpected frequency distributions, as depicted in figure 7.1(a)–7.3(a) representing the estimated neutrino momenta associated to the leading tau. Especially the predictions in x- and y-direction suffer from collapses around zero where a peak is expected (see figure 5.3–5.5). The attempts to optimize the network's hyperparameters did not eliminate the effects. Only the training on truth level inputs reduced the observed deficits in the prediction significantly, as represented by the figures 7.1(b)–7.3(b). The corresponding frequency distributions of the neutrino originating from the subleading tau are depicted in figure C.1–C.3 in appendix C .

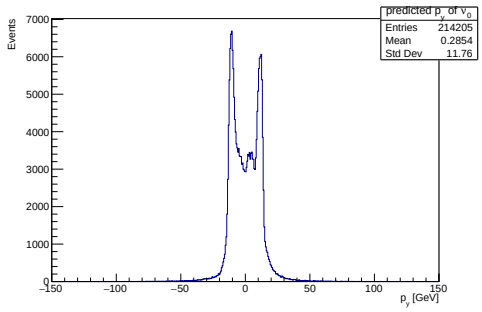


(a) Predicted p_x. The training was performed with inputs on **reconstruction level**.

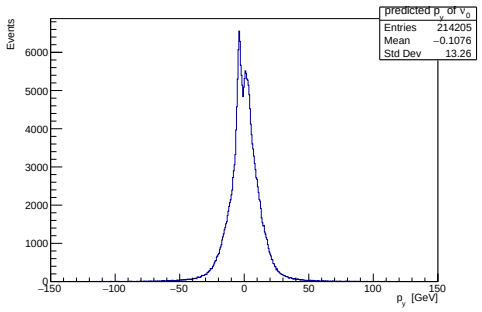


(b) Predicted p_x. The training was performed with inputs on **truth level**.

Figure 7.1: Predicted momentum in x-direction for the neutrino originating from the leading tau. The training was performed on the basic setup introduced in section 6.1. The presented prediction is based on the events selected for testing.

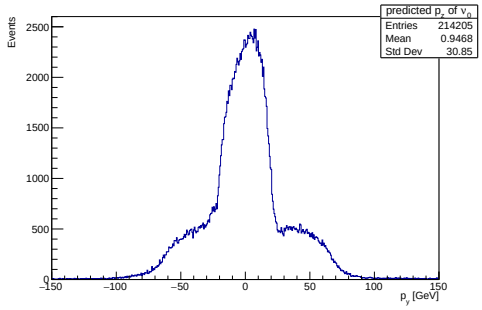


(a) Predicted p_y. The training was performed with inputs on **reconstruction level**.

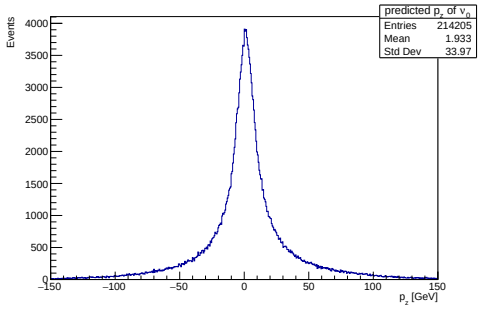


(b) Predicted p_y. The training was performed with inputs on **truth level**.

Figure 7.2: Predicted momentum in y-direction for the neutrino originating from the leading tau. The training was performed on the basic setup introduced in section 6.1. The presented prediction is based on the events selected for testing.



(a) Predicted p_z. The training was performed with inputs on **reconstruction level**.



(b) Predicted p_z. The training was performed with inputs on **truth level**.

Figure 7.3: Predicted momentum in z-direction for the neutrino originating from the leading tau. The training was performed on the basic setup introduced in section 6.1. The presented prediction is based on the events selected for testing.

Since the observed effects were significantly reduced by the training on truth level inputs, resolution effects seem to play a decisive role. The neutrino momenta of the transverse plane are highly correlated with the missing transverse energy E_T^{Miss} which is affected by a low resolution as described in section 3.4.5. Consequently, the network needs to learn the general physical connection between the missing transverse energy and the neutrino momenta as well as a correction for deficits in the E_T^{Miss} reconstruction. Another aspect is the allocation of momentum to the individual neutrinos. Since the missing transverse energy corresponds to the sum of the neutrino momenta, the contribution of an individual neutrino is only indirect available and has to be learned.

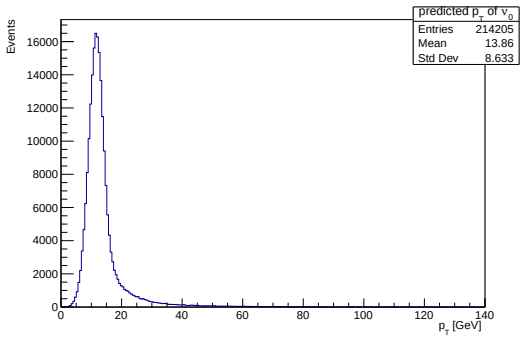
The impact of these effects might be reduced by the alternative target set (p_T, η, ϕ). The reparametrization of the information content might be beneficial to training the network. While p_x as well as p_y are predominantly correlated with the missing transverse energy, their dependence is combined in the transverse momentum p_T. The remaining features η and ϕ are highly correlated with the angles of the visible tau. At this point, resolution effects might play a minor role which provides a suitable learning environment for these variables. Additionally, general aspects considering the distribution shape are beneficial. The uniform distribution of the azimuthal angle introduces no specific characteristics which facilitates the learning process. Furthermore, methodological aspects favor the choice of (p_T, η, ϕ). As already mentioned in section 4.5.3, the mass constraint can be solved unambiguously for the transverse momentum but not for the momenta in the cartesian representation. Therefore, considering the previous discussion, the set (p_T, η, ϕ) was chosen as target. Further issues arising from its individual characteristics will be part of the following sections.

7.2 Transformation of the Transverse Momentum

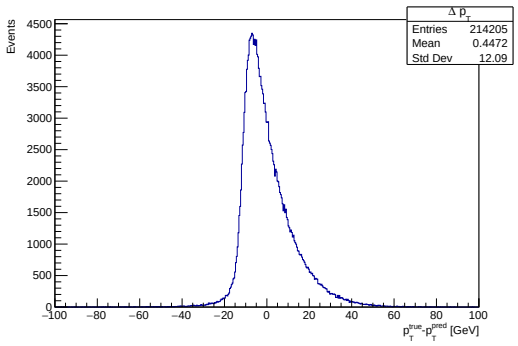
As already discussed in section 7.1, it is a complex task for the ANN to transfer knowledge of the missing transverse energy to the neutrino momenta in x- and y-direction. Since p_T is the momentum in the transverse plane, the same should apply for its estimation if the assumption is correct. The detailed investigation of this target variable will be topic of the following section.

The expected frequency distribution of the transverse momentum carried by the neutrino originating from the leading tau is depicted in figure 5.7(a). The general shape is characterized by a maximum at zero and a decreasing distribution resulting in a large tail towards high momenta. The prediction of the neural network, though, depicted in figure 7.4(a), does not reproduce the expectation. Neither the maximum at zero nor the large tails can be observed. Figure 7.4(b) represents the corresponding resolution and also reflects the described deficits by its large width and its tail representing a dominant underestimation of transverse momenta. A similar behavior can be observed for the predicted transverse momentum of the neutrino originating from the subleading tau, depicted in figure C.4 in appendix C.

The observation is in agreement with the results of section 7.1 and the hypothesis of limited learning due to the E_T^{Miss} resolution. Especially the absence of predicted values at zero matches the observation of figure 7.1(a) and 7.2(a).
However, further aspects concerning the characteristics of the target distribution might have an impact on the outcome. Avoiding zero might be supported by boundary effects induced by the value range of the variable. According to the hypothesis, predictions in direct proximity to zero are avoided and negative solutions are excluded completely since only positive values are presented as desired outputs



(a) Frequency distribution of the predicted p_T.



(b) Resolution of p_T.

Figure 7.4: Predicted transverse momentum of the neutrino originating from the leading tau and its resolution. The prediction is based on the events selected for testing.

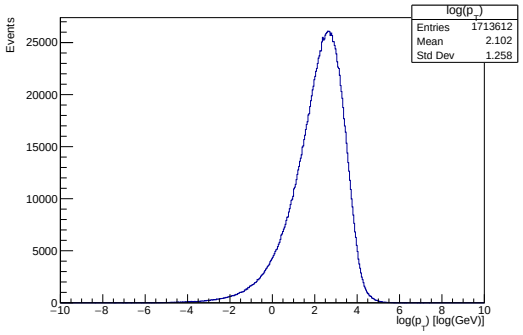
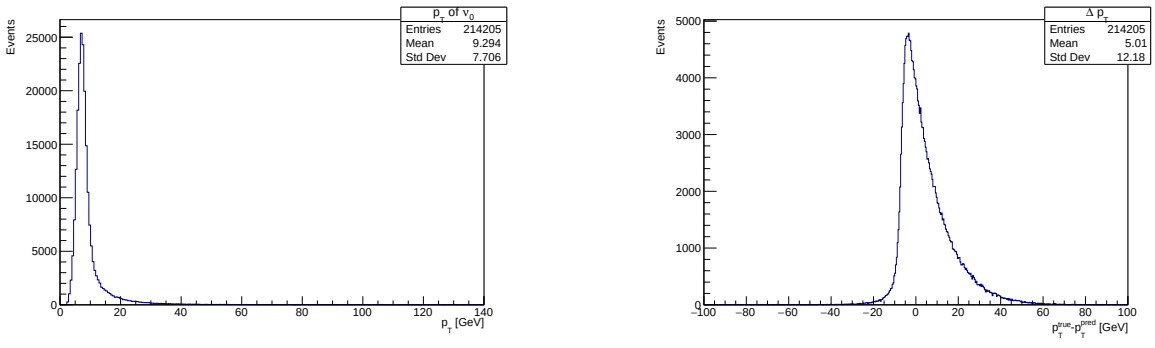


Figure 7.5: Logarithmic transformation of the transverse momentum. The depicted frequency distribution belongs to the neutrino originating from the leading tau. The events selected for training are depicted.

to the network. Furthermore, the prediction of high momenta might be prevented due to large losses induced by false outlier predictions. In order to address this issue, modified loss functions were tested which are more suited for the modellation of tails than the *mean-squared-error*. Exemplary, the *Log-Cosh-loss* described in section 4.3.3 was considered. These different approaches, however, did not lead to significant improvements.
Therefore, another attempt considering a direct transformation of the transverse momentum was performed, in order to produce a more symmetric starting point. Exemplary, the study of a logarithmic transformation will be presented. The application of the natural logarithm on the transverse momentum is depicted in figure 7.5 for the neutrino originating from the leading tau. The transformation eliminates the large tail by squeezing the momenta in a smaller value range. Furthermore, a more symmetric distribution is achieved and negative values are introduced.

However, no significant improvements were achieved for the training with the logarithm of the transverse momentum as target. The backtransformed prediction is depicted in figure 7.6. The applied backtransformation corresponds to the exponential of the model's output. Again, the tails towards high momenta are hardly represented and the lower bound is not reached as well. The overall performance is comparable to the one achieved by the model considering the original p_T distribution as target,



(a) Backtransformed prediction of p_T.

(b) Resolution of the backtransformed p_T.

Figure 7.6: Backtransformed prediction of the transverse momentum. The basic model was trained with the logarithm of the transverse momentum as target. Thus, the backtransformation corresponds to the exponential of the model's output. The prediction was performed on the events selected for testing and corresponds to the neutrino originating from the leading tau.

depicted in figure 7.4. Similar results can be observed for the transverse momentum of the neutrino associated to the subleading tau depicted in figure C.5 in appendix C.

Since different kinds of transformations within the loss function as well as directly on the transverse momentum distribution did not have large impact on the final performance, the E_T^{Miss} resolution is still the limiting factor in the prediction. The issue cannot be resolved for the moment and might be an interesting subject for future studies.

7.3 Loss Contributions

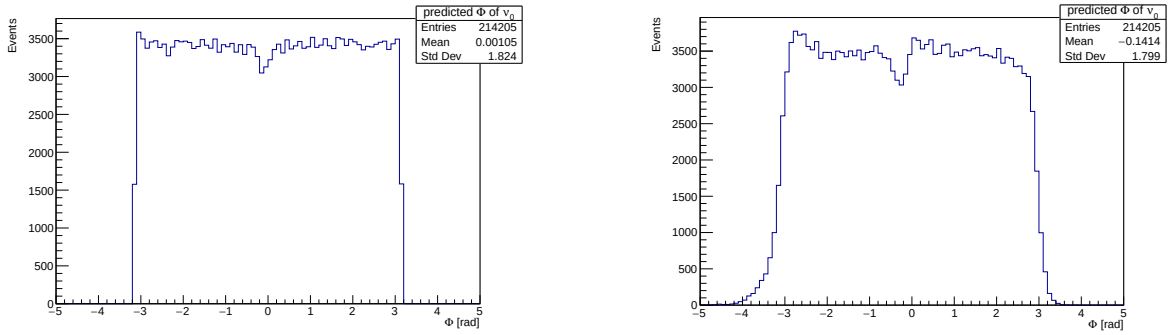
The total loss $\mathcal{L}_{\text{total}}$ of a neural network consisting of multiple output nodes is the combination of the individual loss contributions $\mathcal{L}_{\text{i}}$. For the chosen setup, six summands have to be considered, namely three for each neutrino.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{p_{T,0}} + \mathcal{L}_{\phi_0} + \mathcal{L}_{\eta_0} + \mathcal{L}_{p_{T,1}} + \mathcal{L}_{\phi_1} + \mathcal{L}_{\eta_1} \tag{7.1}$$

For compound losses, the scales of the single components are of interest. If the contributions of some nodes dominate the others, the training focuses on their reduction. Consequently, an imbalanced learning occurs which neglects the lowly weighted output nodes.
From the chosen targets, the losses of the transverse momenta contribute with mean-squared-error values of about 200 during the first training epochs while the losses of the angles are close to one. Thus, the single components have to be scaled to similar values. For the sake of simplicity, the loss of the transverse momentum is scaled down.

$$\frac{1}{200}\mathcal{L}_{p_{T,i}} \approx \mathcal{L}_{\phi_i} \approx \mathcal{L}_{\eta_i} \approx 1$$

The impact of an unscaled training is depicted in figure 7.7. The prediction of the azimuthal angle is shown with 7.7(a) and without 7.7(b) loss scaling. While the former reproduces the expected uniform distribution within the valid range, the latter distribution is characterized by blurred edges and



(a) Prediction **with** loss scaling.

(b) Prediction **without** loss scaling.

Figure 7.7: Influence of a loss scaling on the predicted azimuthal angle. The prediction was performed on the events selected for testing and corresponds to the neutrino originating from the leading tau.

bumps in the plateau region. Furthermore, a hole can be observed at zero which will be addressed in section 7.4. In both cases, a periodicity correction was applied which will be introduced in section 7.4 as well. Since the loss scaling induces a more balanced training, it will always be applied unless stated otherwise.

7.4 Periodicity of the Azimuthal Angle

The azimuthal angle Φ is a 2π-periodic quantity defined in the range from $-\pi$ to π. This property highly affects the learning achievement as depicted in figure 7.9(a). It represents the prediction of the azimuthal angle using the basic setup trained with the mean-squared-error as loss. Ideally, a uniform distribution is expected as depicted in figure 5.8. The prediction, though, is characterized by two distinctive peaks near the edges. This effect is a measure of the ANN in order to avoid large losses. Figure 7.8 illustrates the angle assignment in space which might improve the comprehension of the observed effects. For true angles close to π or $-\pi$, a reasonable prediction of the wrong sign causes an unjustified large loss. Therefore, the network prefers predictions with a buffer to the edges so that sign flips can be avoided. In order to correct for this effect, the loss function has to be modified appropriately. If the predicted $\hat{\Phi}$ lies in the expected region from $-\pi$ to π, 2π need to be subtracted from the deviation if it is larger than π or added if it is smaller than $-\pi$

$$
\begin{aligned}
&\text{if } \hat{\Phi} \in [-\pi, \pi] : \\
&\qquad \underbrace{\Phi - \hat{\Phi}}_{\Delta\Phi} = \Delta\Phi - 2\pi, \qquad \text{if } \Delta\Phi > \pi \\
&\qquad \underbrace{\Phi - \hat{\Phi}}_{\Delta\Phi} = \Delta\Phi + 2\pi, \qquad \text{if } \Delta\Phi < -\pi.
\end{aligned}
$$

The estimated azimuthal angle of the model profiting from the periodicity correction is depicted in figure 7.9(b). The frequency distribution becomes predominantly flat as expected. Only a small dip around zero is observable which also occurs in figure 7.9(a). It is less pronounced for the corresponding

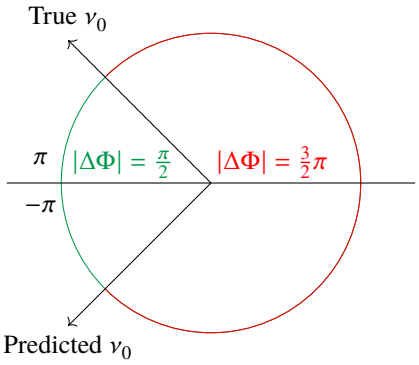
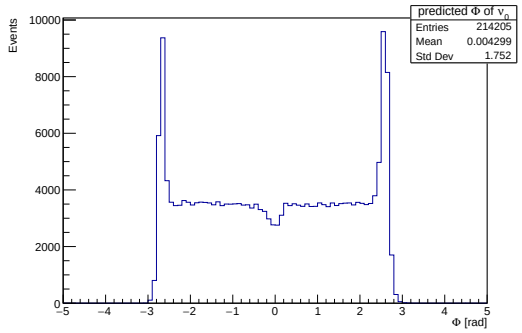
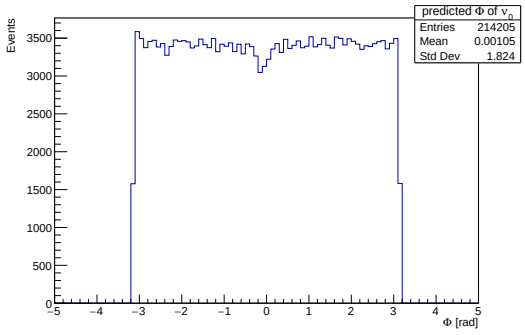


Figure 7.8: Illustration of the Φ-periodicity. The difference between the true and predicted azimuthal angle needs to be smaller than π. [12]



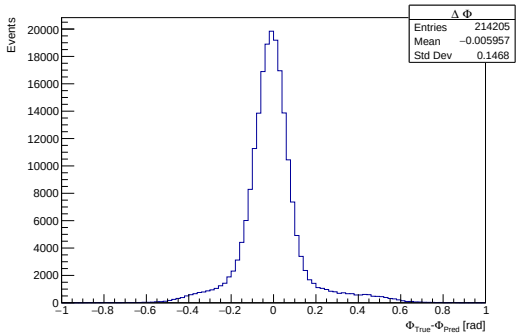
(a) Prediction based on a model trained with the common mean-squared-error.



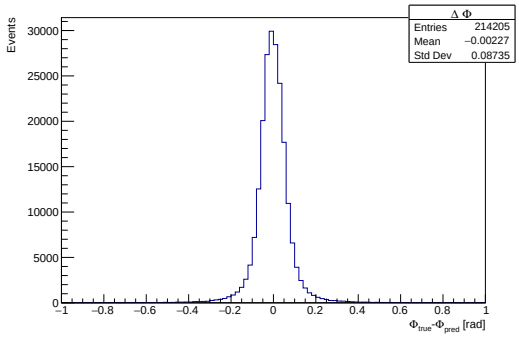
(b) Prediction based on a model considering a Φ-periodicity correction.

Figure 7.9: Influence of the periodicity correction on the azimuthal angle. The prediction is based on the evaluation with the events selected for testing and represents the azimuthal angle of the neutrino originating from the leading tau.

prediction of the subleading tau in figure C.6(b) and can be further reduced or disappears with the realization of physical constraints (see chapter 8). For predictions around zero, the effect cannot be attributed to the loss function. In order to give a reasonable explanation, though, further studies are necessary.
Additionally to the the impact of the correction on the frequency distributions, the resolutions are investigated in figure 7.10. While in figure 7.10(a) no periodicity correction was applied, it was considered in figure 7.10(b). The standard deviation of the latter is smaller which favors the application of the correction. Similar observations were made for the predictions corresponding to the neutrino of the subleading tau, depicted in figure C.6 and C.7 in appendix C. In the following, the Φ loss will always be adapted correspondingly unless indicated otherwise.



(a) Φ-resolution based on a model trained with the common mean-squared-error.



(b) Φ-resolution based on a model considering a periodicity correction.

Figure 7.10: Influence of the Φ-periodicity correction on the resolution of the azimuthal angle. The prediction was performed on the events selected for testing and corresponds to the neutrino originating from the leading tau.

7.5 Ambiguities: A Toy Model Study

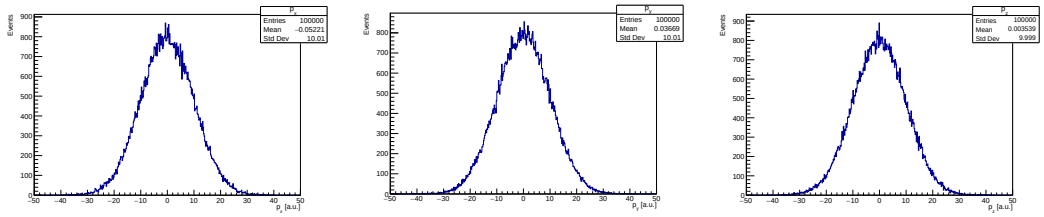
As described in section 4.5.3, ambiguities are caused by the mathematical solutions of the quadratic mass constraints. Since this issue appears as a hard coded aspect in the constraint layers, it can be dealt with. However, it is uncertain how the network deals with ambiguous mappings internally. Besides the constraints, quadratic relations also occur within the variables, for example in the energy, which comprises the sum of the squared momenta. No direct indications of an inappropriate processing of such aspects were observed for investigations concerning the di-tau mass reconstruction. A toy model study, though, revealed issues in this domain, which need to be further investigated.

The Setup

The idea behind the toy model study is the comprehension of the interplay of overlapping effects for multidimensional outputs and the influence of constraints on the learning process. It was investigated in what extent the incorporation of a physical constraint can replace the information content of a whole input variable or how different noise levels influence the overall performance. During these studies, an ambiguity related issue appeared which will be presented in the following.
The simplified toy model is designed for predictions related to a single particle. In this case, the physical constraint of interest is the energy-momentum relation:

$$E^2 = m^2 + p_x^2 + p_y^2 + p_z^2 \tag{7.2}$$

The generated data for the training is depicted in figure 7.11 and 7.12. For the momenta, 100000 independent samples were drawn from normal distributions with a mean of 0 a.u. and a standard deviation of 10 a.u.. Under the assumption of a mass of 1.8 a.u., the corresponding energies were calculated using equation 7.2. Furthermore 10000 events were produced for validation and testing, following the same scheme.
The architecture and choice of hyperparameters imitates the basic setup described in section 6.1 but task specific adaptations like three input nodes and one output node had to be considered.



(a) Generated p_x (b) Generated p_y (c) Generated p_z

Figure 7.11: Generated momenta. For each momentum, 100000 samples are drawn from a normal distribution with $\mu = 0$ a.u. and $\sigma = 10$ a.u..

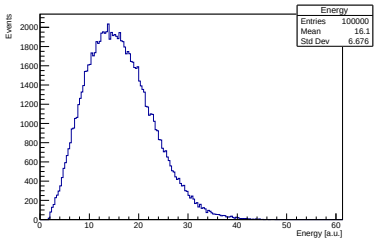


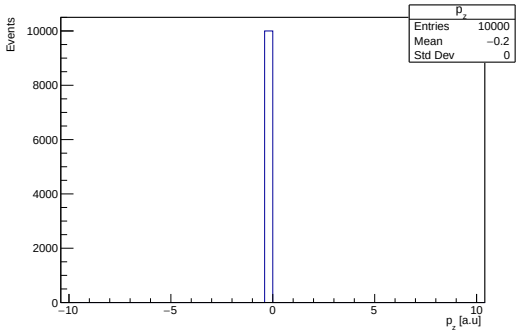
Figure 7.12: Energy determined with the toy momenta and an assumed mass of m=1.8 a.u. (see equation 7.2).

The Sign Study

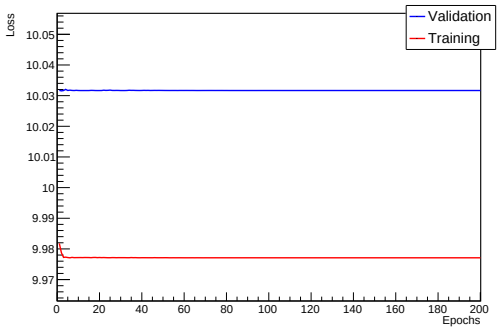
A sign related study was performed by providing the energy as well as the momentum in x- and y-direction as input and formulating the task so that p_z was used as target. In principal, most of the information is provided directly so that the mapping of the input variables to the output should be straightforward. However, no satisfactory learning was achieved as depicted in figure 7.13(a). Instead of reproducing the expected gaussian distribution, the momentum is stuck at zero, which is the mean of the expectation. The constant losses of the training as well as the one of the validation testify no learning process (see figure 7.13(b)). Their slight deviation might result from the evaluation after each batch or each epoch respectively (see section 6.2.2).

Since every information except the mass, which is only a constant, and the sign of p_z, which is quadratically included in the energy term, is given, ambiguity related issues are assumed.
Therefore, the sign of p_z, namely +1 for $p_z \geq 0$ and -1 for $p_z < 0$ are provided as additional input feature. The result of the prediction is depicted in figure 7.14. Like expected, a gaussian frequency distribution is produced (figure 7.14(a)) and the loss decreases with progressing epochs (figure 7.14(b)). Thus, the network is certainly able to extract the magnitude of p_z out of the input data, but it is overstrained by the ambiguous nature of the information content. Therefore, it is favorable for the optimization process to choose the impartial average between the positive and negative solution rather than preferring one side (figure 7.13(a)).

The described studies were also performed on the simulated data of the leading tau in order to confirm the result on data containing real physical correlations. The observations can be reproduced as depicted in figure C.8 in appendix C. Since many variables used for the di-tau mass reconstruction contain quadratic information, ambiguity related effects may occur.

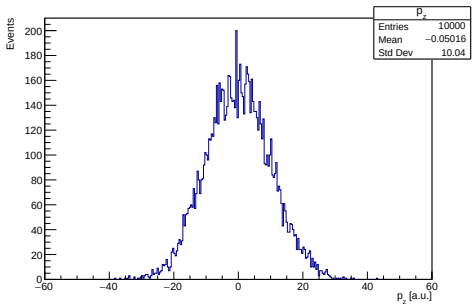


(a) Predicted p_z.

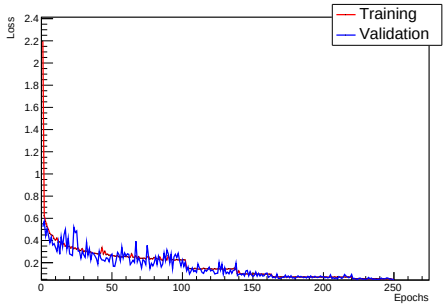


(b) Loss development.

Figure 7.13: Toy model study: p_x, p_y and E are used as input and p_z as target. The prediction was performed on the events generated for testing. Apart from the estimated momentum (a), the loss development (b) is depicted as well.



(a) Predicted p_z.



(b) Loss development.

Figure 7.14: Toy model study: p_x, p_y, E as well as the sign of p_z are used as input and p_z as target. The prediction was performed on the events generated for testing. Apart from the estimated momentum (a), the loss development is depicted as well (b).

CHAPTER 8

Di-Tau Mass Reconstruction Based on Artificial Neural Networks

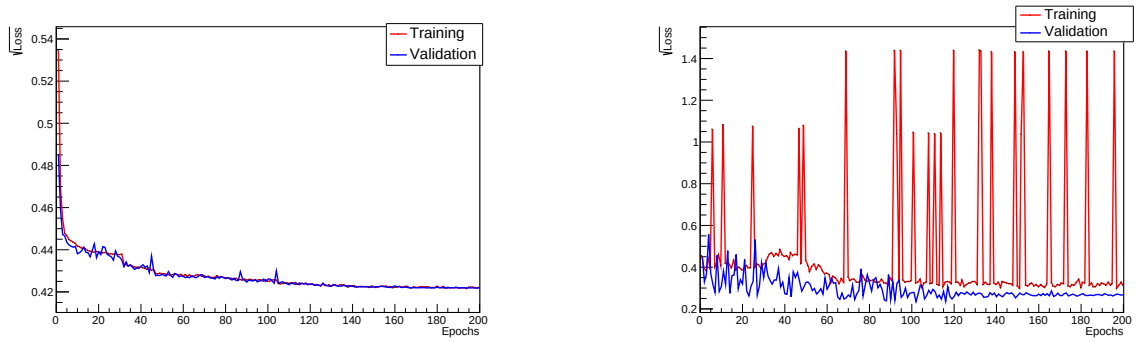
Considering the task and target related issues, described in chapter 7, regression based ANNs will be developed in order to perform the di-tau mass reconstruction. In this chapter, the main focus is on the differentiation of two specific approaches. One of them aims to provide the same starting point for the MMC and the ANNs. For this purpose, the input variables provided to the MMC (see section 5.3.1) have to be used for training and the invariant mass constraint of the tau needs to be realized as well. The latter is one of the fundamental assumptions of the current technique used for di-tau mass reconstruction and might provide valuable information for training.
Apart from these studies, models will be investigated based on input features beyond the MMC. The straightforward incorporation of additional or deviating input variables offers a beneficial flexibility of the ANN which favors its application compared to the MMC.

8.1 On par with the MMC

In this section, the two methods of constraint realization, introduced in section 4.5, will be explored and setups for the final comparison of different di-tau mass reconstruction techniques will be chosen. The investigation focuses on effects induced by the incorporation of the tau mass requirements into the learning process using the set of default hyperparameters introduced in section 6.1. The models were trained on the training sample γ^* and evaluated on the events selected for testing.

8.1.1 The Soft Constraint Approach

The soft constraint approach was introduced in section 4.5.2. It penalizes the network during the training process depending on the disregard of the mass constraint. The following investigation focuses on the performance in dependence on the weighting parameter α which determines the ratio of the common mean-squared-error and the penalty term in the loss.
The penalty term was applied in every one of the six output nodes, but its influence on the single loss terms differs because of their individual magnitudes before the loss scaling described in section 7.3. While the loss associated to the angles (≈ 1) is in a similar range as the penalty term, the loss of the transverse momentum is roughly 200 times larger. Applying the penalty first and scaling the p_T loss



(a) Loss corresponding to the transverse momentum after the scaling.

(b) Loss corresponding to the azimuthal angle.

Figure 8.1: Loss development for the output nodes corresponding to the transverse momentum (a) and the azimuthal angle (b) of the neutrino originating from the leading tau. A penalty weighting of 0.1 was applied.

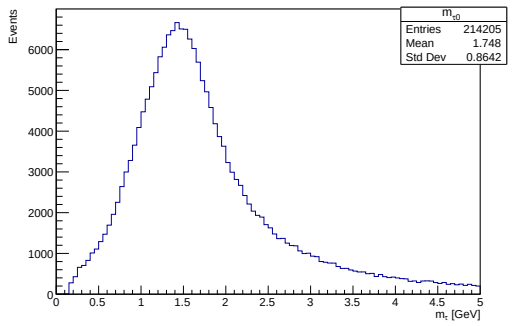
afterwards, leads to the smooth loss development of figure 8.1(a). In contrast, the loss of the azimuthal angle is characterized by distinctive peaks induced by the additional term in the loss function (see figure 8.1(b)). Applying the scaling of p_T before adding the penalty term reduced its performance immensely so that the first order was maintained. This choice supports the learning of the problematic target variable p_T (see section 7.2) and provides a sufficient constraint realization at the same time.

Since the total loss is the sum of the losses of the six output nodes, scaled to similar magnitudes, it is volatile due to the contributions of the angle related loss terms. The extent of the effect, however, can be controlled by a proper choice of the weighting parameter α.
The positive impact of the penalty term on the final constraint realization can be noticed in figure 8.2. In figure 8.2(a), the determined mass of the leading tau is shown for a model trained with the mean-squared-error, while the soft constraint was applied with a penalty weighting of 0.1 in the model producing the underlying prediction for the mass determination in figure 8.2(b). The calculation was performed according to equation 2.16 using the estimated neutrino variables and the reconstructed charged and neutral decay products of the tau.

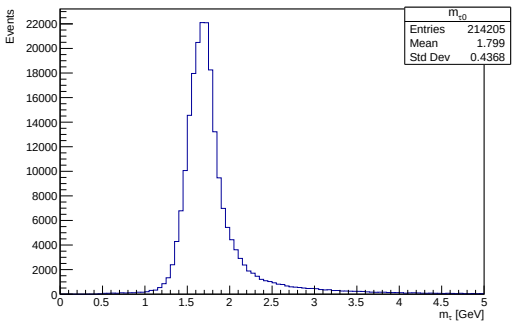
Both frequency distributions peak in the desired region around $m_\tau = 1.77\,\text{GeV}$ but the standard deviation of the latter is roughly half as large. A further increase of the penalty weighting causes merely small changes in the determined tau mass as depicted in figure D.1 in appendix D, but it influences the learning of the underlying variables, as discussed in the following.

In total, three different penalty weighting factors $\alpha = 0.1$, $\alpha = 0.2$ and $\alpha = 0.3$ were considered for training. The predictions of the ANNs using these weightings are described by the correlation factors listed in table 8.1. They comprise the six estimated neutrino variables as well as the determined di-tau mass. The determination is based on relation 3.13 which considers the predicted neutrino variables and the four momenta of the visible taus. The four momenta of the visible taus were chosen because of their superior energy resolution (see section 3.4.4). Using the reconstructed charged and neutral decay products instead would decrease the performance of the di-tau mass determination.

While the performance of the angle estimations is quite comparable for all setups, differences occur for the transverse momenta. The best result is achieved for the smallest weighting factor of $\alpha = 0.1$ with $\rho_{p_{T,0}} = 0.594$ and $\rho_{p_{T,1}} = 0.595$ which also induces the highest correlation of the di-tau mass,



(a) Determined tau mass for a model trained with the mean-squared-error.



(b) Determined tau mass for a soft constrained model using a penalty weighting of 0.1.

Figure 8.2: Impact of the soft constraint on the mass of the leading tau determined from the predicted neutrino variables using equation 2.16.

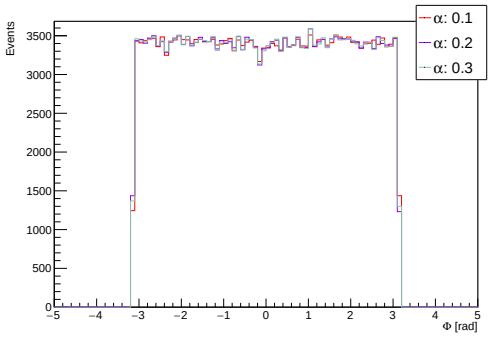
α	p_T^0	p_T^1	η^0	η^1	ϕ^0	ϕ^1	$m_{\tau\tau}$
0.1	0.594	0.595	0.997	0.998	0.950	0.958	0.804
0.2	0.591	0.579	0.997	0.998	0.950	0.958	0.798
0.3	0.571	0.573	0.997	0.998	0.950	0.958	0.797

Table 8.1: Correlation factors of the estimated neutrino variables as well as the one of the determined di-tau mass. The determination was performed using equation 3.13. Different penalty contributions in the loss function are considered. The model was evaluated using the events selected for testing.

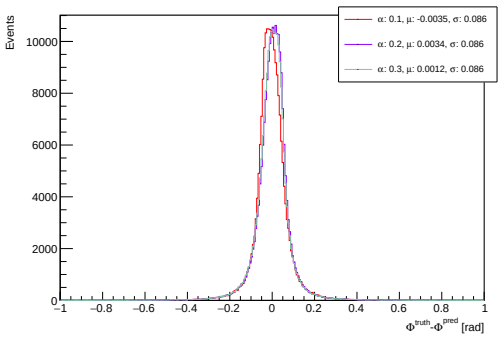
namely $\rho_{m_{\tau\tau}} = 0.804$. The frequency distributions of the predicted neutrino variables originating from the leading tau as well as their resolution are depicted in figure 8.3, 8.4 and 8.5 for all setups. The analogous figures related to the subleading tau are represented in D.2, D.3 and D.4 in appendix D.

While the estimation of the angles corresponds predominantly to the expectation, the predicted transverse momenta suffer from an inadequate reproduction of the tails and missing predictions for low momenta, as already observed in section 7.2.
Furthermore, biases occur in the resolution of the pseudorapidities, which vary with the applied penalty weighting and the belonging to the leading or subleading tau, as depicted in figure 8.4(b) and D.3(b). These biases might be induced by the additional loss term and especially affect the pseudorapidity since its prediction is the most accurate characterized by the smallest mean-squared-error contributions.
These observations emphasize the difficulties of a balanced, constrained learning. Even though the learning of the underlying variable and the constraint realization should accompany each other, compromises are made during the minimization of the loss function.

The determined di-tau masses of all investigated setups are shown in figure 8.6(a). Furthermore, the scatter plot based on a model using a penalty weighting of $\alpha = 0.1$ is depicted in figure 8.6(b). Scatter plots were introduced in section 6.2.1 and represent the pairs of true and determined di-tau masses. The remaining scatter plots for $\alpha = 0.2$ and $\alpha = 0.3$ are depicted in figure D.5(a) and D.5(b) in appendix D. Especially in the region up to 150 GeV, most of the masses are overestimated which is deduced from the accumulation of entries below the black line denoting the desired linear relation. The same conclusion is drawn from the binned resolution plots in figure D.6(a), D.7(a) and

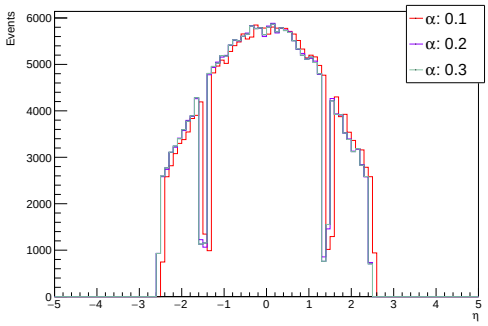


(a) Frequency distributions of the predicted azimuthal angle for different penalty weightings α.

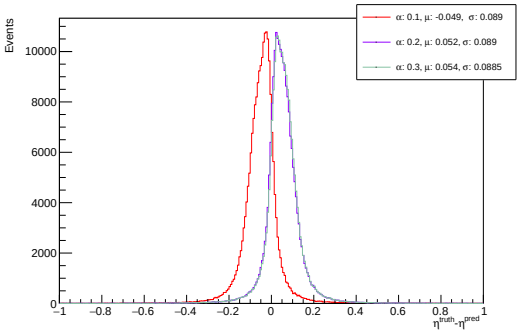


(b) Resolution of the azimuthal angle for different penalty weightings α.

Figure 8.3: Predicted azimuthal angle and its resolution. Different penalty weightings α are considered. The predictions are based on the events selected for testing and correspond to the neutrino of the leading tau.

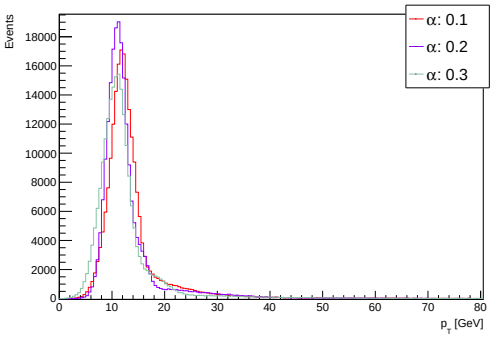


(a) Frequency distributions of the predicted pseudorapidity for different penalty weightings α.

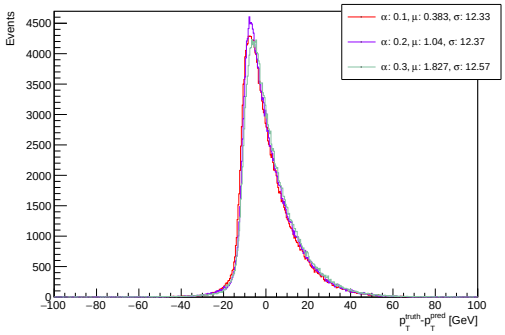


(b) Resolution of the pseudorapidity for different penalty weightings α.

Figure 8.4: Predicted pseudorapidity and its resolution. Different penalty weightings α are considered. The predictions are based on the events selected for testing and correspond to the neutrino of the leading tau.

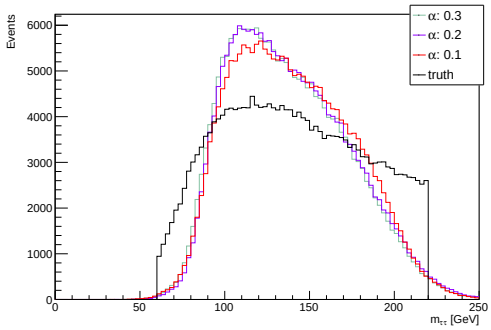


(a) Frequency distributions of the predicted transverse momentum for different penalty weightings α.

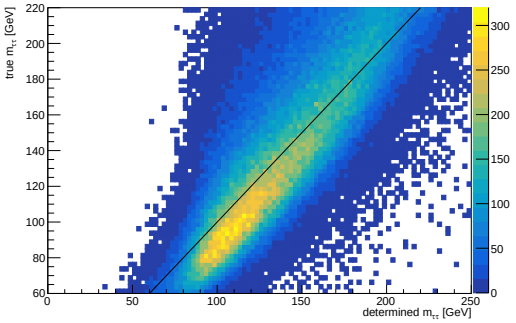


(b) Resolution of the transverse momentum for different penalty weightings α.

Figure 8.5: Predicted transverse momentum and its resolution. Different penalty weightings α are considered. The predictions are based on the events selected for testing and correspond to the neutrino of the leading tau.



(a) Determined di-tau masses for different penalty weightings α as well as the true di-tau mass.



(b) Scatter plot of the di-tau mass for a penalty weighting of $\alpha = 0.1$.

Figure 8.6: Determined di-tau masses for different penalty contributions in the loss function (a) and the scatter plot for a penalty weighting of $\alpha = 0.1$ (b). The correlation between the true and the determined di-tau mass corresponds to 80.4%. For illustration, the desired linear relation is denoted by the black line. All models were evaluated on the events selected for testing.

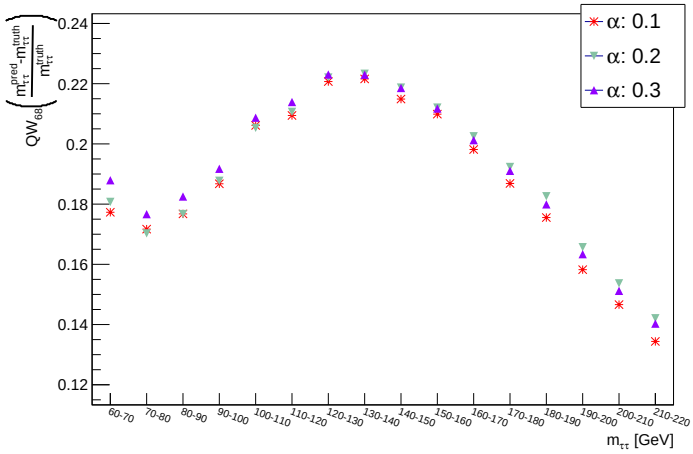


Figure 8.7: Half 68%-quantile widths for the calibrated binned resolution plots for different penalty contributions in the loss function. The applied calibration is described in section 6.2.4. The model was evaluated using the events selected for testing.

D.8(a) in appendix D which are characterized by positive biases for the mentioned mass region. The biases caused by the ANNs trained with different penalty weightings are nearly identical. Therefore, the comparison between their 68%-quantile widths would be valid. However, in order to enable comparability to other techniques, marked by different biases, the calibration described in section 6.2.4 was applied. The calibrated resolution plots are depicted in figure D.6(b), D.7(b) and D.8(b).

The corresponding half 68%-quantile widths are separately shown in figure 8.7 and reveal the best performance for the lowest penalty weighting. The correlation factor of the di-tau mass related to this model is also the highest ($\rho_{m_{\tau\tau}} = 80.4\%$). Therefore, it is chosen for following investigations.

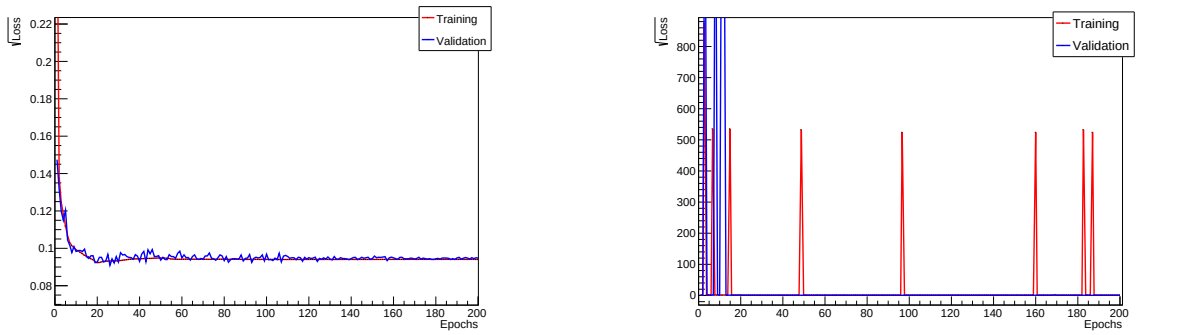
Setup	p_T^0	p_T^1	η^0	η^1	ϕ^0	ϕ^1	$m_{\tau\tau}$
one	0.336	0.375	0.997	0.998	0.950	0.958	0.761
one, smeared	0.314	0.352	0.997	0.998	0.950	0.958	0.743
two	0.370	0.382	0.997	0.998	0.950	0.958	0.762
two, smeared	0.366	0.341	0.997	0.998	0.950	0.958	0.737

Table 8.2: Correlation factors of the neutrino variables as well as the one of the corresponding di-tau mass. The description of the different setups are given in section 8.1.2. For the setups using a gaussian mass distribution, a standard deviation of 0.18 GeV was assumed. They are denoted by the term *smeared*. The models were evaluated on the events selected for testing.

8.1.2 The Hard Constraint Approach

The hard constraint approach was introduced in 4.5.3. It exploits the mass constraint by estimating two of three neutrino variables and determining the third accordingly. Since the estimation of the direction of flight performed quite well (see table 8.1), it might be suitable to use the angle related neutrino variables as direct outputs of the optimizable network part. Furthermore, no differentiation between positive and negative solutions has to be performed if the transverse momentum is determined in the constraint layers (see section 4.5.3).
Following, four different setups will be presented. They are motivated by different observations concerning the learning behavior. The initial setup represents the architecture described in section 4.5.3, which comprises four direct outputs for the angle estimation and two following constraint layers determining the transverse momenta. Despite a careful loss scaling, an unstable learning was observed. The training often developed in favor for quantities related to one tau, whereas the estimations for the second tau system diverged. This issue is ascribed to the hard coded calculation of p_T. As soon as the predicted angles induce a diverging loss in p_T, whether due to outlier predictions or due to the sensitive calculation, the prediction of the whole system of neutrino quantities involved in the calculation diverges and does not return to reasonable values. This issue is counteracted by controlling interventions within the constraint layer. One measure eliminates unphysical negative determinations of p_T by setting them to a fixed value, which was chosen to be zero. Consequently, the loss of the transverse momentum recovers so that the learning of the angle related variables returns to a suitable basis. Since such manipulative settings disturb the learning flow, an additional approach was considered. It uses two separated ANNs for the two tau decays respectively. Thus, the setup might be less dependent on these internal settings. The two ANNs have only two direct outputs for the angle related variables and one constraint layer for the calculation of the transverse momentum. The combined estimations of the neutrino variables can then be used for the di-tau mass determination. By means of this measure, an appropriate estimation for all variables might be established since the two systems do not influence each other anymore. Consequently, each ANN manages to deal with sudden jumps in the loss and to readjust the weights accordingly. The described effect is depicted in figure 8.8 which shows the loss development of the ANN dealing with the system of the subleading tau. The loss development for the predicted pseudorapidity is illustrated in figure 8.8(a) and for the determined transverse momentum in figure 8.8(b).

While the loss of the optimizable network part follows a smooth decreasing line as expected, the loss of the determined output is characterized by extensive peaks. However, already after a few epochs, the system recovers and returns to the baseline.

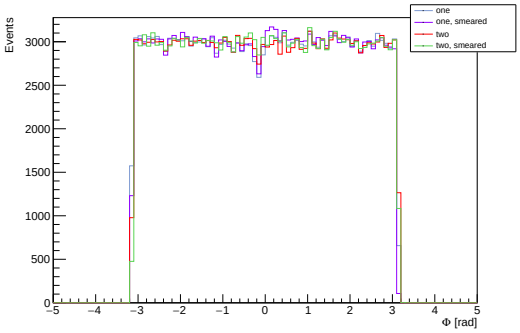


(a) Loss development of the pseudorapidity.

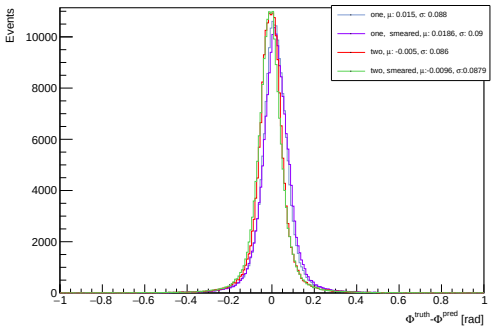
(b) Loss development of the transverse momentum.

Figure 8.8: Loss development of the pseudorapidity (a) and the one of the determined transverse momentum (b). The plots describe the learning of variables associated to the subleading tau within the setup using two ANNs (see section 8.1.2).

As a measure of performance, the correlation factors of the six neutrino variables and the one of the determined di-tau mass can be looked up in table 8.2. The setup descriptions *one* and *two* represent the corresponding number of ANNs used for the overall task. Although a high correlation can be achieved for the angle related variables, the correlation factors of the transverse momenta are rather low. This observation might be a direct consequence of the resolution of the inputs. Since several variables are plugged into the determination of p_T, a multitude of uncertainties contribute in its calculation as well. Consequently, the invariant mass constraint is fulfilled for the outputs but the transverse momentum might still be far off from the true value.
As possible solution, a gaussian distributed tau mass was used for the calculations in the constraint layers. It is centered at the original 1.77 GeV and a standard deviation of $\sigma = 0.18\,\text{GeV}$ was assumed after testing different values .
The correlation factors for the setups *one* and *two* using a gaussian tau mass distribution are additionally listed in table 8.2 and are denoted by the term *smeared*. While the estimation of the angle related variables stays almost unchanged, the performance of the transverse momenta decreases slightly. Thus, replacing the fixed mass by a random value of a normal distributed sample does not improve the final result. According to the correlation factors of the di-tau masses, the setups assuming a fixed tau mass perform best. With factors of 0.761 and 0.762 both are on par.
The frequency distributions as well as the resolutions of the predicted as well as the determined neutrino variables associated to the leading tau are depicted for all setups in figure 8.9, 8.10 and 8.11. Only events with a meaningful transverse momentum ($p_T > 0$) are considered, which induces an exclusion of roughly 25000 events out of the 214205 testing events. The predictions of the angle related variables are quite accurate which causes narrow and symmetric resolution plots. This observation does not apply for the transverse momentum whose frequency distribution is affected by the common deficits of a missing tail towards high momenta and a missing peak at zero. Depending on the variable and the setup, biases occur. In figure 8.9(b), representing the resolution of the azimuthal angle, a shift of the peak positions is noticeable for the setups relying on a single network. The observation might be an artefact of an unstable training due to diverging losses. Furthermore, a shift occurs for all setups in figure 8.10(b), depicting the resolution of the pseudorapidity. It might be caused by the hard coded

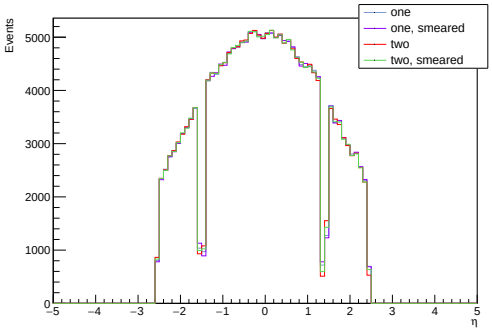


(a) Predicted azimuthal angle for different hard constrained setups.

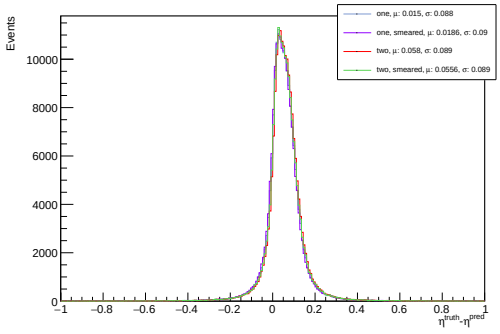


(b) Resolution of the azimuthal angle for different hard constrained setups.

Figure 8.9: Predicted azimuthal angle and its resolution. The results of the different hard constrained setups are depicted. Their description is given in section 8.1.2. The predictions are based on the events selected for testing and correspond to the neutrino associated to the leading tau. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.



(a) Predicted pseudorapidity for different hard constrained setups.



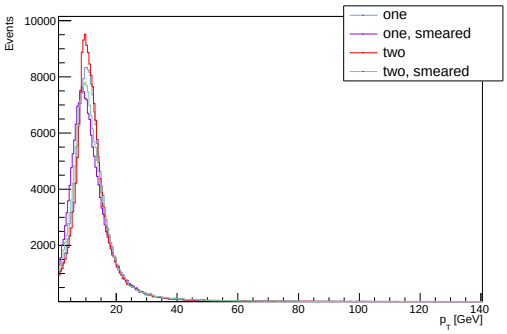
(b) Resolution of the pseudorapidity for different hard constrained setups.

Figure 8.10: Predicted pseudorapidity and its resolution. The results of the different hard constrained setups are depicted. Their description is given in section 8.1.2. The predictions are based on the events selected for testing and correspond to the neutrino associated to the leading tau. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.

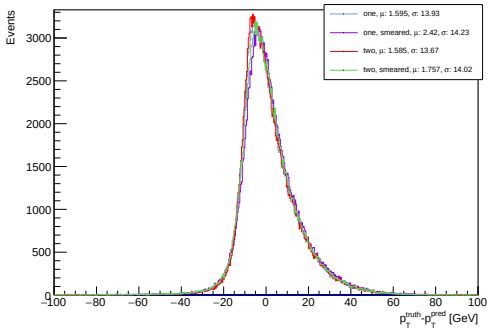
calculation of p_T and the interferences for non-physical determinations.

The analogous figures for the estimations related to the subleading tau are depicted in appendix D in figure D.9, D.10 and D.11. Applying equation 3.13 to the estimated neutrino variables and the reconstructed four momenta of the visible taus leads to the di-tau masses depicted in figure 8.12(a). The corresponding scatter plots are represented by figure 8.12(b), D.12 and D.13.

They reveal an overestimation for low masses which is also confirmed by the positive biases of the binned resolution plots in figure D.14(a), D.15(a), D.16(a) and D.17(a) in appendix D. Since biases have to be eliminated for the comparison of different techniques, the calibration described in section 6.2.4 is applied to all setups. The calibrated plots are depicted in figure D.14(b), D.15(b), D.16(b) and D.17(b). Their half 68%-quantile widths are compared in figure 8.13 in order to rate the

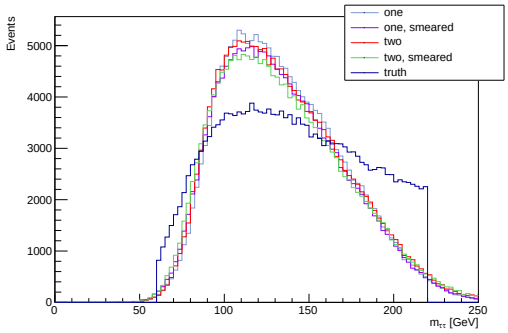


(a) Determined transverse momentum for different hard constrained setups.

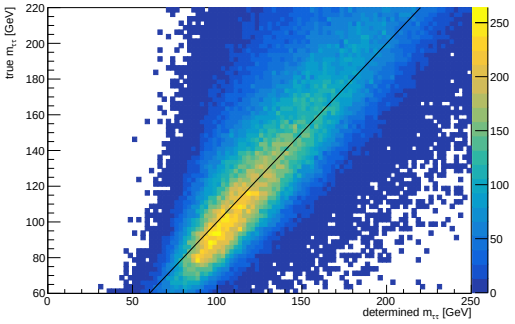


(b) Resolution of the determined transverse momentum for different hard constrained setups.

Figure 8.11: Determined transverse momentum and its resolution. The results of the different hard constrained setups are depicted. Their description is given in section 8.1.2. The predictions are based on the events selected for testing and correspond to the neutrino associated to the leading tau. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.



(a) Frequency distributions of the determined di-tau masses.



(b) Scatter plot based on the setup 'two'.

Figure 8.12: Determined di-tau masses for different setups relying on the hard constraint (a). The models were evaluated on the events selected for testing. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319. Furthermore, the scatter plot of the di-tau mass for the setup 'two' is shown (b), with the black line denoting the desired linear relation. The correlation factor: 76.2%.

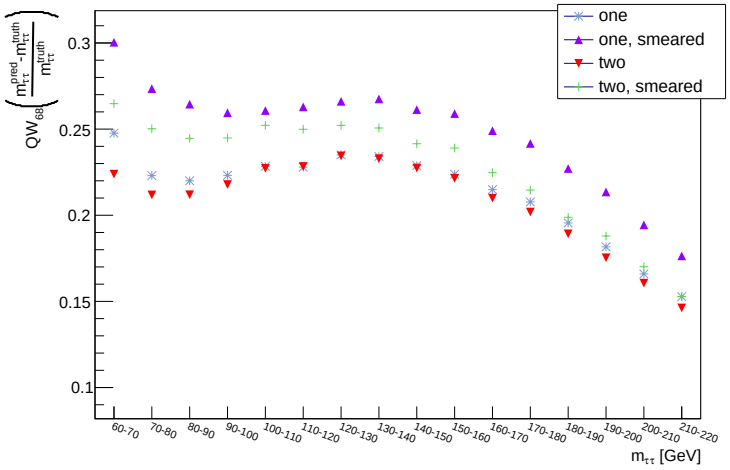


Figure 8.13: Half 68%-quantile widths of the calibrated binned resolution plot. The applied calibration is described in section 6.2.4. The different hard constrained setups were evaluated using the events selected for testing.

mass dependent reconstruction performance. It is noticed that the lowest widths are provided by the setups relying on a fixed mass. From these two setups, the model using two separated ANNs is chosen for further comparisons since its loss development testifies a more stable training.

8.1.3 Separation Power

The calibration introduced in section 6.2.4 enables a direct comparison of different di-tau mass reconstruction techniques. The half 68%-quantile widths of the binned resolution is a measure of the distribution's width. Therefore, it offers insight into the capability to separate different resonances.
The investigated models of this section provide the same starting point as the MMC which comprises the realization of the invariant mass constraint of the tau lepton. In section 8.1.1, a penalty weighting of $\alpha = 0.1$ was finally chosen while the best performing hard constrained model relies on two separated ANNs. Each of them learns the neutrino variables associated to one of the two taus.
The half 68%-quantile widths of these techniques are compared to the MMC in figure 8.14. Furthermore, a model trained with the MMC input variables, but without constraint realization is consulted in order to rate the influence of the invariant mass requirement on the learning achievement. This model is denoted by *no constraint*.
The resolution of the hard constrained model is characterized by the largest widths in most of the bins. A more desirable result is achieved, though, for the soft constrained ANN which performs even better than the MMC. However, the lowest quantile widths are provided by the unconstrained setup, estimating the neutrino variables only on basis of the given inputs. Thus, the incorporation of constraints seems to decrease the overall performance. This effect might result from an increased complexity of the optimization task, an imbalanced learning or unrealistic requirements as the hard coded constraint realization which cannot provide a proper handling of resolution effects.

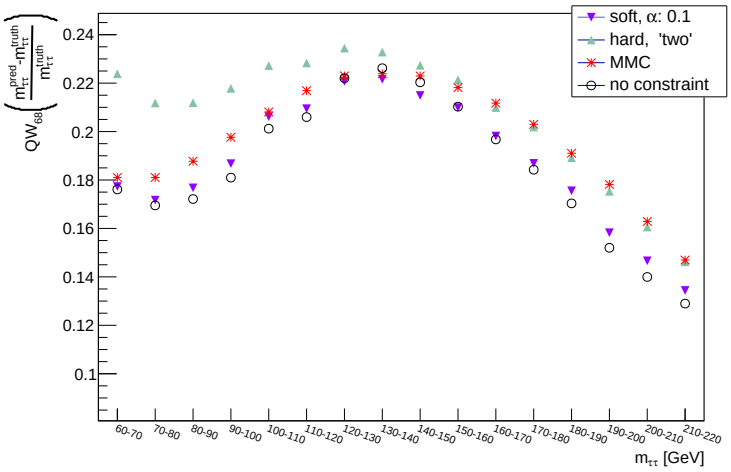


Figure 8.14: Half 68%-quantile widths for different di-tau mass reconstruction techniques in bins of the true mass. The binned resolution is based on the events selected for testing. The underlying reconstructed di-tau masses were calibrated beforehand.

8.2 Beyond the MMC

In this section, one of the main advantages of the ANNs will be exploited which is the easy exchange of input variables as well as the straightforward expansion by additional input nodes. Consequently, the amount as well as the kind of provided knowledge can be handled easily, which is an outstanding property compared to the MMC.

8.2.1 Investigated Input Features

Promising input variables comprise the kinematics of the substructure of the visible tau decay, namely the charged and neutral decay products as well as the jet kinematics. The former is highly correlated with the neutrino momenta which might improve their prediction. Furthermore, the jets of the event are of interest in two respects. Since the performance of the E_T^{Miss} reconstruction might be correlated with the jet kinematics, the ANN might learn the correction of deficits. Additionally, the jet properties might also improve the understanding of the boost of the underlying resonance.
In this section, two sets of input features will be explored. They are denoted by *Set 1* and *Set 2* which differ in the consideration of the jet kinematics.

Set 1

- the transverse momentum p_T
- the pseudorapidity η
- the azimuthal angle ϕ

} of the charged and the neutral decay products of the taus

- decay modes of the taus

Set	p_T^0	p_T^1	η^0	η^1	ϕ^0	ϕ^1	$m_{\tau\tau}$
1	0.554	0.600	0.997	0.998	0.949	0.955	0.808
2	0.560	0.604	0.997	0.998	0.948	0.955	0.807

Table 8.3: Correlation factors of the neutrino variables and the one of the determined di-tau mass for the two sets of inputs described in section 8.2.1. The results are based on the evaluation of the models using the events selected for testing.

- the reconstructed missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$
- the corresponding reconstructed azimuthal angle ϕ^{miss}

Set 2

- the transverse momentum p_T
- the pseudorapidity η
- the azimuthal angle ϕ

} of the charged and the neutral decay products of the taus

- the transverse momentum p_T
- the pseudorapidity η
- the azimuthal angle ϕ

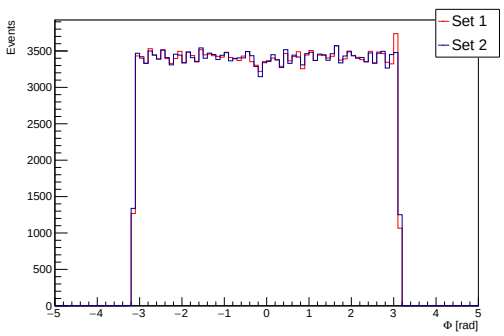
} of the reconstructed jets

- decay modes of the taus
- the reconstructed missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$
- the corresponding reconstructed azimuthal angle ϕ^{miss}

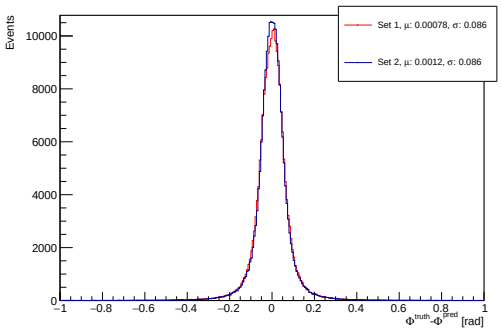
8.2.2 Performance

In order to investigate the influence of the alternative inputs listed in section 8.2.1, the basic setup was trained with the two introduced *Sets 1* and *2* and evaluated using the events selected for testing. The correlation factors of the network's outputs as well as the correlation of the determined di-tau mass are listed in table 8.3. As before, the determination based on equation 3.13. For the angle related variables, a high linear relation ($\geq 95\%$) occurs between the expected values and the predictions. Furthermore, the correlation factors of the transverse momenta related to the neutrino of the subleading tau ($\geq 60\%$) and the ones of the di-tau mass ($\geq 80.7\%$) are the highest compared to the results of the constrained networks (table 8.1 and 8.2).
The frequency distributions as well as the corresponding resolution of the predicted neutrino variables are depicted in figure 8.15, 8.16 and 8.17 for those originating from the leading tau. The figures related to the subleading tau are shown in D.18, D.19 and D.20 in appendix D. Compared to the constrained models investigated in section 8.1, a more consistent prediction without significant biases in the resolution is observed.

In figure 8.18, the frequency distributions of the determined di-tau masses (figure 8.18(a)) and the scatter plot of the di-tau mass based on the model trained with *Set 1* (figure 8.18(b)) are depicted.

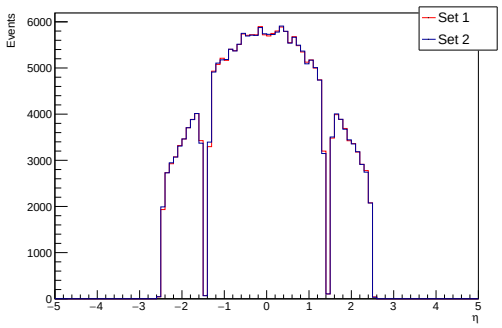


(a) Predicted azimuthal angle.

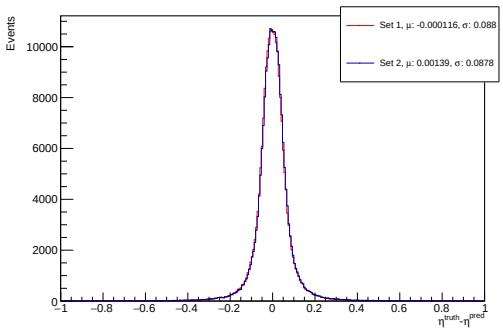


(b) Resolution of the azimuthal angle.

Figure 8.15: Predicted azimuthal angle and its resolution. The results for the models trained with the two sets of input variables beyond the MMC, described in section 8.2.1, are shown. The evaluation of the ANNs was performed on the events selected for testing and is depicted for the neutrino of the leading tau.

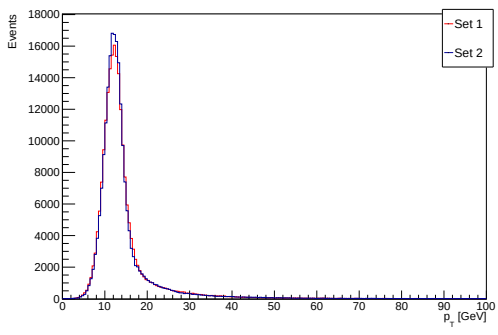


(a) Predicted pseudorapidity.

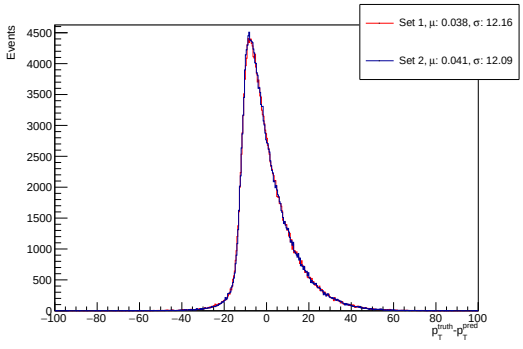


(b) Resolution of the pseudorapidity.

Figure 8.16: Predicted pseudorapidity and its resolution. The results for the models trained with the two sets of input variables beyond the MMC, described in section 8.2.1, are shown. The evaluation of the ANNs was performed on the events selected for testing and is depicted for the neutrino of the leading tau.

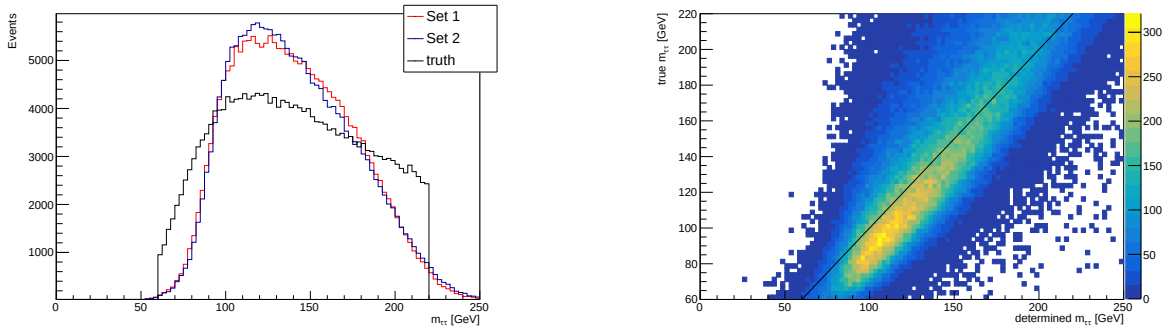


(a) Predicted transverse momentum.



(b) Resolution of the transverse momentum.

Figure 8.17: Predicted transverse momentum and its resolution. The results for the models trained with the two sets of input variables beyond the MMC, described in section 8.2.1, are shown. The evaluation of the ANNs was performed on the events selected for testing and and is depicted for the neutrino of the leading tau.



(a) Frequency distributions of the determined di-tau masses.

(b) Scatter plot of the di-tau mass based on the model trained with input *Set 1*.

Figure 8.18: Determined di-tau masses for the models trained with inputs beyond the MMC (a). The models were evaluated on the events selected for testing. Furthermore, the scatter plot of the di-tau mass based on the model trained with input *Set 1* is shown (b), with the black line denoting the desired linear relation. Correlation factor: 80.8%.

The determination was performed according to equation 3.13 as before. The scatter plot for the model trained with input *Set 2* is shown in figure D.21 in appendix D.
An overestimation of the di-tau masses up to 150 GeV is observed. This overestimation is similar to the one occurring in the models on par with the MMC which were investigated in section 8.1. The binned resolution plots before and after the calibration of the reconstructed di-tau masses are depicted in figure D.22 and D.23 respectively. The 68%-quantile widths of the calibrated resolution will be consulted in the following paragraph.

Comparison

In order to determine the impact of the inputs beyond the MMC on the final performance, a direct comparison of the half 68%-quantile widths after the calibration is provided in figure 8.19. The MMC is compared to the models trained with the input *Set 1* and *Set 2*, introduced in section 8.2.1 and to a model trained with the inputs on par with the MMC, denoted by *MMC input*. It is noticed that the ANNs perform better than the MMC in most of the bins. However, no overall superiority is ascribable to one specific setup since the ANN producing the lowest width varies with the bin. While the setup trained with the MMC inputs performs best for low and high masses, the ANNs based on the input *Set 1* and *Set 2* are characterized by the lowest widths in the range from 120 GeV–170 GeV. The comparison of *Set 1* and *Set 2* among each other also leads to bin dependent results. The model which additionally considers the jet information is characterized by lower widths in the region between 80 GeV–150 GeV but it is outperformed in the remaining bins by the network using input *Set 1*.

Outlook

Because of the good performance of the models beyond the MMC, the usage of further input variables might be suitable in order to improve overall performance. At this point, further investigations

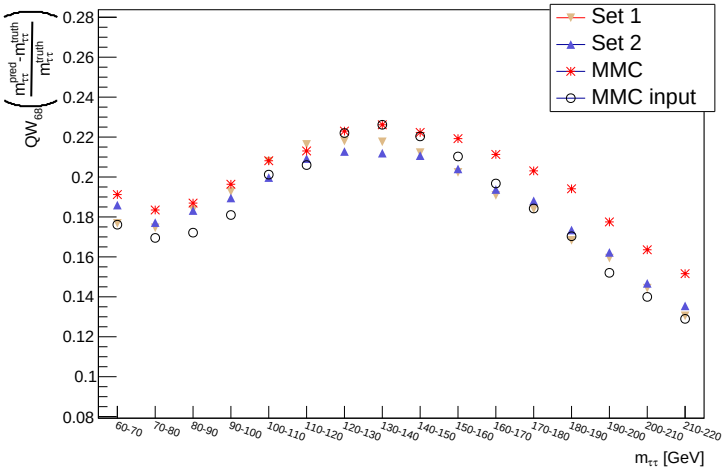


Figure 8.19: Comparison of the 68%-quantile widths after the calibration described in section 6.2.4. The MMC is compared to the performance of the models trained with the input *Set 1* and *Set 2*. Furthermore, the results of a model trained with the MMC inputs are consulted. The underlying predictions are based on the events selected for testing.

are necessary in order to fully exploit their potential. Promising extensions comprise the vertex information of the taus. The direction of flight of the tau, defined by the connection vector of the primary and secondary vertex describing the reconstructed position of production and decay, might be of interest.

CHAPTER 9

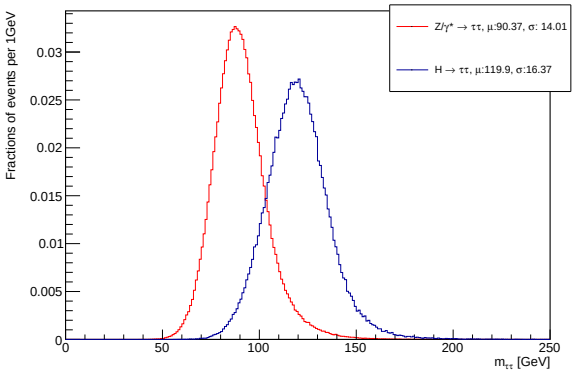
Evaluation

The overall task of the di-tau mass reconstruction is the separation of Higgs and Z. Therefore, the overlap of the reconstructed mass peaks is the closing and most important measure of performance and decides about the applicability of the reconstruction technique. In this context, the performance of the current technique for di-tau mass reconstruction, the MMC, will be compared to the overlap produced by the presented regression based artificial neural networks. In the following, the term *soft* describes the selected setup of section 8.1.1 using a penalty weighting of 0.1. The final choice of section 8.1.2, which consists of two separated ANNs with a fixed tau mass, will be denoted with the expression *hard*. For the sake of completeness, a model based on the MMC inputs, but without constraint realization is included in the comparison. It is denoted by the term *no constraint*. Additionally, the setups trained with input variables beyond the MMC are considered as well. The evaluation is based on the Higgs and the Z sample introduced in chapter 5. Related to the investigation of the overlap, the influence of the deviating spins of the underlying bosons (γ^*, Z, and Higgs) will be discussed.

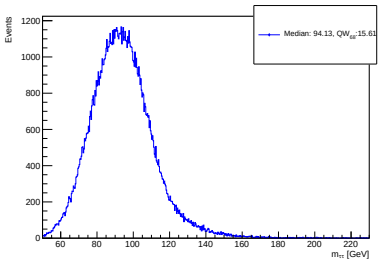
9.1 Performance of the MMC

The overlap of the reconstructed mass peaks measures the separation power of signal and background. In order to extract the percentage of events which fall in the same mass region, the frequency distributions of the reconstructed masses are normalized to one. For the overlap determination of the MMC, the di-tau mass reconstruction approach using the most probable neutrino three-momenta (mlnu3p) was chosen like before since it provides the most suitable comparison to the target choice of the ANNs. In figure 9.1(a), the reconstructed di-tau masses of the MMC are depicted for the Z sample in red, and for the Higgs sample in blue. Their overlap corresponds to 29.21%.

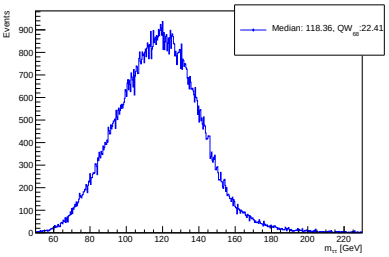
The means of the estimated mass peaks, 90.37 GeV for the Z and 119.9 GeV for the Higgs, are compatible with the expectations of 91.19 GeV and 125.1 GeV [13] respectively. Furthermore, consistency between the Z and the Higgs with the γ^* sample is probed by investigating the reconstructed masses of the training sample in 10 GeV bins of the underlying true masses, as depicted in figure 9.1(b) and 9.1(c). The chosen mass intervals imitate with 90 GeV − 100 GeV in figure 9.1(b) and 120 GeV − 130 GeV in figure 9.1(c) the resonances of interest. The medians of the reconstructed masses, 94.13 GeV and 118.36 GeV respectively, are comparable to the observed peak positions in the Z and the Higgs sample. Thus, the reconstruction technique of the MMC can be assumed to be largely independent on the individual physical properties of the samples.



(a) Normalized reconstructed di-tau masses of the MMC for the Z and the Higgs sample.



(b) Reconstructed di-tau masses of the MMC for true masses between 90 GeV and100 GeV. Median: 94.13 GeV.



(c) Reconstructed di-tau masses of the MMC for true masses between 120 GeV and 130 GeV. Median: 118.36 GeV.

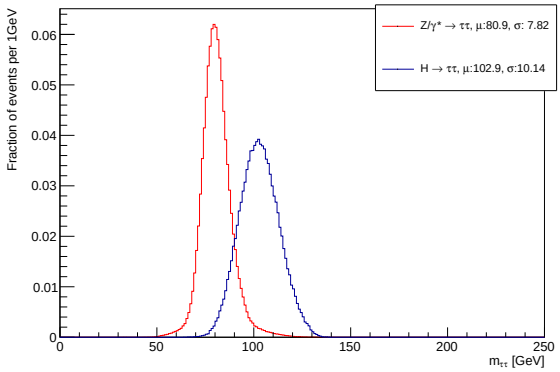
Figure 9.1: Normalized reconstructed di-tau masses for the Z and the Higgs sample using the MMC. The overlap corresponds to 29.21 % (a). Figure (b) and (c) show the reconstructed masses of the MMC in the γ^* sample in 10 GeV bins of the true mass.

9.2 Performance of Artificial Neural Networks

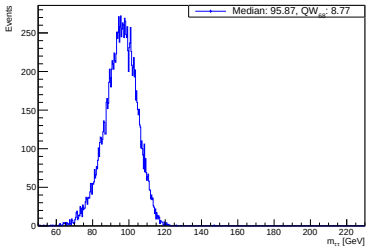
9.2.1 Influence of the Spin

In the following, the sensitivity of the ANN based reconstruction techniques to different physical properties in the samples will be probed. The understanding of these aspects is essential for the final evaluation described in section 9.2.2.
Possible dependencies on the sample were investigated using the basic model described in section 5. It was trained on truth level inputs, excluding any resolution effects. The chosen setup provides an idealized learning environment, focusing on the physical properties. The model evaluation on the Higgs and the Z sample results in an overlap of 20.3 %, as depicted in figure 9.2(a).

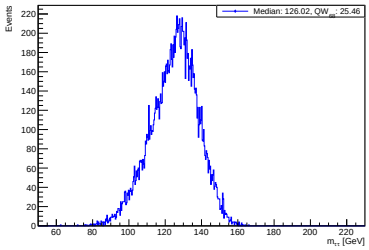
It is noticed that the means of the reconstructed mass peaks located at $\mu_Z = 80.9\,\text{GeV}$ and $\mu_H = 102.9\,\text{GeV}$ deviate from the expectation. If these observations are reproducible for the model evaluation on the γ^* sample, they are attributable to the learning achievement. In order to probe this hypothesis, di-tau masses were reconstructed on the basis of the training sample and were plotted in 10 GeV bins of the true masses, imitating the mass regions of the two investigated massive bosons. In figure 9.2(b), the true mass interval from 90 GeV–100 GeV was covered, whereas in figure 9.2(c) the interval from 120 GeV–130 GeV was chosen. The medians of the distributions are centered with 95.87 GeV and 126.02 GeV respectively, which corresponds to the desired peak positions of an



(a) Normalized reconstructed di-tau masses for the Z and the Higgs sample.



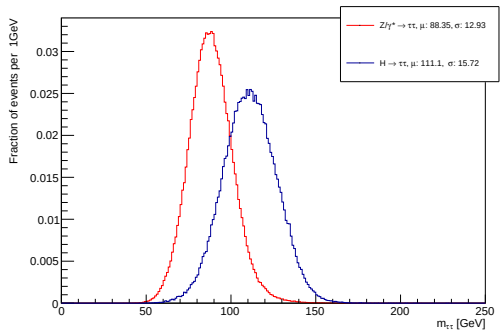
(b) Reconstructed di-tau masses for true masses between 90 GeV and 100 GeV. Median: 95.87 GeV.



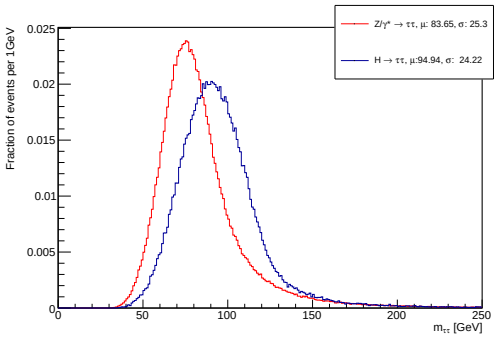
(c) Reconstructed di-tau masses for true masses between 120 GeV and 130 GeV. Median: 126.02 GeV.

Figure 9.2: Determined di-tau masses of the Z and the Higgs sample based on the basic model trained on truth level inputs. The corresponding overlap is 20.3 % (a). Figure (b) and (c) show the reconstructed masses resulting from the model evaluation on the γ^* sample. The depicted frequency distributions comprise 10 GeV bins of the underlying true masses. The events selected for testing were used.

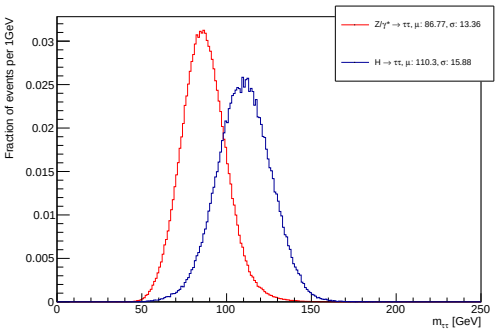
unbiased learning. Thus, the unexpected observations on the Higgs and the Z sample are caused by physical properties of the samples. This hypothesis is also supported by the smaller width of the Z peak (σ_z = 7.82 GeV) compared to the Higgs peak (σ_H = 10.14 GeV) which testifies more accurate predictions of the underlying neutrino variables in the Z sample. These effects might be induced by the different spins of the bosons. While the photon and the Z boson have a spin of 1, the scalar Higgs is spinless. Since the spin influences the decay kinematics of the tau as described in section 2.7.2, the Z sample is more similar to the training sample γ^*. Thus, a better performance of the Z prediction is comprehensible. Another aspect, which differentiates the Z sample from the other samples, is the polarization of its tau leptons. Since this property influences the decay kinematics as well, as described in section 2.7.1, a transferability of the learned properties of the γ^* sample to the Z sample, cannot be guaranteed. Therefore, the unexpected peak positions of the reconstructed Higgs and Z masses might result from deviating spins and polarization states compared to the training data of the ANN. In order to confirm this hypothesis, further investigations regarding these effects need to be pursued. An approach could consist of using events in the Z sample whose polarization configuration mimics the one of the γ^* sample. This selection should resolve the described issues for the chosen setup.
Furthermore, the observed spin sensitivity is attributable to the target choice of the ANN setup. Initial studies focusing on the di-tau mass as target [4] were not that affected by the described dependencies which might result from the influence of the spin on the decay kinematics of the tau. Whereas the neutrino momenta highly correlate with the overall decay kinematics, the di-tau mass of the whole



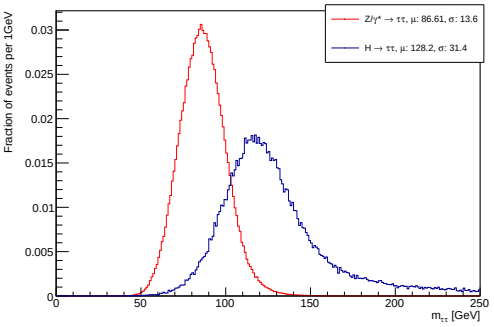
(a) Overlap for the di-tau mass reconstruction based on the soft constrained model: 41.23%.



(c) Overlap for the di-tau mass reconstruction based on the hard constrained model: 72.32%.



(b) Overlap for the di-tau mass reconstruction based on the model considering no constraints: 40.41%.



(d) Overlap for the di-tau mass reconstruction based on the model trained with inputs beyond the MMC (*Set 1*): 28.85%.

Figure 9.3: Normalized determined di-tau masses for the Z and the Higgs sample using different models. Their description is given in the introduction of chapter 9.

system is not that sensitive to this aspect.
Since ANNs seem to be capable to differentiate deviating spin configurations, this fact could be exploited in the discrimination of Higgs and Z in future studies. An interesting approach would be to estimate the di-tau mass using an ANN directly trained on the mass and combining the estimation with another model, which handles the spin information of the underlying resonance.

9.2.2 Final Evaluation

Having the discussed dependencies in mind, the developed ANNs can be evaluated on the Z and the Higgs sample. The normalized reconstructed di-tau mass peaks, belonging to the models listed in the introduction of this chapter, are depicted in figure 9.3. The evaluation of all remaining models, which were part of previous discussions in chapter 8, are depicted in figure E.1, E.2 and E.3 in appendix E.

Depending on the choice of the setup, different observations are made. In figure 9.3(a), the reconstructed di-tau masses of the Z and Higgs peak are depicted for the soft constrained model using a penalty weighting of 0.1. As before, the estimated Z peak, σ_Z =12.93 GeV, is more accurate than the Higgs peak, σ_H =15.72 GeV. Compared to the model using no constraint, as depicted in

figure 9.3(b), the mass peaks of the soft model are narrower but closer to each other. With overlaps of 41.23% and 40.41% respectively, both methods perform similarly. Since the setups based on larger penalty weightings, as depicted in figure E.1(a) and E.1(b), are characterized by even larger overlaps (44.04% for $\alpha = 0.2$ and 43.91% for $\alpha = 0.3$), no gain in the final performance was achieved by the requirement of the invariant tau mass.
In figure 9.3(c), the overlap produced by the hard constrained setup using two ANNs is depicted. Both mass peaks are characterized by large tails towards high masses. Furthermore, the Higgs distribution peaks with $\mu_H = 94.94\,\text{GeV}$ at unexpected low masses which almost induces a concurrence with the Z peak. Thus, the overlap corresponds to 73.0 % which excludes this approach for further studies. Comparable effects occur for the remaining hard constrained setups, introduced in section 8. Their corresponding overlaps are depicted in figure E.2(a), E.2(b) and E.3(a) in appendix E.
A final investigation considers the models trained with inputs beyond the MMC. In figure 9.3(d), the reconstructed di-tau masses are depicted for a training on the input *Set 1*. Whereas the estimation of the Z peak, with a mean of $\mu_Z = 86.61\,\text{GeV}$ and a width of $\sigma_Z = 13.6\,\text{GeV}$, performs reasonably, the estimated Higgs peak is characterized by a large width of $\sigma_H = 31.4\,\text{GeV}$. The produced overlap of 28.85% seems small but it mainly results from the large amount of outliers in the Higgs peak. This observation provides insight into the importance and the potential of the choice of the inputs. The training on the substructure of the tau decay seems to emphasize the spin dependent peculiarities of the different samples which is why the performance of the Higgs estimation decreases further.

CHAPTER 10

Conclusion

During the whole analysis, the di-tau mass reconstruction based on the prediction of the neutrino momenta. In this context, the focus lied on the realization of the invariant mass constraint of the tau. Enforcing the physical relation provided, together with a certain choice of input variables, the same starting point for the current method, the MMC, and the investigated ANNs. Compared to the MMC, neural networks are way more flexible in their expansion and in the consideration of further input variables. Therefore, the performance of models trained with the knowledge of the decay substructure of the tau were additionally investigated.
Training a model with multiple targets increases the complexity of the learning task and causes many overlapping effects, which are hard to disentangle. Different magnitudes of the output variables, boundary effects as well as ambiguities and periodicities of variables influenced the learning. However, the main limiting factor on the final performance can be assigned to the resolution of the input variables. Especially the learning of the transverse momentum, which is highly dependent on the missing transverse energy, could not be further improved.
All these aspects had to be considered in the constraint realization as well. In this thesis, two different approaches were further investigated. Apart from an additional term in the loss function (*soft constraint*), an enforced, hard coded realization of the invariant tau mass requirement (*hard constraint*) was implemented as well. Studies concerning the constraint realization revealed further learning related effects.
The hard coded realization of the invariant tau mass decreased the overall performance which might be attributed to an imbalanced training, caused by the fixed determination of one neutrino variable. The strict requirement of an ideal tau mass is an unrealistic assumption in a setup affected by resolution effects. The soft constraint, though, lead to more promising results. Apart from the constraint realization, also a reasonable learning of the network's outputs was established. From the investigated penalty weightings, however, the lowest performed the best. Therefore, the gain of the constraint realization seems to be outperformed by a proper learning of the underlying neutrino momenta. This hypothesis can also be confirmed in the comparison with the unconstrained setup. Considering the half 68%-quantile widths of the binned resolution plots based on the γ^* sample as well as the overlap resulting from the evaluation on the Z and the Higgs sample, its overall performance is on par or better than the soft constrained setup. The investigations revealed that the training of multiple output nodes at once is quite sensitive to a balanced learning. As soon as constraints are included, compromises between the realization of the physical requirement and the learning of the direct network outputs are

made which affects the final performance.
Apart from the ANNs mimicking the MMC, the choice of input variables was expanded for further investigations. Using the kinematics of the decay products as well as the kinematics of the jets as inputs for the training already led to promising results for the evaluation on the γ^* sample. The half 68%-quantile widths are even lower than the ones of the MMC. In this context, it might be advisable to expand the provided knowledge and include vertex information in the training as well.
However, the closing evaluation on the Z and the Higgs sample revealed further sample and setup related effects. First of all, the general performance, measured by the overlap of the two mass peaks, is outperformed by the MMC. The constrained models, for example, are characterized by overlaps of larger than 40%, which is more than 10% larger than the overlap produced by the current method of di-tau mass reconstruction. Furthermore, the training on the neutrino momenta is sensitive to spin dependent properties of the samples which is why a more accurate estimation of the Z peak can be observed. The performance differences for Z and Higgs are emphasized for the model trained on the decay substructure.
Generally, no suitable signal and background separation was achieved. However, the observed spin sensitivity might be useful for further studies regarding the Higgs and Z discrimination.
In conclusion, the training on multiple targets increases the complexity of the task. Furthermore, the resolution of the input variables influences the final result decisively. Although, the MMC was outperformed considering the 68%-quantile widths of the binned resolution based on the γ^* sample, the evaluation on the Z and Higgs sample excluded the investigated models as a replacement for the current method. Nevertheless, the application of ANNs is still a promising alternative. Initial studies on the topic [4] used the di-tau mass directly as target which eliminates or reduces some of the mentioned aspects. Therefore, it might be advisable for future studies to focus on a small complexity of the learning task and think about exploiting the advantages of the two different target sets.

Bibliography

[1] A. Witze, *Higgs found: Last particle in physics's standard model falls into place*, Science News, magazine of the Society for Science & the Public **Vol. 182 No. #2** (July 4, 2012) (cit. on p. 1).

[2] Particle Data Group, M. Tanabashi et al., *Status of Higgs Boson Physics*, Phys. Rev. D **98** (2019) 030001, URL: https://pdg.lbl.gov/2019/reviews/rpp2019-rev-higgs-boson.pdf (cit. on pp. 1, 10, 14).

[3] A. Elagin et al., *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nuclear Instruments and Methods in Physics **654** (2011) 481, arXiv: 1012.4686 (cit. on pp. 1, 24, 25).

[4] M. Werres, *Estimating the Mass of Di-Tau Systems in the ATLAS Experiment Using Neural Network Regression*, MA thesis: Universität Bonn, 2019, URL: https://web.physik.uni-bonn.de/group/view.php?&group=1&lang=de&c=t&id=102 (cit. on pp. 2, 36, 59, 87, 92).

[5] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013, ISBN: 9781107034266 (cit. on pp. 3, 4, 6–9, 13–15, 42).

[6] Andrej B. Arbuzov, *Quantum Field Theory and the Electroweak Standard Model*, hep-ph (2018) 35, arXiv: 1801.05670 (cit. on p. 5).

[7] S. Povh Rith and Zetsch, *Particles and nuclei*, Springer, 2003, ISBN: 9783540201687 (cit. on p. 5).

[8] Steven Weinberg, *Conceptual foundation of the unified theory of weak and electromagnetic interactions*, Nobel lecture (1979), URL: https://www.nobelprize.org/uploads/2018/06/weinberg-lecture.pdf (cit. on p. 6).

[9] The ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. B716 1-29 (2012), arXiv: 1207.7214 (cit. on p. 7).

[10] The CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 30 (2012), arXiv: 1207.7235 (cit. on p. 7).

[11] Peter W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964), URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.13.508 (visited on 17/10/2020); F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964), URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.13.321 (visited on 17/10/2020), cit. on p. 7.

[12] *Created with the TikZ package in LaTex* (cit. on pp. 8, 10, 17, 18, 31–33, 41, 48, 65).

[13] P. A. Z. e. a. Particle Data Group, *Review of Particle Physics*, J. Phys. G (2020), URL: http://pdg.lbl.gov (visited on 18/10/2020) (cit. on pp. 9, 10, 14, 24, 85).

[14] K.A. Olive et al., *Review of Particle Physics*, Chinese Physics C (2014), URL: https://iopscience.iop.org/article/10.1088/1674-1137/38/9/090001 (cit. on p. 10).

[15] Anderson et al., *Tau polarization asymmetry as a probe of Z' couplings*, Physical Review Letters **69** (1992) (cit. on p. 11).

[16] Monoranjan Guchait, et al., *Using Tau Polarization for Charged Higgs Boson and SUSY searchesat LHC*, hep-ph (2009), arXiv: 0808.0438v2 (cit. on p. 11).

[17] Paul T C, *A measurement of tau lepton polarization using the L3 detector at LEP*, PhD Thesis, 1994, URL: https://cds.cern.ch/record/348644 (cit. on p. 11).

[18] The ATLAS Collaboration, ed., *ATLAS High-Level Trigger,Data Acquisitionand Controls - Technical Design Report*, ATLAS TDR-016, 2003, URL: https://cds.cern.ch/record/616089/files/cer-002375189.pdf (cit. on p. 14).

[19] Lyndon Evans and Philip Bryant, *LHC Machine*, JINST 3 S08001 (2008), URL: https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001/meta (cit. on p. 15).

[20] official CERN webpage, *High-Luminosity LHC*, URL: https://home.cern/science/accelerators/high-luminosity-lhc (visited on 14/12/2020) (cit. on p. 15).

[21] webpage, *LHC*, URL: http://www.lhc-facts.ch/ (visited on 14/12/2020) (cit. on p. 15).

[22] C. webpage, *HL-LHC Industry*, URL: https://project-hl-lhc-industry.web.cern.ch/content/project-schedule (visited on 14/12/2020) (cit. on p. 15).

[23] officiel CERN webpage, *CERN's accelerator complex*, URL: https://home.cern/science/accelerators/accelerator-complex (visited on 19/09/2020) (cit. on pp. 15, 16).

[24] officiel CERN webpage, *LHC faq the guide*, URL: https://home.cern/sites/home.web.cern.ch/files/2018-07/CERN-Brochure-2017-002-Eng_0.pdf (visited on 19/09/2020) (cit. on p. 15).

[25] officiel CERN webpage, *Experiments*,
URL: https://home.cern/science/experiments (visited on 19/09/2020) (cit. on p. 15).

[26] CERN Document Server, *The CERN accelerator complex - 2019*,
URL: https://cds.cern.ch/record/2684277?ln=de (visited on 18/09/2020)
(cit. on p. 16).

[27] ATLAS Collaboration et al, *The ATLAS Experiment at the CERN Large Hadron Collider*,
JINST **3** (2008) S08003,
URL: https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003
(cit. on pp. 15–17, 19, 20).

[28] J. Pequenao, *Computer generated image of the whole ATLAS detector*,
URL: https://cds.cern.ch/images/CERN-GE-0803012-01 (visited on 19/09/2020)
(cit. on p. 17).

[29] A. Collaboration,
Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment,
URL: https://cds.cern.ch/images/ATLAS-PHOTO-2020-018-1 (visited on 22/09/2020)
(cit. on p. 19).

[30] The ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. C72 (2012), arXiv: 1110.3174
(cit. on p. 20).

[31] The ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, Eur. Phys. J. C 76 (2016), arXiv: 1606.01813
(cit. on p. 20).

[32] The ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s}$ = 13 *TeV*, Eur. Phys. J. C (2016), arXiv: 1603.05598
(cit. on p. 21).

[33] Cacciari et al., *The anti-k_t jet clustering algorithm*, JHEP (2008), arXiv: 0802.1189
(cit. on p. 21).

[34] The ATLAS collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s}$ = 13 *TeV*, ATL-PHYS-PUB-2015-015, 2016,
URL: http://cds.cern.ch/record/2037613 (cit. on p. 21).

[35] The ATLAS collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using p p collisions at* $\sqrt{s}$ = 13 *TeV*,
ATLAS-CONF-2017-029, 2017,
URL: https://cds.cern.ch/record/2261772/files/ATLAS-CONF-2017-029.pdf
(cit. on p. 21).

[36] Joern Mahlstedt, *The ATLAS Hadronic Tau Trigger*,
Journal of Physics Conference Series (2014), URL: https://www.researchgate.net/publication/273030152_The_ATLAS_Hadronic_Tau_Trigger (visited on 15/10/2020)
(cit. on p. 22).

[37] The ATLAS collaboration,
Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment,
Eur. Phys. J C 76(5) (2016), arXiv: 1512.05955 (cit. on pp. 21–23).

[38] The ATLAS collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s}$ = 13 *TeV*,
CONF-TAUP-2016-01, 2016, URL:
https://cds.cern.ch/record/2199788/files/ATL-COM-PHYS-2016-929.pdf?
(cit. on p. 22).

[39] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at* $\sqrt{s}$ = 13 *TeV*, Eur. Phys. J. C **78** (2018) 903,
arXiv: 1802.08168 (cit. on p. 23).

[40] ATLAS Collaboration, *Expected performance of missing transverse momentum reconstruction for the ATLAS detector at* $\sqrt{s}$ = 13 *TeV*, (2015),
URL: https://cds.cern.ch/record/2013489/files/ATL-COM-PHYS-2015-347.pdf
(visited on 28/09/2020) (cit. on p. 23).

[41] M. Hübner,
Effects of tau decay product reconstruction in a Higgs CP analysis with the ATLAS experiment,
MA thesis: Universität Bonn, 2016, URL: https://web.physik.uni-bonn.de/group/view.php?&group=1&lang=de&c=t&id=81 (cit. on p. 25).

[42] "exchange of information with M. Hübner; The topic will be covered soon in his PhD thesis" (cit. on pp. 25, 44).

[43] A. E. Jensen, *Dilepton final states with ATLAS at* $\sqrt{s}$ = 7 *TeV*,
MA thesis: University of Copenhagen, 2011,
URL: https://inspirehep.net/files/b802ad9a6004fe958ed9cdbfc30f0520
(cit. on p. 25).

[44] Habib et al., *ASCR/HEP Exascale Requirements Review Report*, physics.comp-ph (2016),
arXiv: 1603.09303v2 (cit. on p. 26).

[45] Gleisberg et al., *Event generation with SHERPA 1.1*, hep-ph (2008), arXiv: 0811.4622
(cit. on p. 26).

[46] Sjöstrand et al., *A Brief Introduction to PYTHIA 8.1*, hep-ph (2016), arXiv: 0710.3820
(cit. on p. 26).

[47] Kuhn et al., *APACIC++, A PArton Cascade In C++, version 1.0*, hep-ph (2000),
arXiv: hep-ph/0004270 (cit. on p. 26).

[48] Barberio et al., *Photos — a universal Monte Carlo for QED radiative corrections in decays*, Computer Physics Communications **66** (1991) 115 (cit. on p. 26).

[49] M. J. Schultens, *Search for Supersymmetry with Hadronically and Leptonically Decaying Tau Leptons at the ATLAS Experiment*, BONN-IR-2012-06, PhD Thesis: University of Bonn, 2016,
URL: https://www.lhc-ilc.physik.uni-bonn.de/ergebnisse/dateien/t00000078.pdf?c=t&id=78 (cit. on p. 26).

[50] B.Andersson et al., *Parton fragmentation and string dynamics*, Physics Reports **97** (1993) 31
(cit. on p. 26).

[51] Frixione, et al.,
Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,
hep-ph (2007), arXiv: 0709.2092 (cit. on pp. 26, 27).

[52] The ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, hep-ex (2010),
arXiv: 1005.4568 (cit. on p. 27).

[53] S.Agostinelli et al., *Geant4—a simulation toolkit*,
Nuclear Instruments and Methods in Physics Research **506** (2003) 250 (cit. on p. 27).

[54] Hinton et al., *Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups*, IEEE Signal Processing Magazine (2012),
URL: https://www.researchgate.net/publication/260637318_Deep_Neural_Networks_for_Acoustic_Modeling_in_Speech_Recognition_The_Shared_Views_of_Four_Research_Groups (visited on 20/10/2020) (cit. on p. 29).

[55] Wu et al., *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*, Computation and Language (2016), arXiv: 1609.08144v2
(cit. on p. 29).

[56] Gbenga Dada et al.,
Machine learning for email spam filtering: review, approaches and open research problems,
ScienceDirect **5** (2019) (cit. on p. 29).

[57] Silver et al., *Mastering the Game of Go without Human Knowledge*, Nature (2017),
URL: https://www.researchgate.net/publication/320473480_Mastering_the_game_of_Go_without_human_knowledge (visited on 20/10/2020) (cit. on p. 29).

[58] Ghoddusi et al., *Machine learning in energy economics and finance: A review*,
Elsevier **81** (2019) 709 (cit. on p. 29).

[59] Yun et al., *Detecting Cancer Metastases on Gigapixel Pathology Images*,
Computer Vision and Pattern Recognition (2017), arXiv: 1703.02442 (cit. on p. 29).

[60] Rasp et al., *Deep learning to represent subgrid processes in climate models*, PNAS (2018)
(cit. on p. 29).

[61] Wehle, *Machine Learning, Deep Learning, and AI: What's the Difference?*,
Conference: Data Scientist Innovation Day (2017) (cit. on p. 29).

[62] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*,
http://www.deeplearningbook.org, MIT Press, 2016 (cit. on pp. 29–31, 33–36).

[63] S. Raschka and V. Mirjalili, *Python Machine Learning*, Packt, 2017, ISBN: 9781787125933
(cit. on pp. 29–33, 36, 48).

[64] Tom M. Mitchell, *Machine Learning*, McGraw-Hill Science/Engineering/Math, 1997,
ISBN: 0070428077 (cit. on p. 29).

[65] Rumelhart et al., *Learning internal representations by error propagation*,
Institute for cognitive science (1986),
URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a164453.pdf (cit. on p. 32).

[66] R. Rojas, *Neural Networks - A Systematic Introduction*, Springer-Verlag, 1996, URL: http://www.inf.fu-berlin.de/inst/ag-ki/rojas_home/documents/1996/NeuralNetworks/neuron.pdf (visited on 14/12/2020) (cit. on p. 32).

[67] N. Zhang et al., *Investigation on Performance of Neural Networks Using Quadratic Relative Error Cost Function*, IEEE Access **7** (2019) 106642 (cit. on p. 34).

[68] B.T.Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Computational Mathematics and Mathematical Physics **4** (1964) 1 (cit. on p. 35).

[69] Duchi et al., *Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*, Journal of Machine Learning Research (2011) (cit. on p. 35).

[70] Hinton et al., *Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*, Coursera Class (2012), URL: http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf (cit. on p. 35).

[71] Diederik P. Kingma, Jimmy Ba, *Adam: A Method for Stochastic Optimization*, Machine Learning (cs.LG) (2014), arXiv: 1412.6980 (cit. on p. 35).

[72] Masters et al., *Revisiting Small Batch Training for Deep Neural Networks*, Computer Vision and Pattern Recognition (2018), arXiv: 1804.07612 (cit. on p. 36).

[73] Tom Beucler et al., *Achieving Conservation of Energy in Neural Network Emulators for Climate Modeling*, Atmospheric and Oceanic Physics (physics.ao-ph) (2019), arXiv: 1906.06622 (cit. on p. 38).

[74] Tom Beucler et. al., *Enforcing Analytic Constraints in Neural-Networks Emulating Physical Systems*, 2019, arXiv: 1909.00912 (cit. on p. 38).

[75] The ATLAS Collaboration, ed., *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, 2019, URL: http://cdsweb.cern.ch/record/2688062/files/ATL-PHYS-PUB-2019-033.pdf (cit. on pp. 39, 43).

[76] ATLAS Collaboration, *MC15.425200.Pythia8EvtGen_A14NNPDF23LO_Gammatautau_MassWeight.py*, 2016, URL: https://gitlab.cern.ch/atlas-physics/pmg/infrastructure/mc15joboptions/-/blob/master/share/DSID425xxx/MC15.425200.Pythia8EvtGen_A14NNPDF23LO_Gammatautau_MassWeight.py (visited on 18/10/2020) (cit. on p. 42).

[77] The ATLAS Collaboration, ed., *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 42).

[78] Ball et al., *Parton distributions with LHC data*, hep-ph (2012), arXiv: 1207.1303 (cit. on p. 42).

[79] David J.Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research **462** (2001) 152 (cit. on p. 42).

[80] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, 2019, arXiv: 1811.08856v2 (cit. on p. 43).

[81] M. A. et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015, URL: https://www.tensorflow.org/ (cit. on p. 49).

[82] F. Chollet and the Keras Community, (), URL: https://keras.io (cit. on p. 49).

[83] F. James, *Statistical Methods in Experimental Physics*, Worlds Scientific, 2012, ISBN: 9789812705273 (cit. on pp. 50, 52).

APPENDIX A

Monte Carlo Samples

In the following, the full descriptions of the MC Samples used in this thesis are listed. While the $\gamma^* \rightarrow \tau\tau$ sample is based on Pythia, $Z\backslash\gamma^* \rightarrow \tau\tau$ relies on Sherpa and $H\rightarrow \tau\tau$ was generated by the Powheg generator. A more detailed description of the tools and the sample properties is given in section 3.6 and chapter 5.

A.1 Pythia Sample

- mc16_13TeV.425200.Pythia8EvtGen_A14NNPDF23LO_Gammatautau_MassWeight.deriv.DAOD_HIGG4D3.e5468_s3126_r10201_r10210_p3703

A.2 Sherpa Samples

- mc16_13TeV.308094.Sherpa_221_NNPDF30NNLO_Ztautau2jets_Min_N_TChannel.deriv.DAOD_HIGG4D3.e5767_s3126_r10201_p3749
- mc16_13TeV.308094.Sherpa_221_NNPDF30NNLO_Ztautau2jets_Min_N_TChannel.deriv.DAOD_HIGG4D3.e5767_s3126_r10724_p3749
- mc16_13TeV.308094.Sherpa_221_NNPDF30NNLO_Ztautau2jets_Min_N_TChannel.deriv.DAOD_HIGG4D3.e5767_s3126_r9364_p3749
- mc16_13TeV.344775.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r10201_p3749
- mc16_13TeV.344775.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r10724_p3749
- mc16_13TeV.344775.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r9364_p3749
- mc16_13TeV.344779.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r10201_p3749

- mc16_13TeV.344779.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r10724_p3749
- mc16_13TeV.344779.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r9364_p3749
- mc16_13TeV.344782.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r10201_p3749
- mc16_13TeV.344782.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r10724_p3749
- mc16_13TeV.344782.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_h30h20.deriv.DAOD_HIGG4D3.e5585_s3126_r9364_p3749

A.3 Powheg Samples

- mc16_13TeV.345123.PowhegPy8EG_NNLOPS_nnlo_30_ggH125_tautauh30h20.deriv.DAOD_HIGG4D3.e5814_s3126_r10201_p3749
- mc16_13TeV.345123.PowhegPy8EG_NNLOPS_nnlo_30_ggH125_tautauh30h20.deriv.DAOD_HIGG4D3.e5814_s3126_r10724_p3759
- mc16_13TeV.345123.PowhegPy8EG_NNLOPS_nnlo_30_ggH125_tautauh30h20.deriv.DAOD_HIGG4D3.e5814_s3126_r9364_p3749
- mc16_13TeV.346193.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_tautauh30h20.deriv.DAOD_HIGG4D3.e7259_s3126_r10201_p3759
- mc16_13TeV.346193.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_tautauh30h20.deriv.DAOD_HIGG4D3.e7259_s3126_r10724_p3759
- mc16_13TeV.346193.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_tautauh30h20.deriv.DAOD_HIGG4D3.e7259_s3126_r9364_p3759

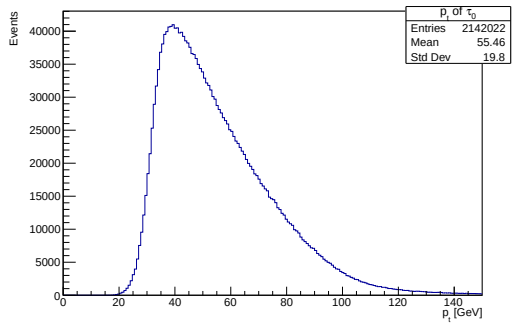
APPENDIX B

Further Plots of chapter 5

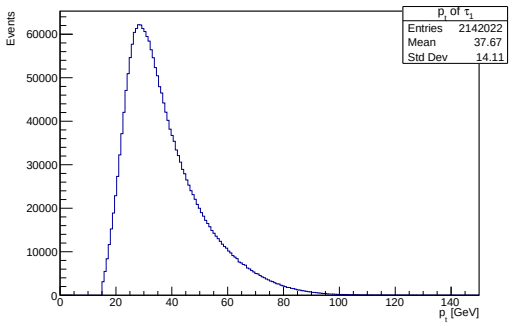
B.1 Input Data

The following figures B.1–B.11 represent the frequency distributions of the input variables used for the ANN training. All events passing the selection described in section 5.1.1 are considered.

B.1.1 On par with the MMC

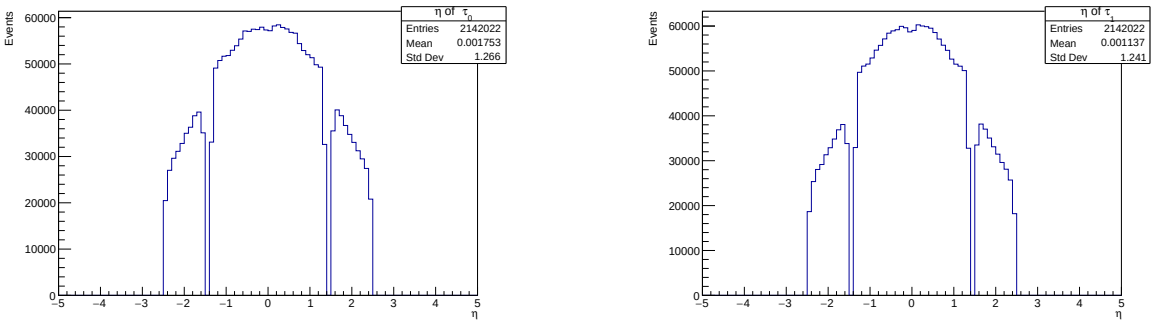


(a) Transverse momentum of the leading tau.



(b) Transverse momentum of the subleading tau.

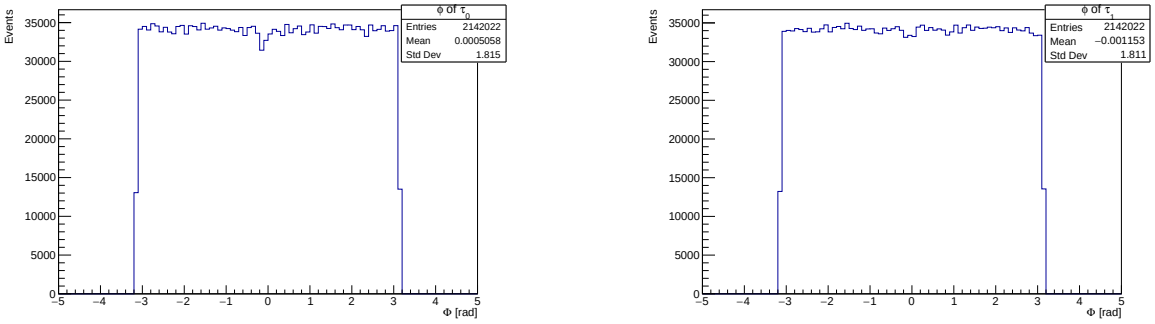
Figure B.1: Reconstructed transverse momentum of the taus. All events passing the selection described in section 5.1.1 are depicted.



(a) Pseudorapidity of the leading tau.

(b) Pseudorapidity of the subleading tau.

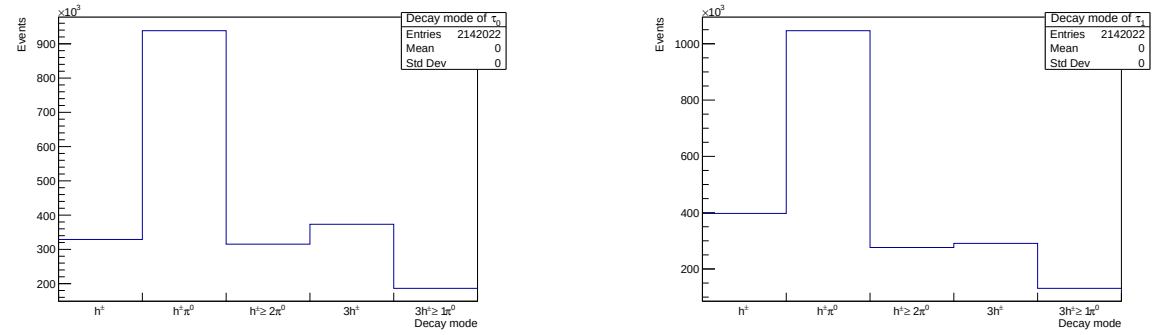
Figure B.2: Reconstructed pseudorapidity of the taus. All events passing the selection described in section 5.1.1 are depicted.



(a) Azimuthal angle of the leading tau.

(b) Azimuthal angle of the subleading tau.

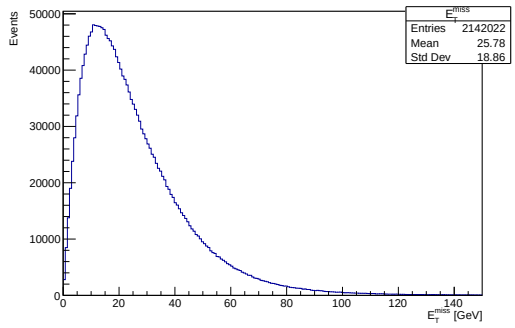
Figure B.3: Reconstructed azimuthal angle of the taus. All events passing the selection described in section 5.1.1 are depicted.



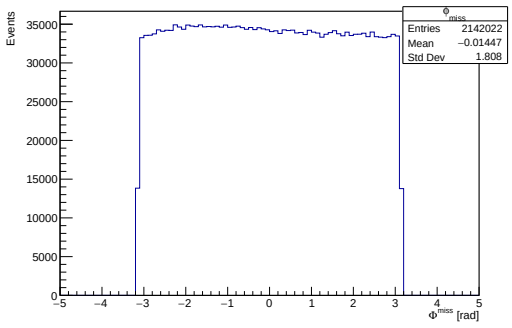
(a) Decay modes of the leading tau.

(b) Decay modes of the subleading tau.

Figure B.4: Reconstructed decay mode of the taus. All events passing the selection described in section 5.1.1 are depicted. The decay modes of the tau are introduced in section 2.7.



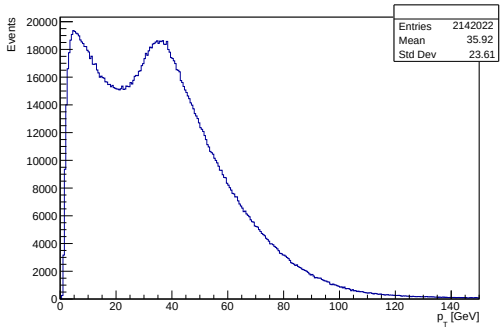
(a) Missing transverse energy.



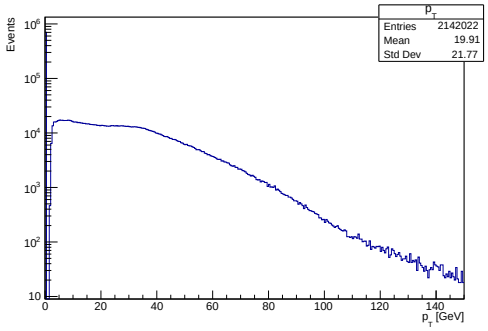
(b) Azimuthal angle of the missing transverse energy.

Figure B.5: Reconstructed missing transverse energy and the corresponding azimuthal angle. All events passing the selection described in section 5.1.1 are depicted.

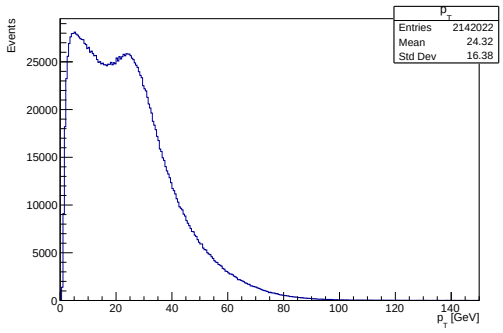
B.1.2 Beyond the MMC



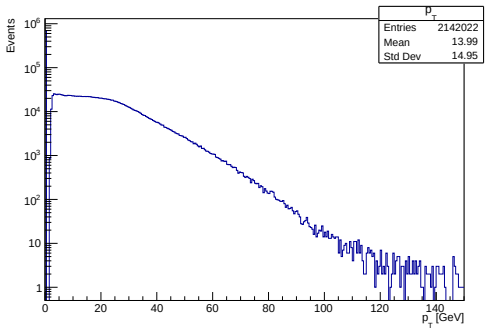
(a) Transverse momentum of the charged decay products of the leading tau.



(c) Transverse momentum of the neutral decay products of the leading tau.

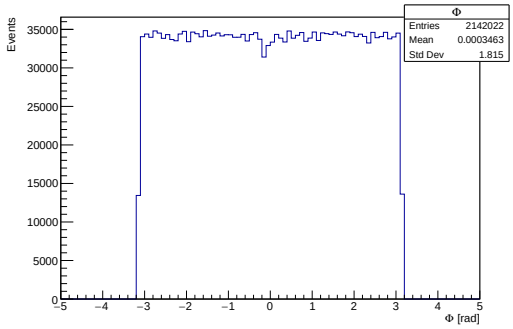


(b) Transverse momentum of the charged decay products of the subleading tau.

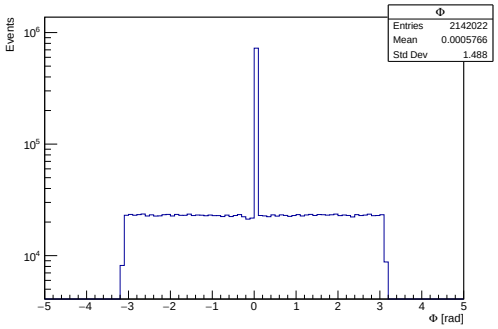


(d) Transverse momentum of the neutral decay products of the subleading tau.

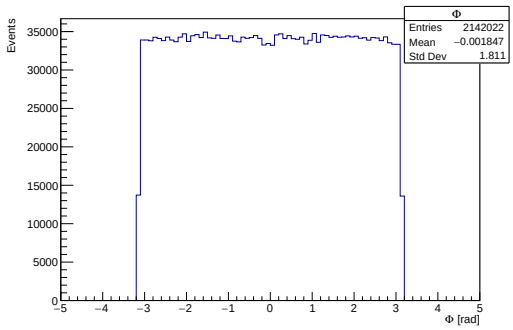
Figure B.6: Transverse momentum of the charged and neutral decay products. All events passing the selection described in section 5.1.1 are depicted. If no object was reconstructed, a zero is considered.



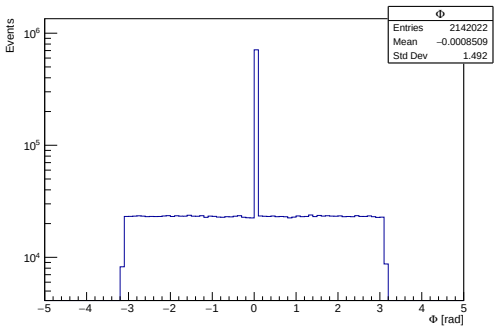
(a) Azimuthal angle of the charged decay products of the leading tau.



(c) Azimuthal angle of the neutral decay products of the leading tau.

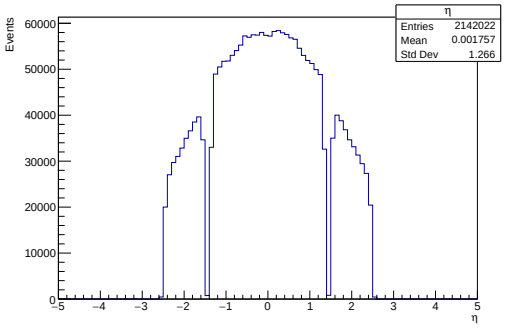


(b) Azimuthal angle of the charged decay products of the subleading tau.

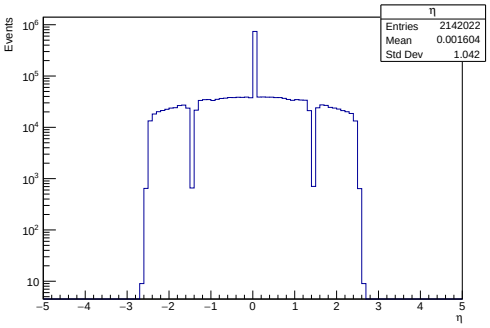


(d) Azimuthal angle of the neutral decay products of the subleading tau.

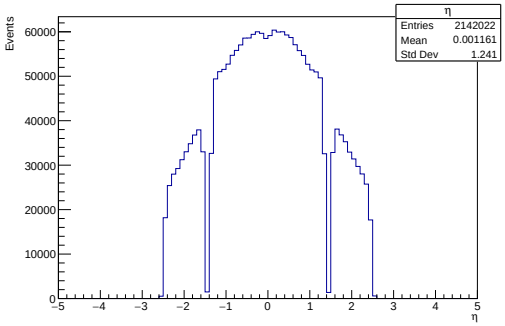
Figure B.7: Azimuthal angle of the charged and neutral decay products. All events passing the selection described in section 5.1.1 are depicted. If no object was reconstructed, a zero is considered.



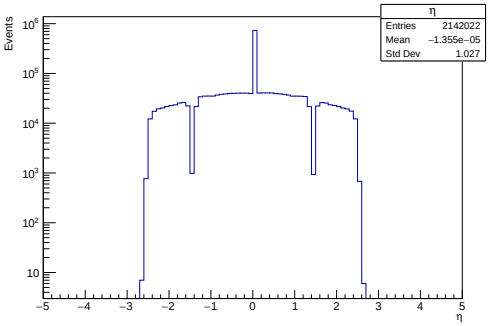
(a) Pseudorapidity of the charged decay products of the leading tau.



(c) Pseudorapidity of the neutral decay products of the leading tau.

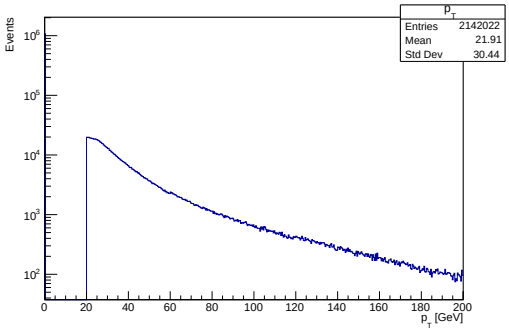


(b) Pseudorapidity of the charged decay products of the subleading tau.

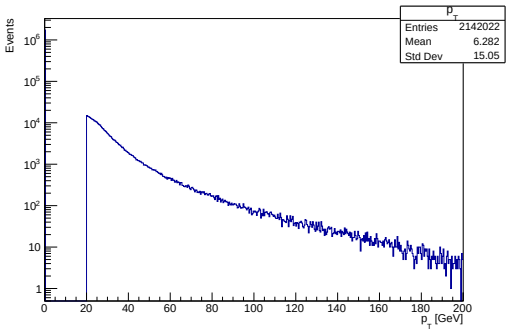


(d) Pseudorapidity of the neutral decay products of the subleading tau.

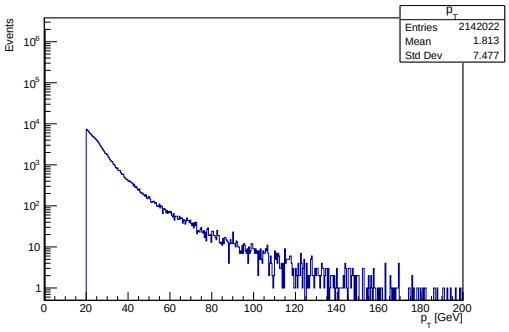
Figure B.8: Pseudorapidity of the charged and neutral decay products. All events passing the selection described in section 5.1.1 are depicted. If no object was reconstructed, a zero is considered.



(a) Transverse momentum of jet 0.

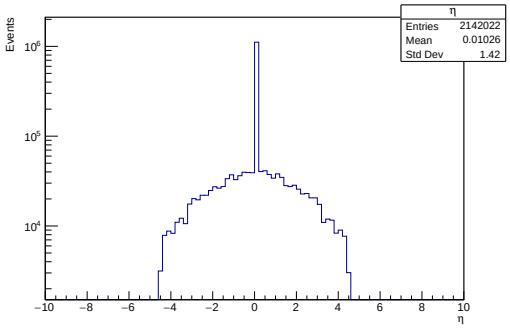


(b) Transverse momentum of jet 1.

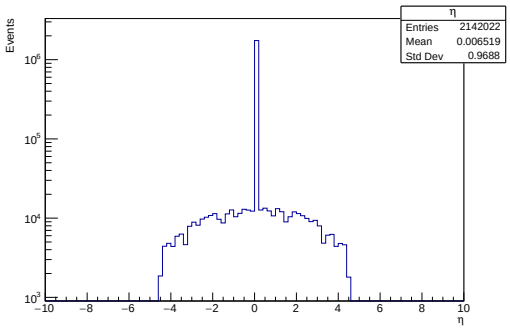


(c) Transverse momentum of jet 2.

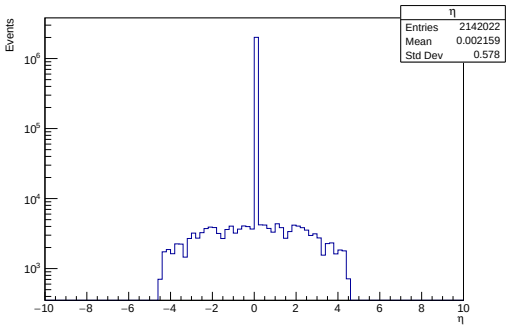
Figure B.9: Transverse momentum of the reconstructed jets. All events passing the selection described in section 5.1.1 are depicted. If no jet was reconstructed in the event, a zero is considered.



(a) Pseudorapidity of jet 0.

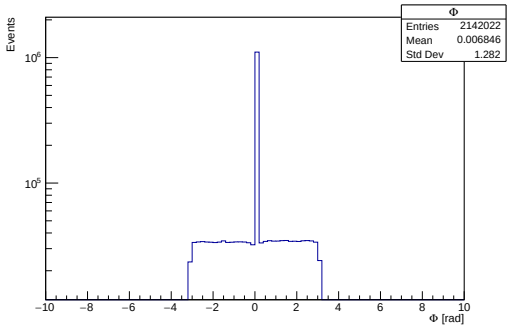


(b) Pseudorapidity of jet 1.

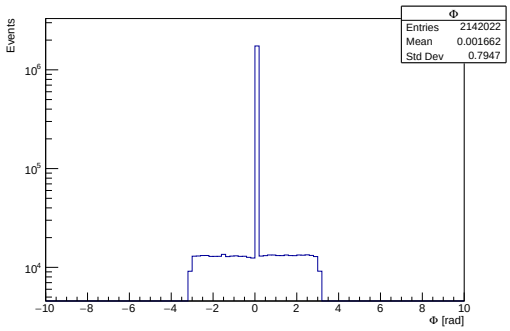


(c) Pseudorapidity of jet 2.

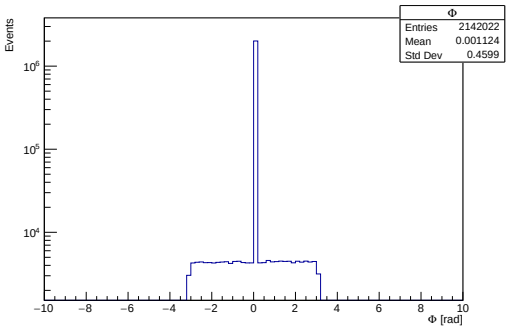
Figure B.10: Pseudorapidity of the reconstructed jets. All events passing the selection described in section 5.1.1 are depicted. If no jet was reconstructed in the event, a zero is considered.



(a) Azimuthal angle of jet 0.



(b) Azimuthal angle of jet 1.

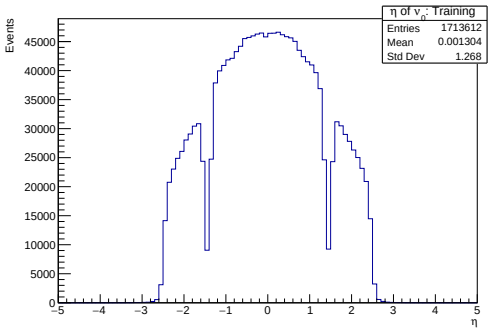


(c) Azimuthal angle of jet 2.

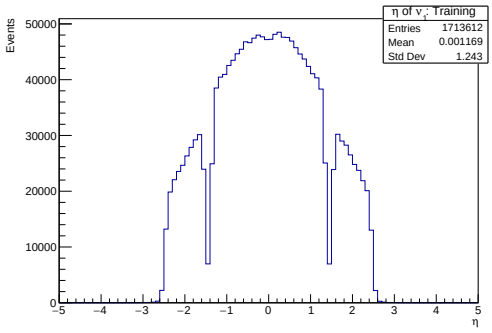
Figure B.11: Azimuthal angle of the reconstructed jets. All events passing the selection described in section 5.1.1 are depicted. If no jet was reconstructed in the event, a zero is considered.

B.2 Target Distributions

The following figures B.12–B.20 show the target variables for the events selected for training, validation and testing.

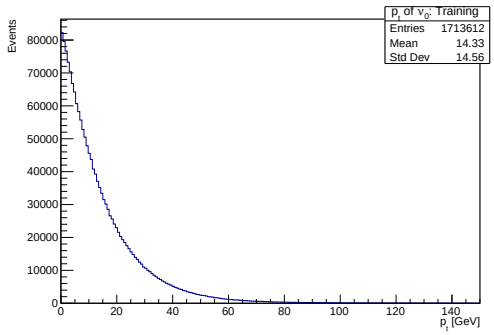


(a) Pseudorapidity of the neutrino originating from the leading tau.

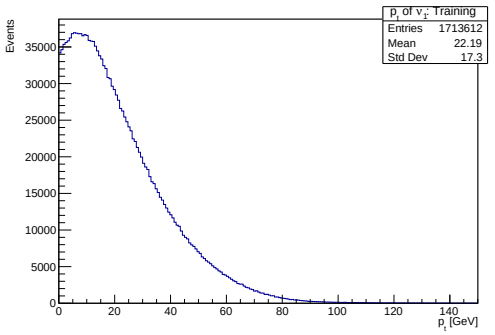


(b) Pseudorapidity of the neutrino originating from the subleading tau.

Figure B.12: True pseudorapidity of the neutrinos. The events selected for **training** are depicted.

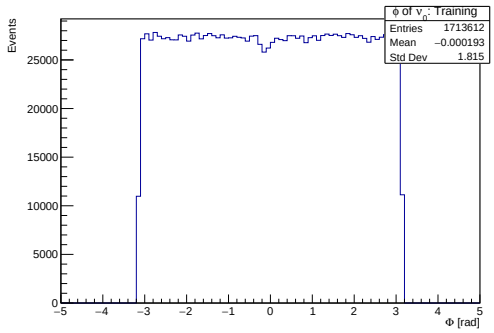


(a) Transverse momentum of the neutrino originating from the leading tau.

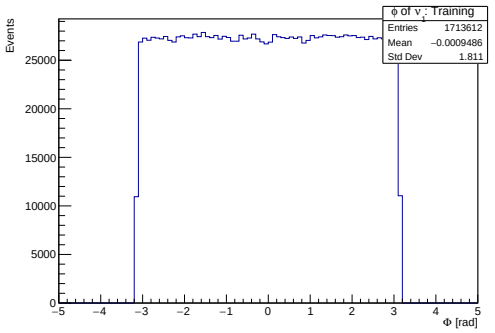


(b) Transverse momentum of the neutrino originating from the subleading tau.

Figure B.13: True transverse momentum of the neutrinos. The events selected for **training** are depicted.

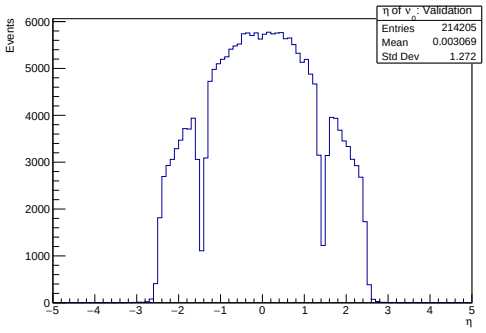


(a) Azimuthal angle of the neutrino originating from the leading tau.

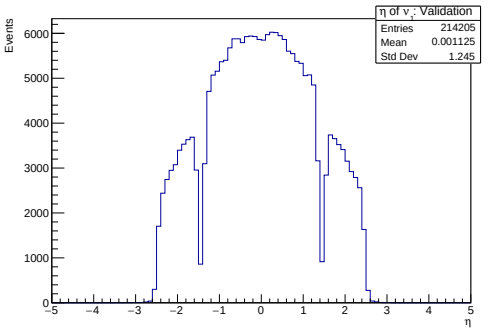


(b) Azimuthal angle of the neutrino originating from the subleading tau.

Figure B.14: True azimuthal angle of the neutrinos. The events selected for **training** are depicted.

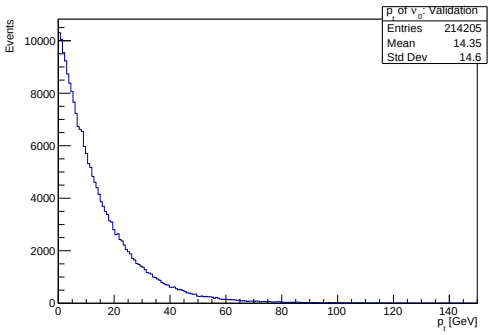


(a) Pseudorapidity of the neutrino originating from the leading tau.

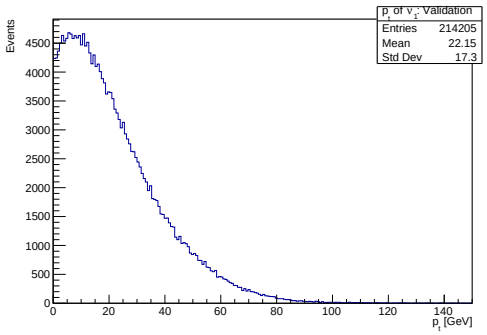


(b) Pseudorapidity of the neutrino originating from the subleading tau.

Figure B.15: True pseudorapidity of the neutrinos. The events selected for **validation** are depicted.

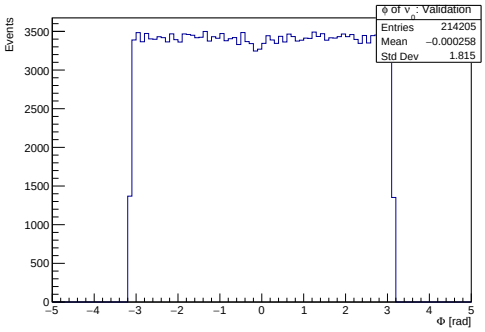


(a) Transverse momentum of the neutrino originating from the leading tau.

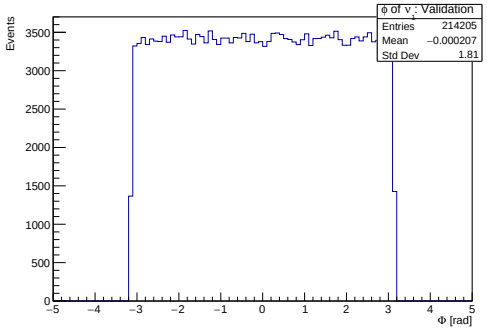


(b) Transverse momentum of the neutrino originating from the subleading tau.

Figure B.16: True transverse momentum of the neutrinos. The events selected for **validation** are depicted.

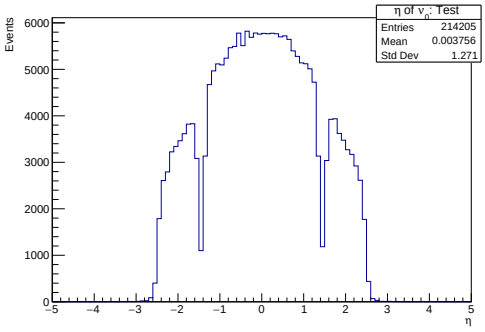


(a) Azimuthal angle of the neutrino originating from the leading tau.

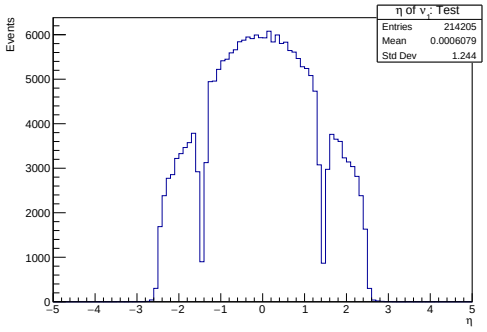


(b) Azimuthal angle of the neutrino originating from the subleading tau.

Figure B.17: True azimuthal angle of the neutrinos. The events selected for **validation** are depicted.

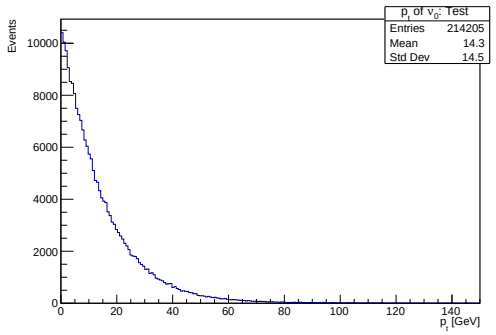


(a) Pseudorapidity of the neutrino originating from the leading tau.

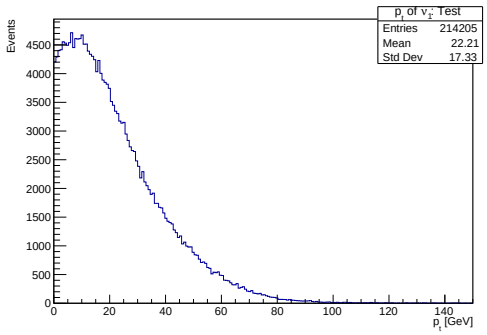


(b) Pseudorapidity of the neutrino originating from the subleading tau.

Figure B.18: True pseudorapidity of the neutrinos. The events selected for **testing** are depicted.

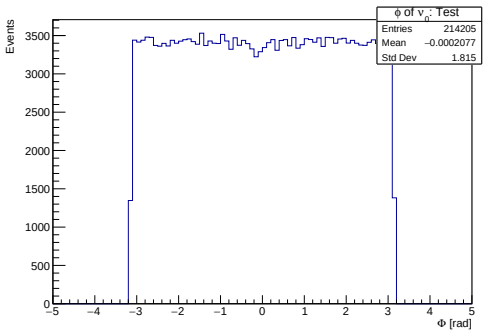


(a) Transverse momentum of the neutrino originating from the leading tau.

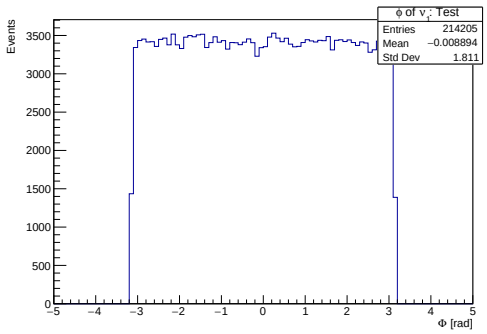


(b) Transverse momentum of the neutrino originating from the subleading tau.

Figure B.19: True transverse momentum of the neutrinos. The events selected for **testing** are depicted.



(a) Azimuthal angle of the neutrino originating from the leading tau.

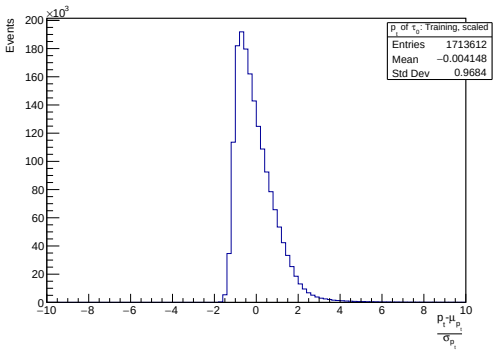


(b) Azimuthal angle of the neutrino originating from the subleading tau.

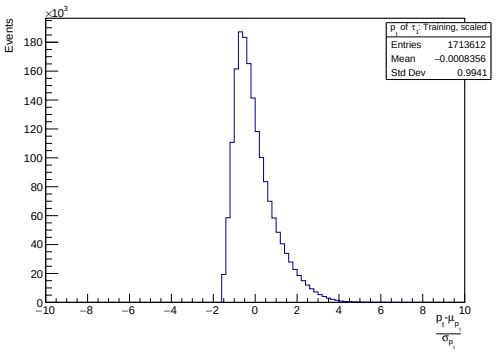
Figure B.20: True azimuthal angle of the neutrinos. The events selected for **testing** are depicted.

B.3 Scaled Input Distributions

The following figures B.21–B.25 belong to section 5.3.4 and illustrate standardized input variables. The events selected for training are depicted for the input variables on par with the MMC (section 5.3.1).

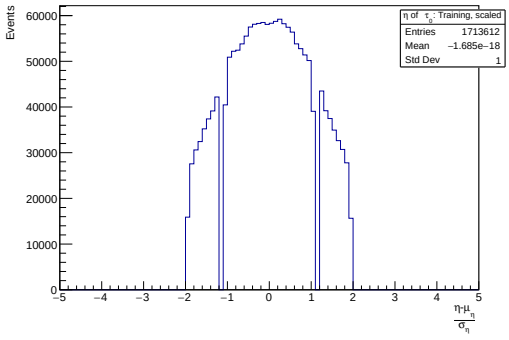


(a) Scaled transverse momentum of the leading tau.

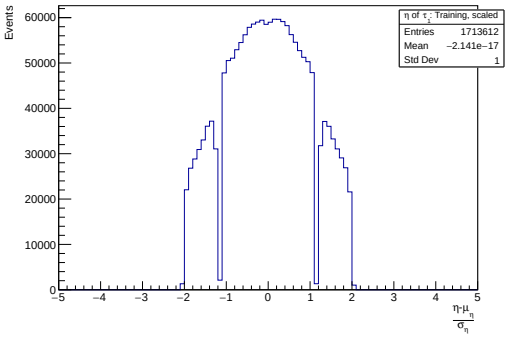


(b) Scaled transverse momentum of the subleading tau.

Figure B.21: Scaled transverse momentum of the taus. The events selected for training are depicted.

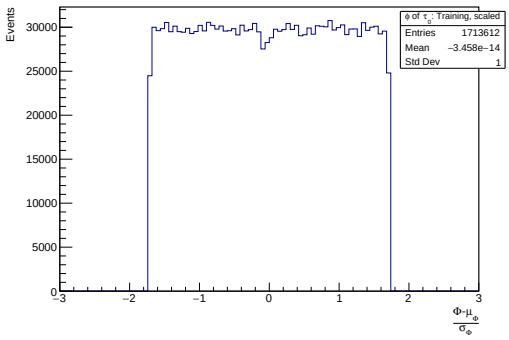


(a) Scaled pseudorapidity of the leading tau.

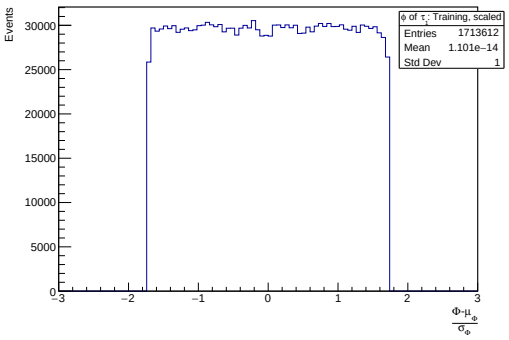


(b) Scaled pseudorapidity of the subleading tau.

Figure B.22: Scaled pseudorapidity of the taus. The events selected for training are depicted.

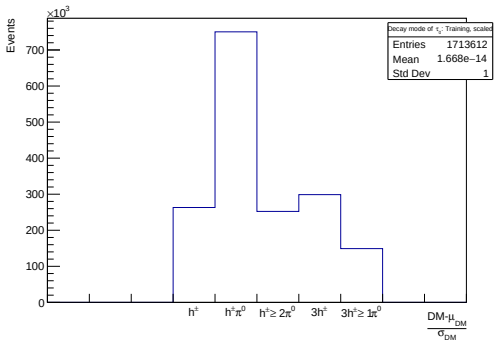


(a) Scaled azimuthal angle of the leading tau.

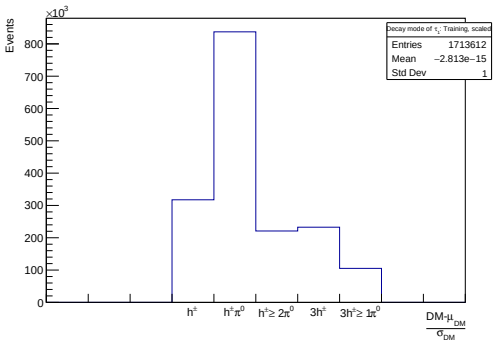


(b) Scaled azimuthal angle of the subleading tau.

Figure B.23: Scaled azimuthal angle of the taus. The events selected for training are depicted.

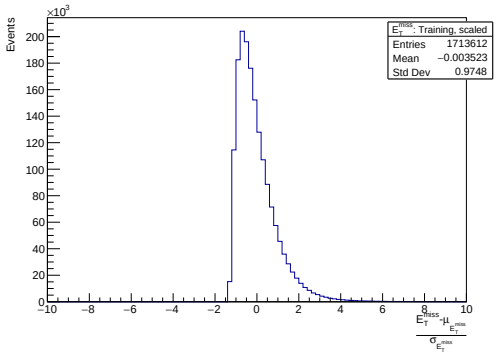


(a) Scaled decay mode of the leading tau.

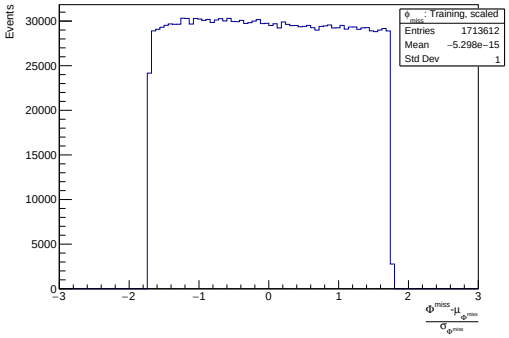


(b) Scaled decay mode of the subleading tau.

Figure B.24: Scaled decay mode of the taus. The events selected for training are depicted.



(a) Scaled missing transverse energy



(b) Scaled azimuthal angle of the missing transverse energy.

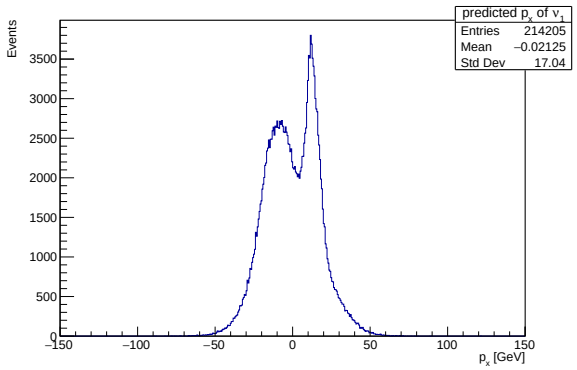
Figure B.25: Scaled missing transverse energy and the corresponding scaled azimuthal angle. The events selected for training are depicted.

APPENDIX C

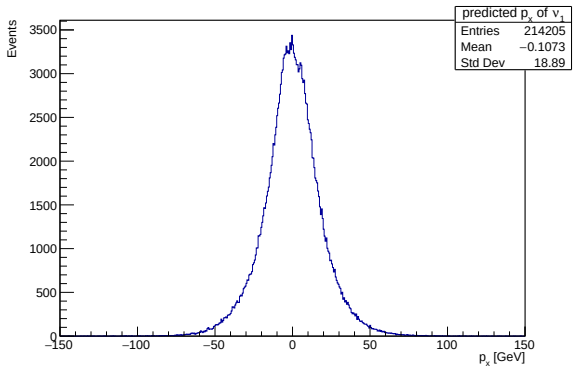
Further Plots of Chapter 7

C.1 The Target: Influence of the Momentum Representation

The figures C.1–C.3 can be considered additionally to section 7.1.

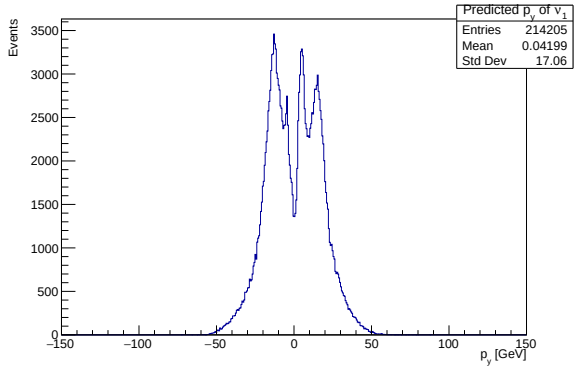


(a) Predicted p_x. The training was performed with inputs on **reconstruction level**.

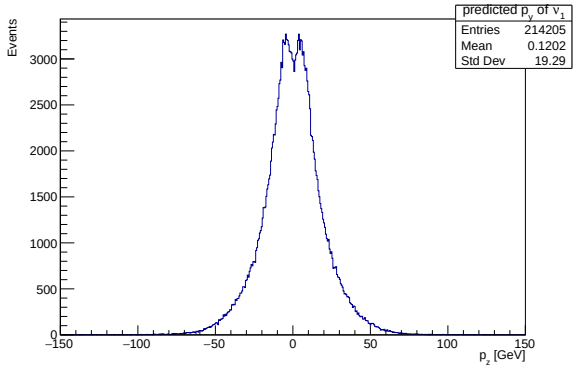


(b) Predicted p_x. The training was performed with inputs on **truth level**.

Figure C.1: Predicted momentum in x-direction for the neutrino originating from the subleading tau. The training was performed on the basic setup introduced in section 6.1. The presented prediction is based on the events selected for testing.

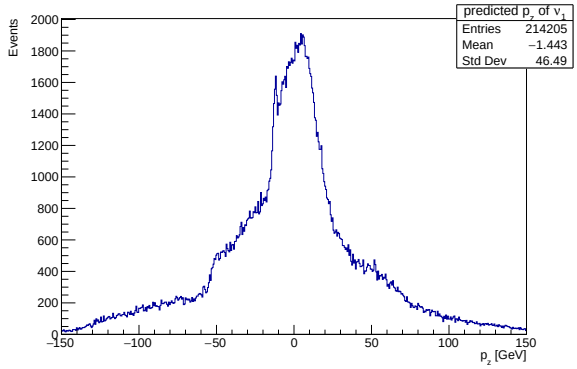


(a) Predicted p_y. The training was performed with inputs on **reconstruction level**.

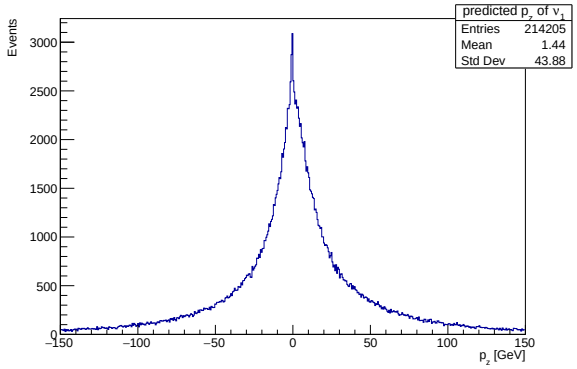


(b) Predicted p_y. The training was performed with inputs on **truth level**.

Figure C.2: Predicted momentum in y-direction for the neutrino originating from the subleading tau. The training was performed on the basic setup introduced in section 6.1. The presented prediction is based on the events selected for testing.



(a) Predicted p_z. The training was performed with inputs on **reconstruction level**.

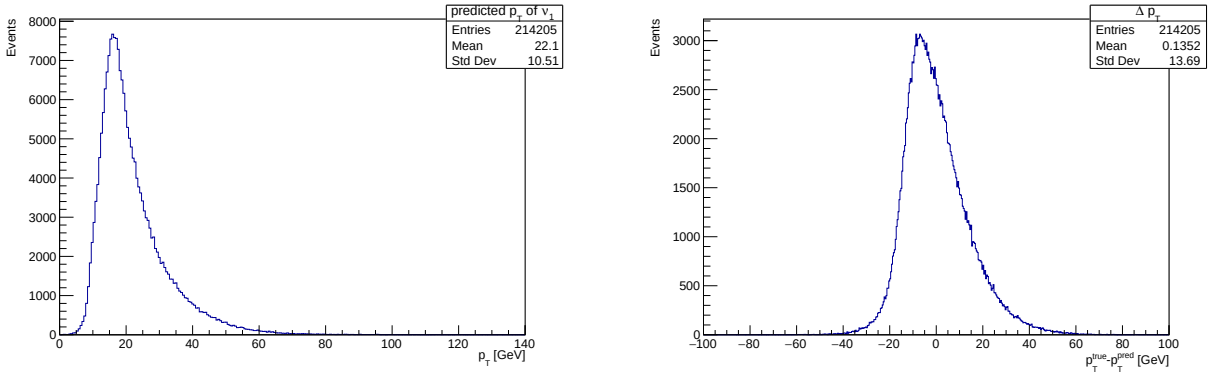


(b) Predicted p_z. The training was performed with inputs on **truth level**.

Figure C.3: Predicted momentum in z-direction for the neutrino originating from the subleading tau. The training was performed on the basic setup introduced in section 6.1. The presented prediction is based on the events selected for testing.

C.2 Transformation of the Transverse Momentum

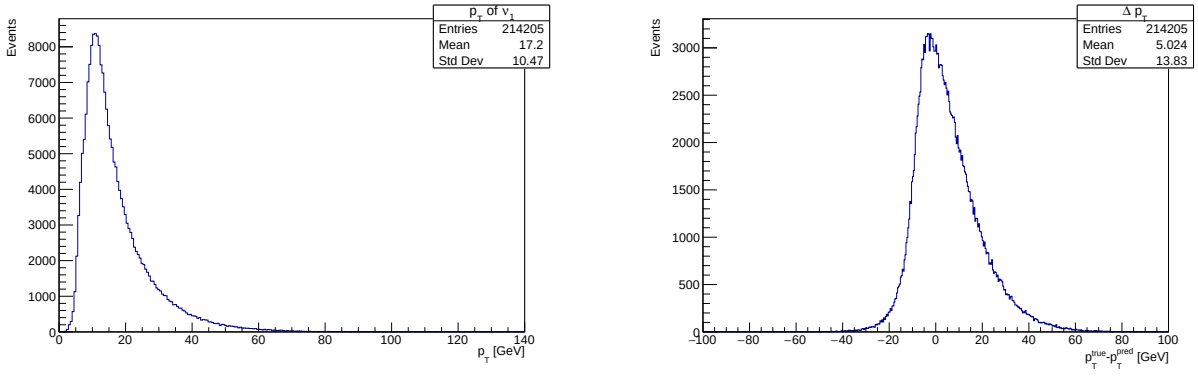
The figures C.4 and C.5 can be considered additionally to section 7.2.



(a) Frequency distribution of the predicted p_T.

(b) Resolution of p_T.

Figure C.4: Predicted transverse momentum of the neutrino originating from the subleading tau for the training on the basic setup. The prediction is based on the events selected for testing.



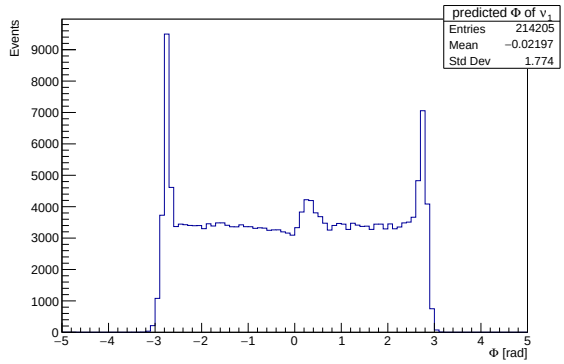
(a) Backtransformed prediction of p_T.

(b) Resolution of the backtransformed p_T.

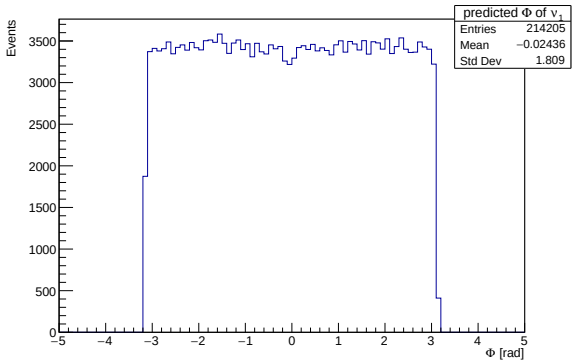
Figure C.5: Backtransformed prediction of the transverse momentum. The basic model was trained with the logarithm of the transverse momentum as target. Thus, the backtransformation corresponds to the exponential of the model's output. The prediction was performed on the events selected for testing and corresponds to the neutrino originating from the subleading tau.

C.3 Periodicity of the Azimuthal Angle

The following figures belong to section 7.4 and show the influence of the Φ-periodicity correction for the neutrino originating from the subleading tau.

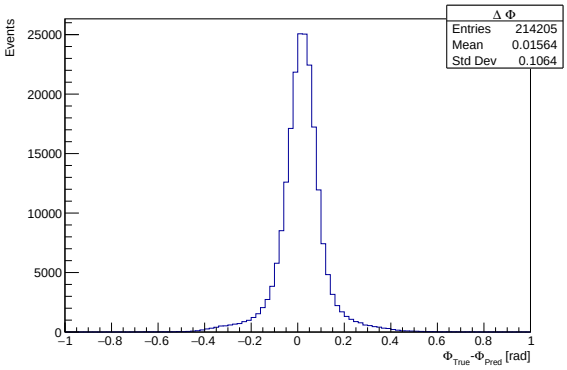


(a) Prediction based on a model trained with the common mean-squared-error.

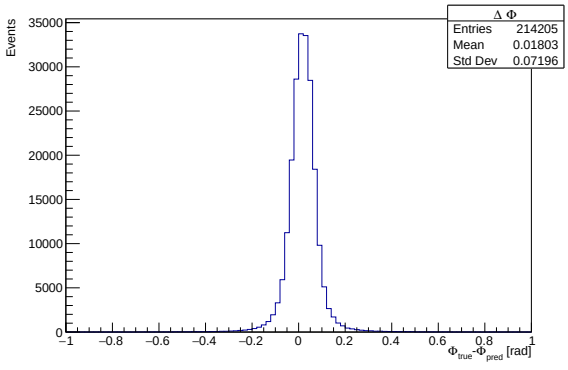


(b) Prediction based on a model considering a Φ-periodicity correction.

Figure C.6: Influence of the periodicity correction on the azimuthal angle. The prediction is based on the evaluation with the events selected for testing and represents the azimuthal angle of the neutrino originating from the subleading tau.



(a) Φ-resolution based on a model trained with the common mean-squared-error.

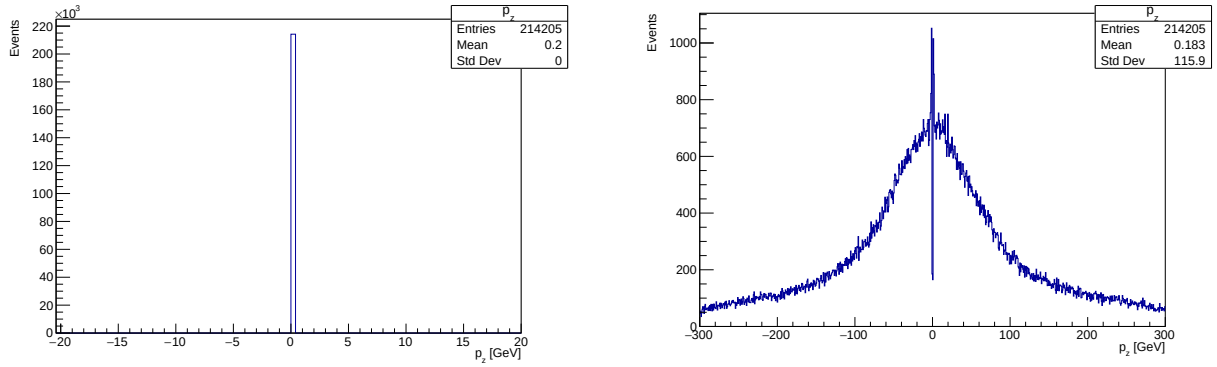


(b) Φ-resolution based on a model considering a periodicity correction.

Figure C.7: Influence of the Φ-periodicity correction illustrated by resolution plots of the azimuthal angle. The prediction was performed on the events selected for testing and corresponds to the neutrino originating from the subleading tau.

C.4 Ambiguities

The following figure belongs to section 7.5 and visualizes the influence of ambiguities in the prediction of p_z.



(a) Predicted p_z for the inputs E, p_x and p_y.

(b) Predicted p_z for the inputs E, p_x,p_y and the sign of p_z.

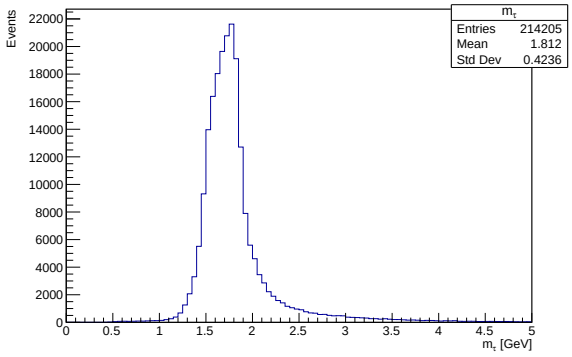
Figure C.8: Verification of the toy model study using the γ^* sample. The predicted p_z of the neutrino originating from the leading tau is depicted for a training based on the inputs E, p_x and p_y (a) as well as E, p_x,p_y and the sign of p_z (b). In the latter, predictions of zero seem to be avoided and small negative or small positive momenta are preferred instead. Nevertheless, the general observations are in accordance with those of figure 7.14.

APPENDIX D

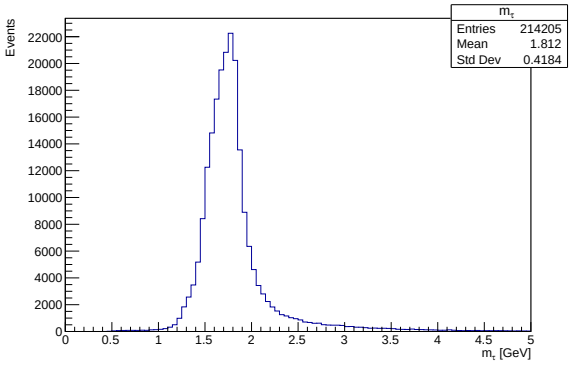
Further Plots of Chapter 8

D.1 The Soft Constraint Approach

The following figures belong to section 8.1.1 investigating soft constrained ANNs for the di-tau mass reconstruction.

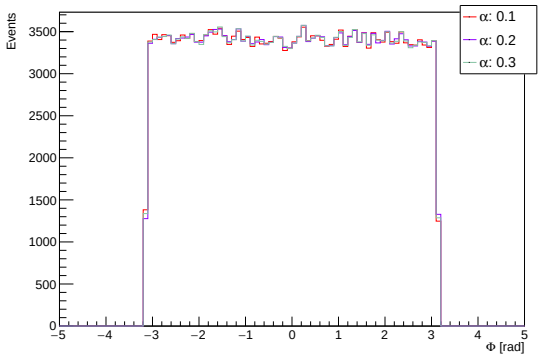


(a) Determined tau mass for a soft constrained model using a penalty weighting of 0.2.

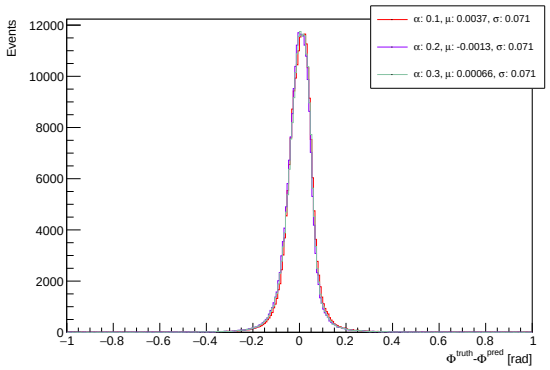


(b) Determined tau mass for a soft constrained model using a penalty weighting of 0.3.

Figure D.1: Impact of different penalty weightings on the mass of the leading tau determined from the predicted neutrino variables using equation 2.16.

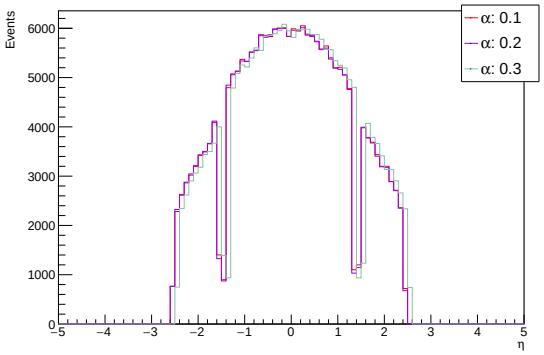


(a) Frequency distributions of the predicted azimuthal angle for different penalty weightings α.

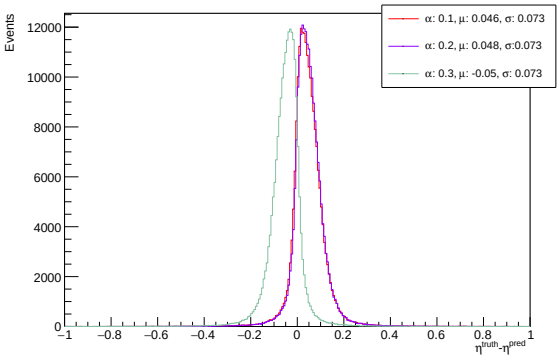


(b) Resolution of the azimuthal angle for different penalty weightings α.

Figure D.2: Predicted azimuthal angle and its resolution. Different penalty weightings α are considered. The predictions are based on the events selected for testing and correspond to the neutrino of the subleading tau.

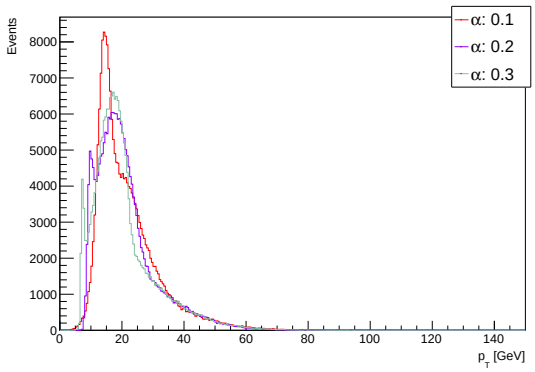


(a) Frequency distributions of the predicted pseudorapidity for different penalty weightings α.

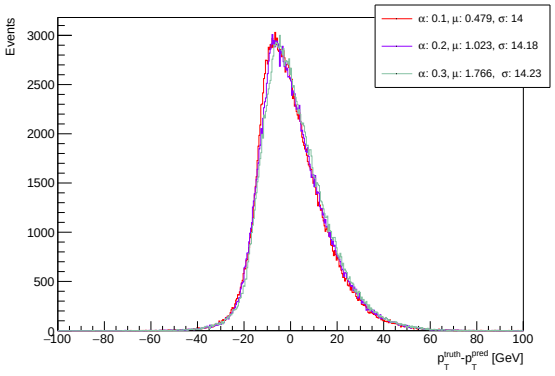


(b) Resolution of the pseudorapidity for different penalty weightings α.

Figure D.3: Predicted pseudorapidity and its resolution. Different penalty weightings α are considered. The predictions are based on the events selected for testing and correspond to the neutrino of the subleading tau.

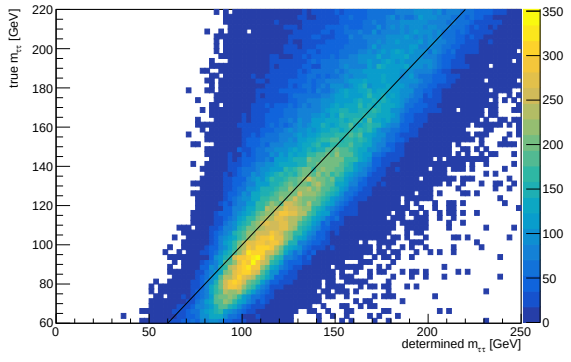


(a) Frequency distributions of the transverse momentum for different penalty weightings α.

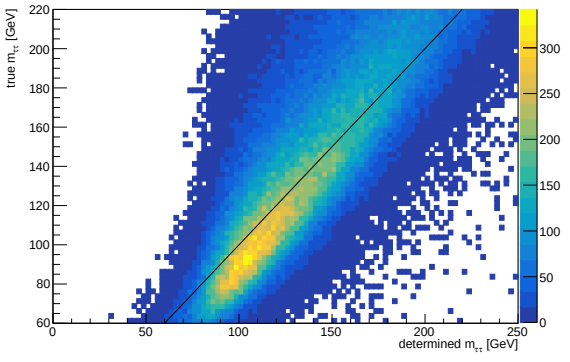


(b) Resolution of the transverse momentum for different penalty weightings α.

Figure D.4: Predicted transverse momentum and its resolution. Different penalty weightings α are considered. The predictions are based on the events selected for testing and correspond to the neutrino of the subleading tau.

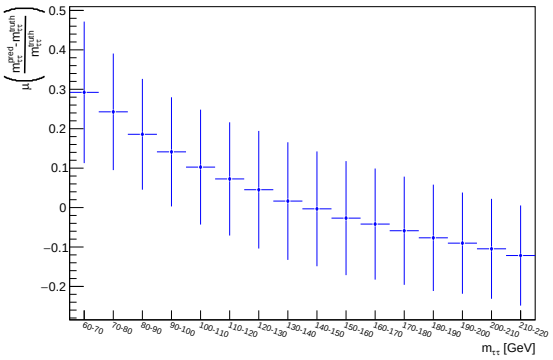


(a) Scatter plot based on the model trained with a penalty weighting of $\alpha = 0.2$. Correlation: 79.8%.

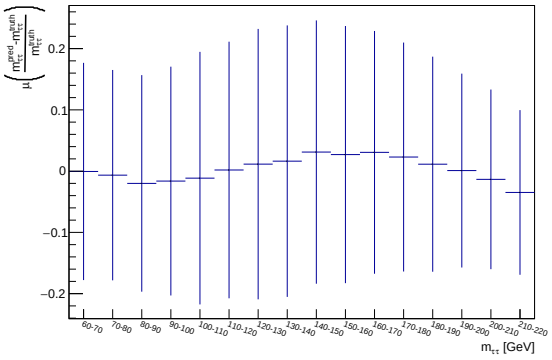


(b) Scatter plot based on the model trained with a penalty weighting of $\alpha = 0.3$. Correlation: 79.7%.

Figure D.5: Scatter plots of the di-tau mass for different penalty weightings. For illustration, the desired linear relation is denoted by the black line. The models were evaluated on the events selected for testing.

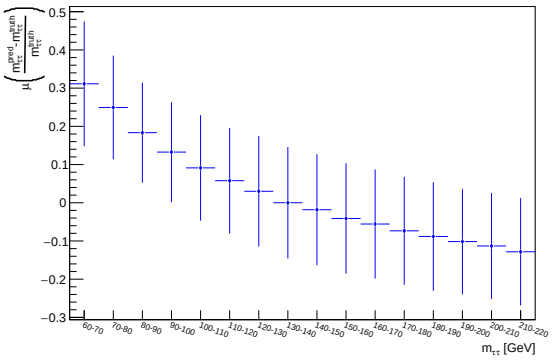


(a) Binned resolution plot.

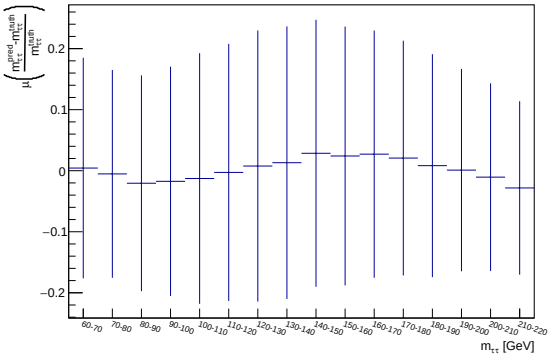


(b) Calibrated binned resolution plot.

Figure D.6: Binned resolution plots in bins of the true mass fo a penalty weighting of 0.1. Figure (a) is produced from the direct network output, whereas in (b) the calibration introduced in section 6.2.4 was applied. The model was evaluated using the events selected for testing.

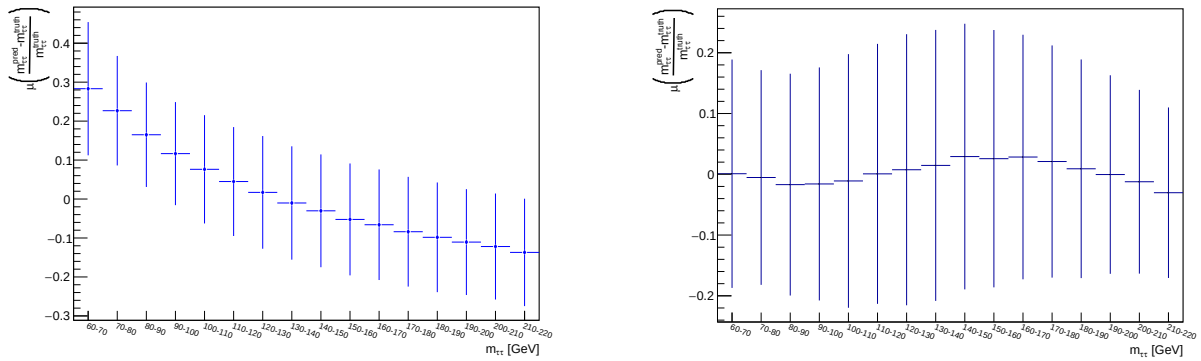


(a) Binned resolution plot.



(b) Calibrated binned resolution plot.

Figure D.7: Binned resolution plots in bins of the true mass fo a penalty weighting of 0.2. Figure (a) is produced from the direct network output, whereas in (b) the calibration introduced in section 6.2.4 was applied. The model was evaluated using the events selected for testing.



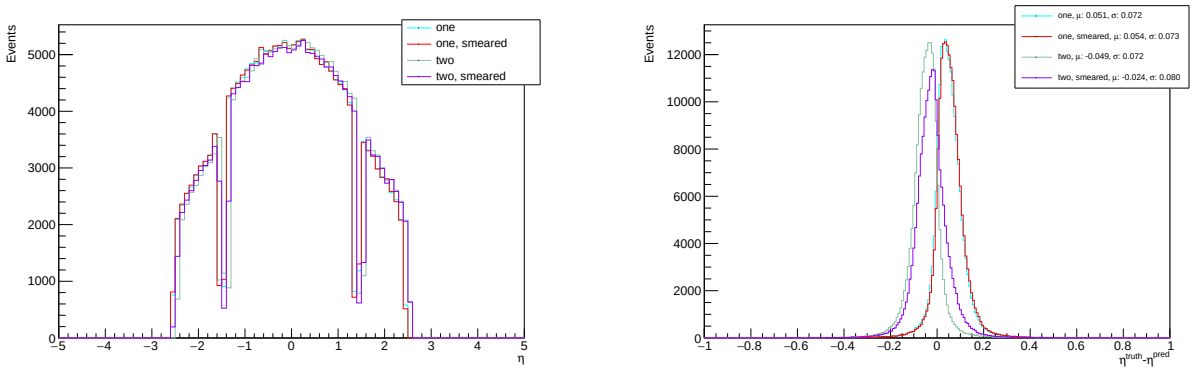
(a) Binned resolution plot.

(b) Calibrated binned resolution plot.

Figure D.8: Binned resolution plots in bins of the true mass fo a penalty weighting of 0.3. Figure (a) is produced from the direct network output, whereas in (b) the calibration introduced in section 6.2.4 was applied. The model was evaluated using the events selected for testing.

D.2 The Hard Constraint Approach

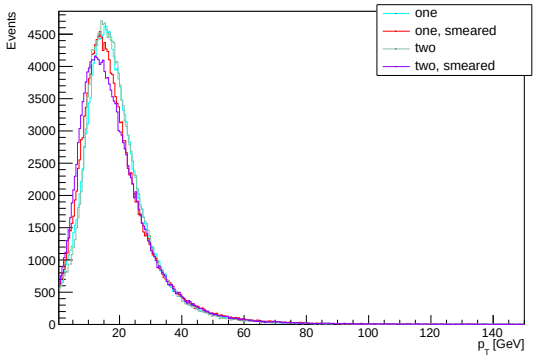
The following figures belong to section 8.1.2 investigating setups relying on the hard constraint.



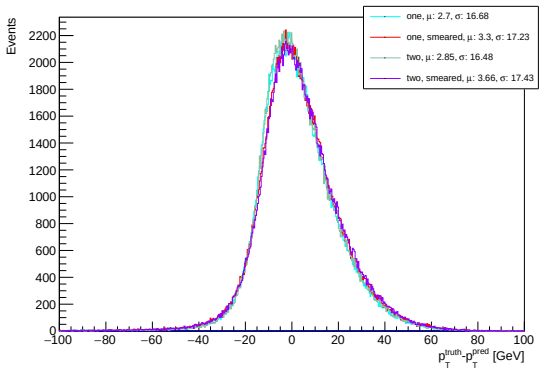
(a) Frequency distribution of the predicted pseudorapidities.

(b) Resolutions of the predicted pseudorapidities.

Figure D.9: Predicted pseudorapidity and its resolution. The results of the different hard constrained setups are depicted. Their description is given in section 8.1.2. The predictions are based on the events selected for testing and correspond to the neutrino associated to the subleading tau. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.

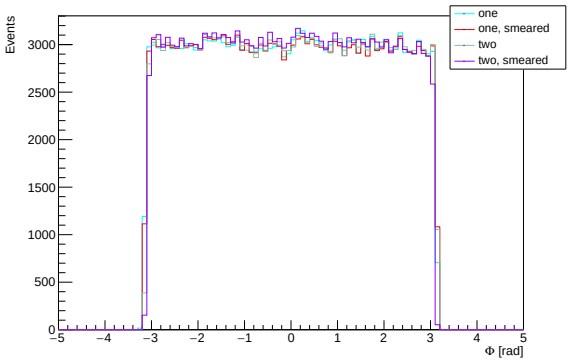


(a) Frequency distribution of the determined transverse momentum.

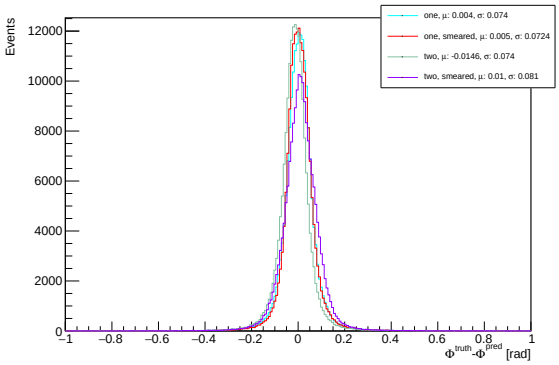


(b) Resolutions of the determined transverse momentum.

Figure D.10: Determined transverse momentum and its resolution. The results of the different hard constrained setups are depicted. Their description is given in section 8.1.2. The predictions are based on the events selected for testing and correspond to the neutrino associated to the subleading tau. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.

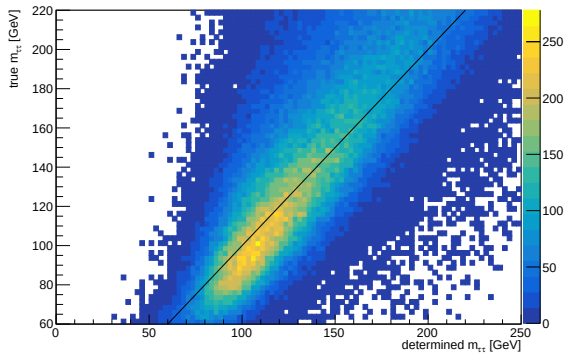


(a) Frequency distribution of the azimuthal angle momentum.

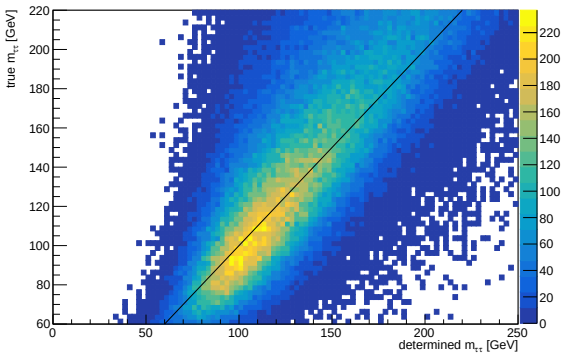


(b) Resolutions of the azimuthal angle.

Figure D.11: Predicted azimuthal angle and its resolution. The results of the different hard constrained setups are depicted. Their description is given in section 8.1.2. The predictions are based on the events selected for testing and correspond to the neutrino associated to the subleading tau. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.



(a) Scatter plot based on the hard constrained model. Correlation: 76.1%.



(b) Scatter plot based on the hard constrained model assuming a gaussian mass distribution of $\sigma = 0.18$. Correlation: 74.3%.

Figure D.12: Scatter plots for the hard constrained models using one ANN. For illustration, the desired linear relation is denoted by the black line. The model was evaluated on the events selected for testing. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.

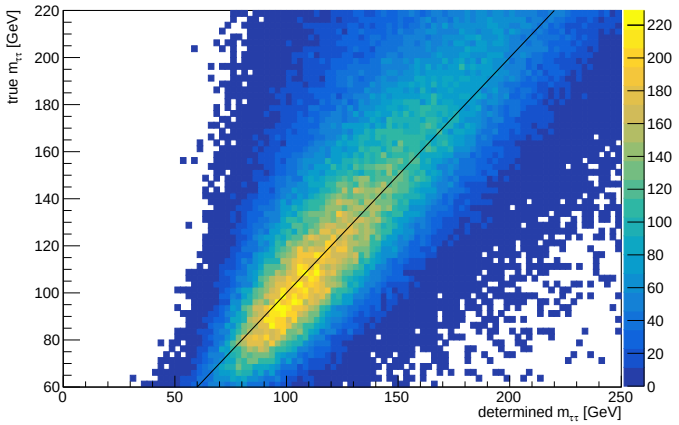
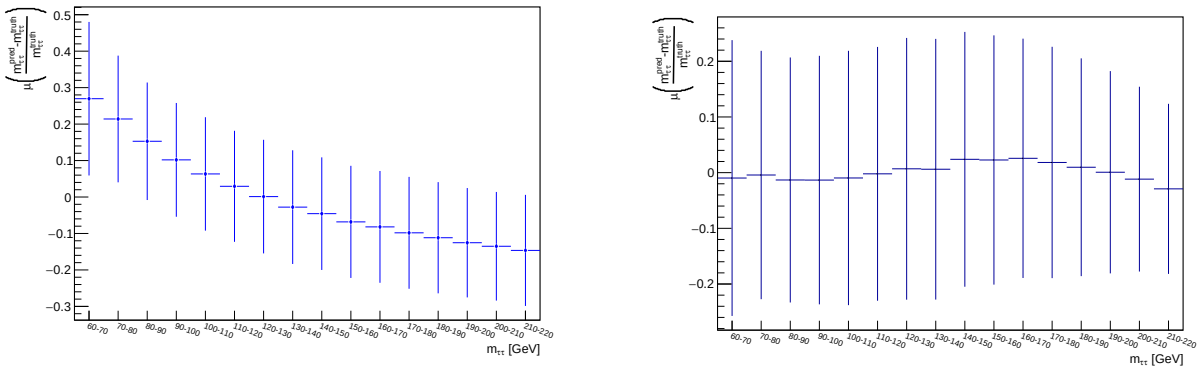


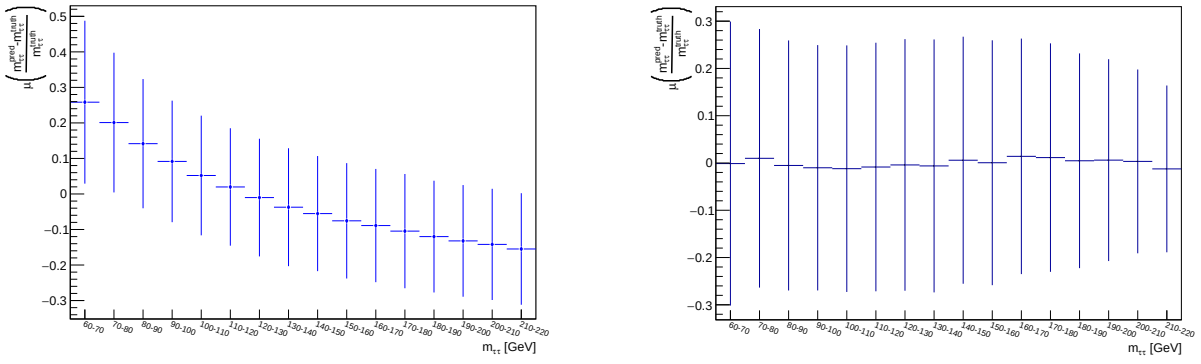
Figure D.13: Scatter plot of the hard constrained model using two ANNs and a gaussian mass distribution of $\sigma = 0.18$. For illustration, the desired linear relation is denoted by the black line. The model was evaluated on the events selected for testing. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319. Correlation: 73.7%.



(a) Binned resolution plot.

(b) Calibrated binned resolution plot

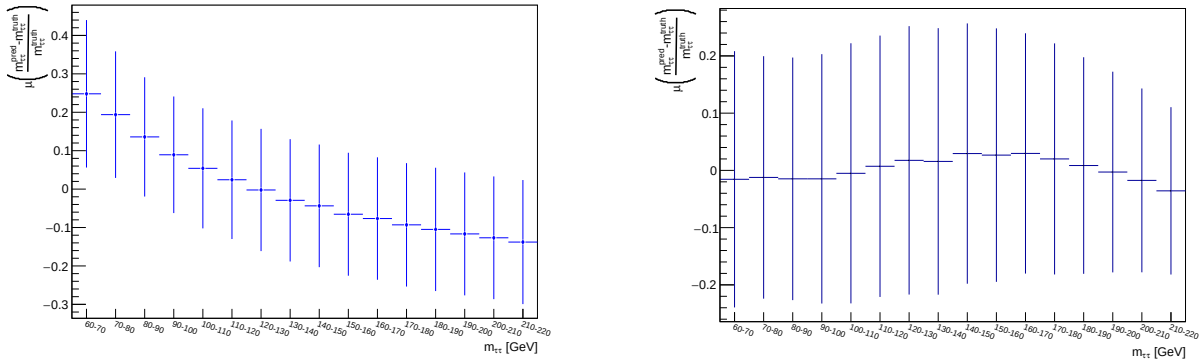
Figure D.14: Binned resolution plot for the hard constrained model without (a) and with (b) the calibration introduced in section 6.2.4. The model was evaluated on the events selected for testing. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.



(a) Binned resolution plot.

(b) Calibrated binned resolution plot

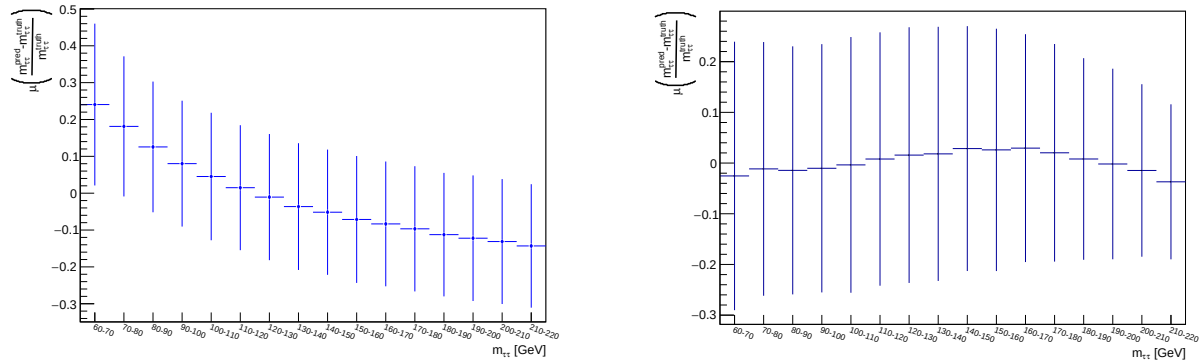
Figure D.15: Binned resolution plot for the hard constrained model assuming a gaussian tau mass distribution with $\sigma = 0.18$ GeV. The plot is shown without (a) and with (b) the calibration introduced in section 6.2.4. The model was evaluated on the events selected for testing. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.



(a) Binned resolution plot.

(b) Calibrated binned resolution plot

Figure D.16: Binned resolution plot for the di-tau mass determination based on the two separated hard constrained models, without (a) and with (b) the calibration introduced in section 6.2.4. The model was evaluated on the events selected for testing. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.



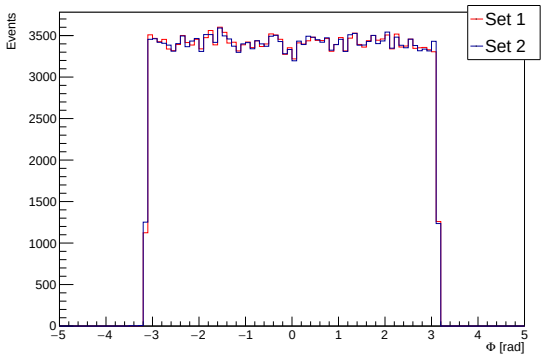
(a) Binned resolution plot.

(b) Calibrated binned resolution plot

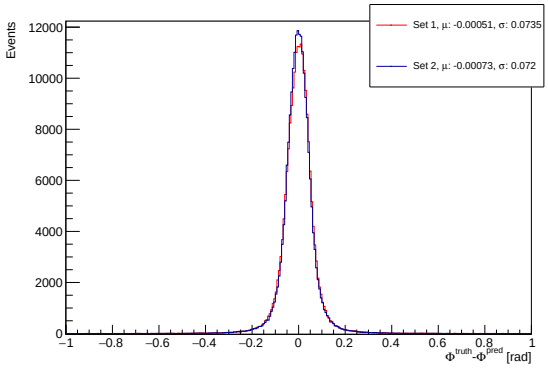
Figure D.17: Binned resolution plot for the di-tau mass determination based on the separated hard constrained models assuming a gaussian tau mass distribution with $\sigma = 0.18\,\text{GeV}$. The plot is shown without (a) and with the calibration (b) introduced in section 6.2.4. The model was evaluated on the events selected for testing. Events with unphysical calculations of p_T are excluded which reduces the initial 214205 events to 187319.

D.3 Beyond the MMC

The following figures belong to section 8.2.2 investigating setups relying on inputs beyond the MMC.

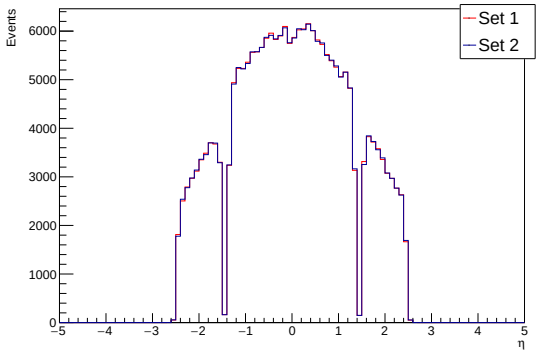


(a) Predicted azimuthal angle.

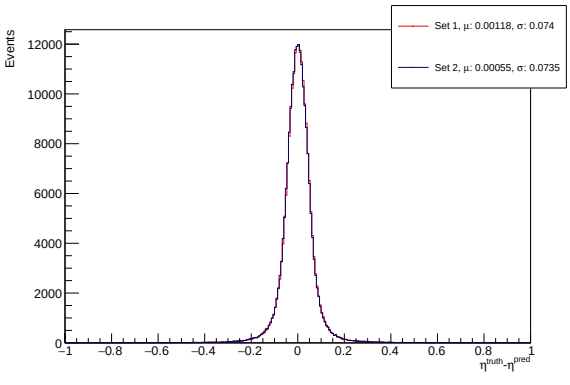


(b) Resolution of the azimuthal angle.

Figure D.18: Predicted azimuthal angle and its resolution. The results for the models trained with the two sets of input variables beyond the MMC, described in section 8.2.1, are shown. The evaluation of the ANNs was performed on the events selected for testing and is depicted for the neutrino of the subleading tau.

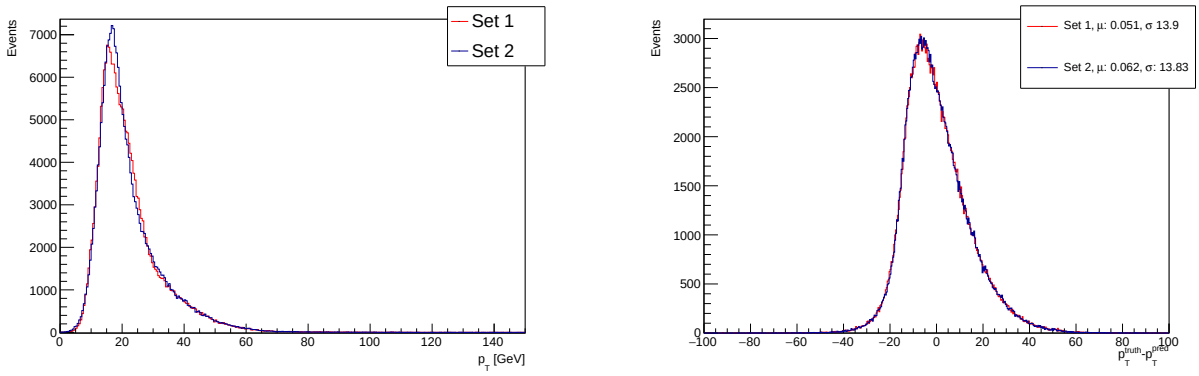


(a) Predicted pseudorapidity.



(b) Resolution of the pseudorapidity.

Figure D.19: Predicted pseudorapidity and its resolution. The results for the models trained with the two sets of input variables beyond the MMC, described in section 8.2.1, are shown. The evaluation of the ANNs was performed on the events selected for testing and is depicted for the neutrino of the subleading tau.



(a) Predicted transverse momentum.

(b) Resolution of the transverse momentum.

Figure D.20: Predicted transverse momentum and its resolution. The results for the models trained with the two sets of input variables beyond the MMC, described in section 8.2.1, are shown. The evaluation of the ANNs was performed on the events selected for testing and and is depicted for the neutrino of the subleading tau.

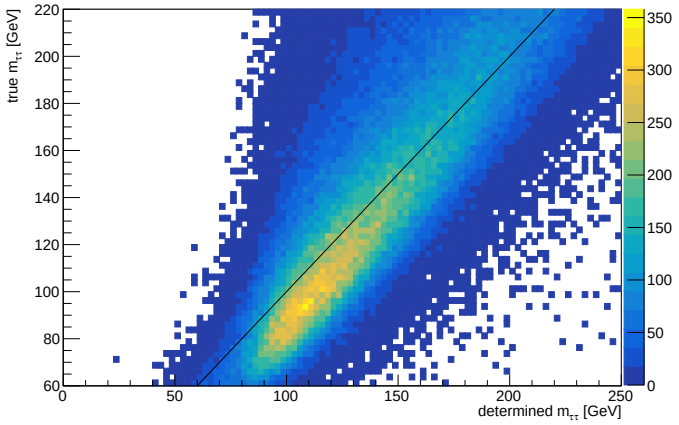
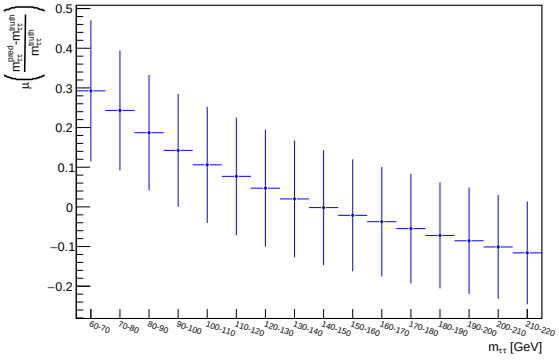
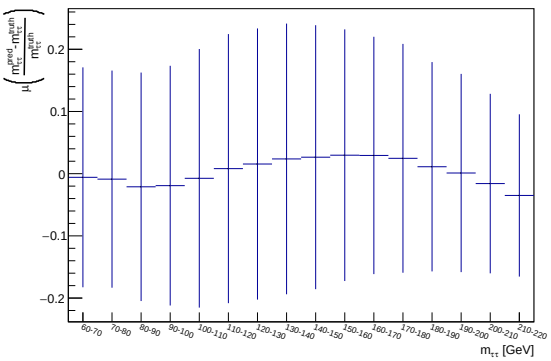


Figure D.21: Scatter plot of the di-tau mass based on the model trained with the input *Set 2*. The black line denotes the desired linear relation. Correlation factor: 80.7%.

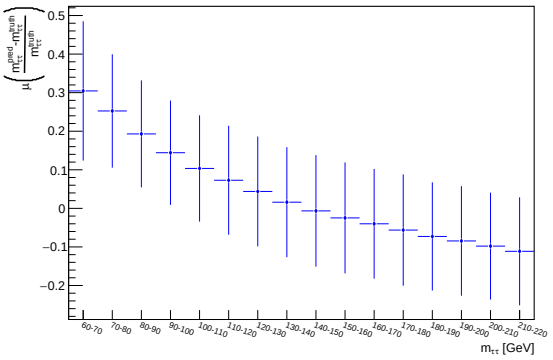


(a) Binned resolution plot.

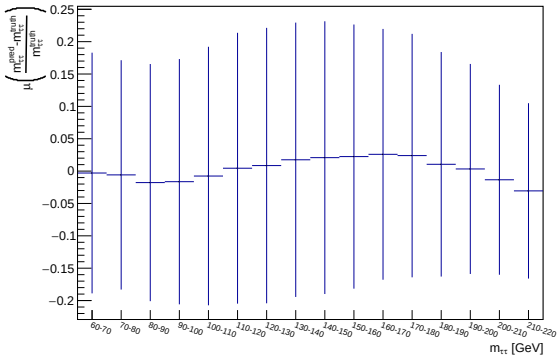


(b) Calibrated binned resolution plot

Figure D.22: Binned resolution plots before and after the calibration described in section 6.2.4. The underlying model was trained with input *Set 1* defined in section 8.2.1 and evaluated using the events selected for testing.



(a) Binned resolution plot.

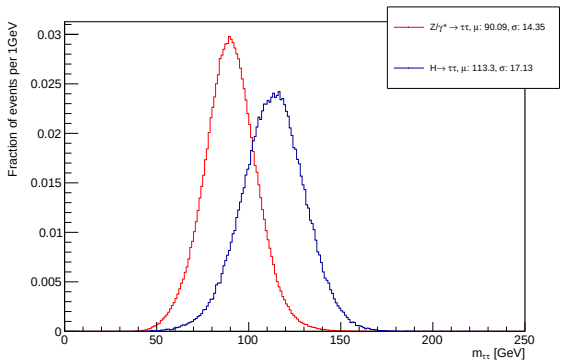


(b) Calibrated binned resolution plot

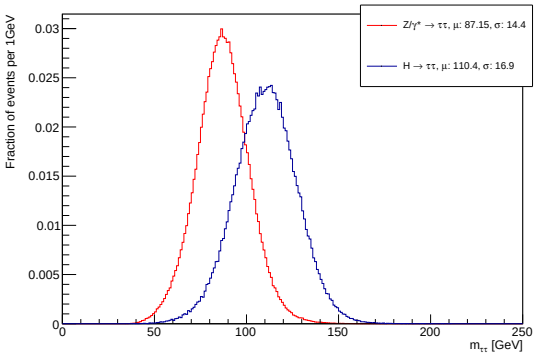
Figure D.23: Binned resolution plots before and after the calibration described in sesction 6.2.4. The underlying model was trained with input *Set 2* defined in section 8.2.1 and evaluated using the events selected for testing.

APPENDIX E

Further Plots of Chapter 9

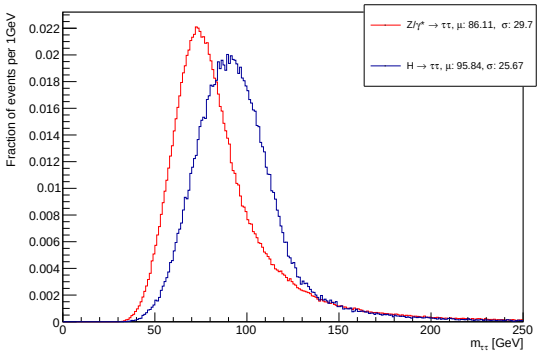


(a) Overlap for the di-tau mass reconstruction based on the soft constrained model using a penalty weighting of 0.2: 44.04%.

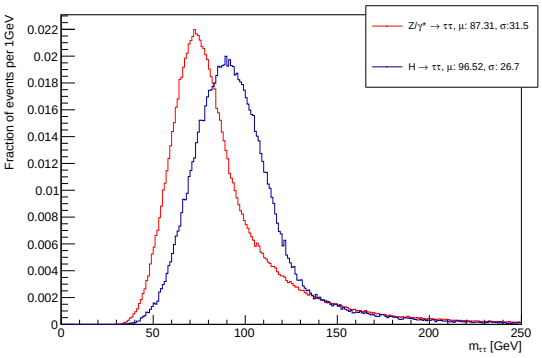


(b) Overlap for the di-tau mass reconstruction based on the soft constrained model using a penalty weighting of 0.3: 43.91%.

Figure E.1: Normalized determined di-tau masses for the Z and the Higgs sample using different soft constrained models.

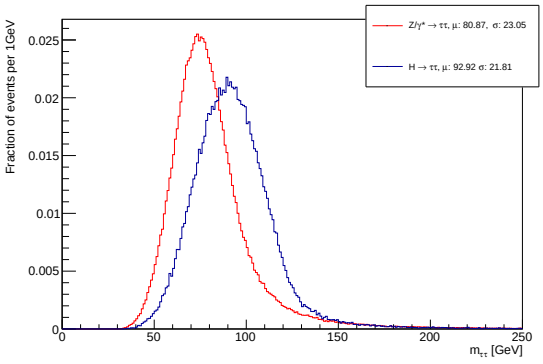


(a) Overlap for the di-tau mass reconstruction based on the hard constrained model using one network and a gaussian distributed tau mass: 73.28%.

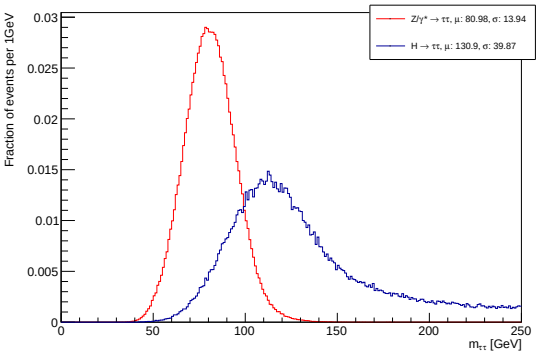


(b) Overlap for the di-tau mass reconstruction based on the hard constrained model using one network: 73.23%.

Figure E.2: Normalized determined di-tau masses for the Z and the Higgs sample using different hard constrained models.



(a) Overlap for the di-tau mass reconstruction based on the hard constrained model using two networks and a gaussian distributed tau mass: 69.01%.



(b) Overlap for the di-tau mass reconstruction based on the model trained with the input data beyond the MMC (*Set 2*): 29.45%.

Figure E.3: Normalized determined di-tau masses for the Z and the Higgs sample using different models.

List of Figures

List of Tables

Acronyms

ANNs artificial neural networks. 1, 61, 68, 71

BDTs Boosted Decision Trees. 21

BRT Boosted Regression Tree. 22

CB Combined Muon. 21

CSCs Cathode Strip Chambers. 20

CT Calorimeter-Tagged Muon. 21

FSR final state radiation. 25

ID Inner Detector. 21

ISR initial state radiation. 25

LHC Large Hadron Collider. 1

LO leading order. 26

MC Monte Carlo. 2

MDTs Monitored Drift Tubes. 20

ME Extrapolated Muon. 21

MMC Missing Mass Calculator. 1

MS Muon Spectrometer. 21

NLO next-to-leading order. 27

PDFs parton distribution functions. 43

PDFs probability density functions. 25

pp collisions proton-proton collisions. 13

QCD Quantum Chromodynamics. 6

QED Quantum Electro Dynamics. 6

RPCs Resistive Plate Chambers. 20

SCT Semiconductor Tracker. 19

SM Standard Model. 1

ST Segment-Tagged Muon. 21

TGCs Thin Gap Chambers. 20

TRT Transition Radiation Tracker. 19