Ditau Reconstruction with Deep Neural Networks Trained on Spin-1 and Spin-0 Samples

Ramy Hmaid

Masterarbeit in Physik
angefertigt im Physikalischen Institut

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

November 2022

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

Bonn, 18.11.2022 ..... Rawy Hmaid .....

Date Signature

1. Gutachter: Prof. Klaus Desch
2. Gutachter: Prof. Ian Brock

Acknowledgements

I would like to thank my family and friends for supporting me throughout my studies. I would also like give a special thanks to Priv.-Doz. Dr. Philip Bechtle, Prof. Klaus Desch and Dr. Christian Grefe for supporting and supervising me throughout the last year and giving me the opportunity to write my master thesis. I also want to thank Florian Sagolla for helping me with uproot and the preparation for my master colloquium. Furthermore, I would like to thank Jan-Eric Heinrichs for supporing me ever since I started writing my bachelor thesis. Finally, I would like to thank Cirno for helping me with various math problems I had over the last year.

Contents

CHAPTER 1

Introduction

The Standard Model (SM) of particle physics is a theory that was developed throughout the 20th century. It is able to successfully describe three of the four fundamental forces, gravity being the only one that cannot be explained by it yet. In doing so it manages to explain a wide variety of fundamental phenomena of nature with high precision, such as the life of stars and nuclear power. It can even give high precision theoretical values for effects that require a combined analysis of the weak interaction, the strong interaction and the electromagnetic interaction. Some of its successful predictions are the discovery of the Higgs boson, the properties of the weak neutral current, and the W and Z boson. It is also able to combine the various theories of fundamental physics in an internally consistent manner. Despite the significant accomplishments of the Standard Model, there are still a growing number of subtle effects measured by experiments all over the world that leave questions unanswered and demand theoretical explanations that it can not yet offer. These range from more fundamental problems (e.g. how to create a unified theory of gravity and quantum mechanics [25], baryon assymetry in the universe [34] and the origin of the neutrino mass [24]) to divergences between theory and measurement (e.g. the anomalous magnetic dipole moment of the muon [18]). While there are plenty of theoretical approaches to many of these outstanding issues, there is still a lack of proof of these theories.

Many of these effects require very high precision measurements to be detected, which in turn require very accurate experiments and rigorous analysis of the data. The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) is one complex where such experiments can be conducted. It can accelerate particles to energies of up to 13.6 TeV [31]. Only cosmic ray experiments are able to offer higher energies, but they have vastly lower luminosities at those energies. The various experiments at the LHC can measure particle production cross sections or branching ratios to infer the underlying coupling strengths and then compare them to theoretical expectations. It can also measure the masses of known particles or find resonances in the data that might be created by as of yet unknown particles. The greatest achievement of the LHC is the discovery of the Higgs boson by the ATLAS and CMS experiment at the LHC. It was originally measured with a mass of $m_H = 126.0 \pm 0.4(\text{stat.}) \pm 0.4(\text{sys.})$ GeV [36] and is now measured to have a mass of $m_H = 125.10 \pm 0.14$ GeV [12].

In this thesis the problems arising from reconstructing $\text{H} \rightarrow \tau\tau$ events are discussed and different methods of reconstructing ditau events are presented. The focus of this thesis are neural networks trained to optimize the mass reconstruction of ditau events.

Chapter 2 gives a brief introduction of the Standard Model. An overview of the particles of the SM and their properties is presented. In addition, the properties of the fundamental forces described by the SM are explained. The tau lepton, the Higgs boson and the Z boson are described in further detail as they are the most relevant for this thesis, including their production cross sections and their decay channels. In chapter 3 the LHC and the ATLAS detector are introduced. Their capabilities, limitations and the way they function are explained. In chapter 4 the reconstruction methods and capabilities of the ATLAS experiment for some particles are explained. First, the reconstruction of jets is explained, since they are important for the tau lepton reconstruction. Then, an overview of the tau lepton reconstruction is given.

In chapter 5 methods of reconstructing the mass of ditau events are introduced. The first method only includes the visible reconstructed mass of the event. After that, the collinear approximation is presented. Finally, the Missing Mass Calculator (MMC) is introduced. Chapter 6 introduces a final approach for the ditau mass reconstruction. This approach is the focus of this thesis and works by using regressional neural networks trained on Monte Carlo data. In this chapter the general concept of how neural networks function and a more in depth description of the parameters and methods that are relevant for this analysis are discussed.

Chapter 7 introduces the Monte Carlo datasets that are used in the training and evaluation of the neural networks. The mass spectrum of these datasets, the cuts applied to the them, and the spin of the ditau resonances of the datasets is discussed. In chapter 8 the neural network analysis is presented. The performance of neural networks trained on different samples, targets, input variables and loss functions are compared to each other and to the MMC reconstruction. Chapter 9 gives a conclusion and outlook of this thesis.

CHAPTER 2

The Standard Model

The Standard Model (SM) of particle physics combines and accurately explains three of the four fundamental forces of nature in a mathematically coherent manner. It is able to give very accurate predictions of most physical effects on a most fundamental level. Unless otherwise specified, the source for this chapter is [92]

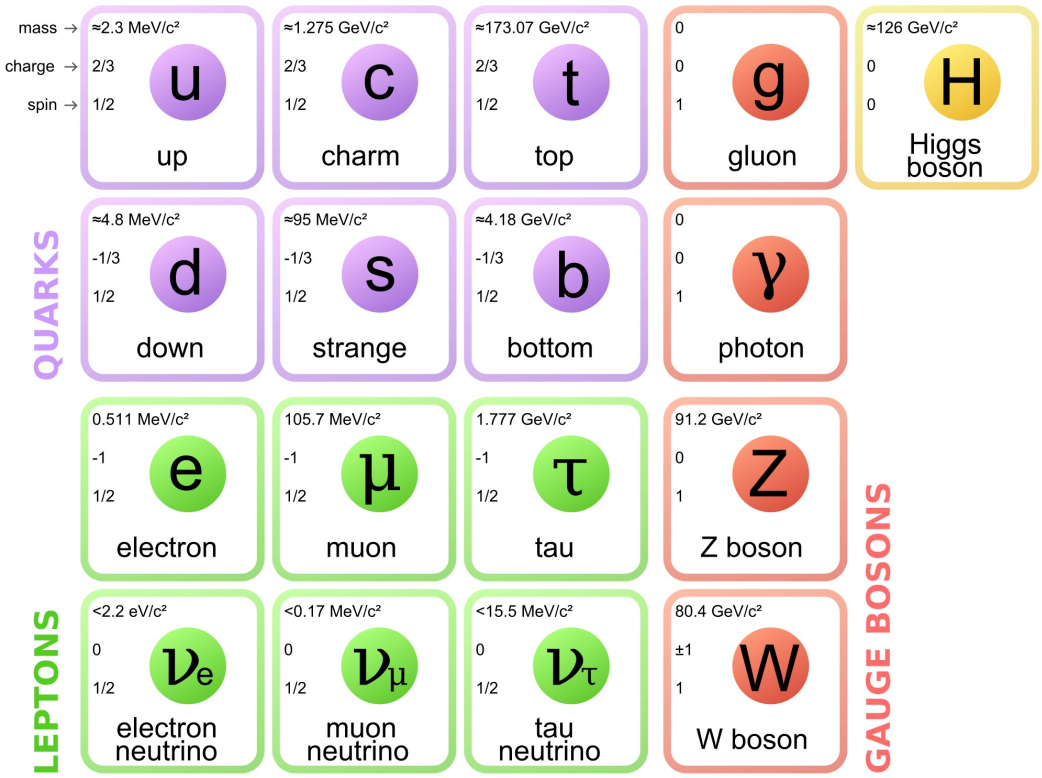


Figure 2.1: Particles of the Standard Model, showing their mass, charge and spin [74].

Fig. 2.1 shows the elementary particles that make up the SM. It also shows the mass, charge and spin of each of the particles. They can be grouped into twelve fermions (left), characterized by

Interaction	Propagators	Relative strength	Range
EM-Interaction	Photon	$\alpha = \frac{1}{137}$ [68]	inf
Weak Interaction	$W^{\pm}, Z$	$\alpha_W \sim 10^{-6}$ [75]	$\sim 10^{-18}$
Strong Interaction	Gluon	$\alpha_S \sim 1$ [47]	$\sim 10^{-15}$
Gravity	Graviton(?)	$\alpha_g \sim 10^{-39}$ [75]	inf

Table 2.1: The four fundamental forces of nature and some of their properties. As the graviton is not proven to exist yet it is marked with a question mark.

their half-integer spin (in this case always $s = 1/2$), and five bosons (right), characterized by their full-integer spin (in this case either $s = 0$ or $s = 1$). The fermions are split into two groups: The leptons (bottom) and the quarks (top). Quarks have color charge and can therefore interact with the strong force, while leptons can not. The fermions are also split into three generations (from left to right), each generation having two quarks and two leptons. The first generation consists of the up quark (u), the down quark (d), the electron lepton (e) and the electron neutrino (ν_e). Most visible matter in the universe is made up of the two quarks and the electron of the first generation. The neutrino is by far the lightest and the quarks are the heaviest of the four particles. The up quark has a charge of $q = +2/3$ e and the down quark has a charge of $q = -1/3$ e. The electron has a charge of $q = -1$ e and the neutrino has no charge. This is a pattern that repeats for the other generations. Only the masses of the particles increase with the generation. The particles of the second generation are the charm quark (c), strange quark (s), muon (μ) and the muon neutrino (ν_μ). In the third and heaviest generation are the top quark (t, the heaviest elementary particle discovered with a mass of $m_t = 173\,\text{GeV}$), bottom quark (b), tau lepton (τ) and the tau neutrino (ν_τ). In addition to those particles, every fermion (with the possible exception of the neutrinos as further discussed in [16]) has a partner antiparticle, which has the same mass and quantum numbers as the standard particles but have an opposite charge.

The bosons are split into a group of four spin 1 gauge bosons ($W^{+/-}$, Z, γ (photon), and g (gluon)), which act as the force carriers of the fundamental forces, and the Higgs boson, which has $s = 0$. The $W^{+/-}$ and Z bosons are the force carriers of the weak force. The $W^{+/-}$-bosons have a charge of $q = \pm 1$ e and a mass of $m_W = 80.38\,\text{GeV}$, while the Z boson has no charge and has a mass of $m_Z = 91.19\,\text{GeV}$ [12]. The photon is the force carrier of the electromagnetic force. It has no charge and no mass. The gluon is the force carrier of the strong force and also has no mass but has color charge.

2.1 Fundamental Interactions

The interactions that are described in the standard model are the electromagnetic interaction, the weak interaction and the strong interaction. Gravity can not yet be explained by a coherent theory that combines the four forces. Tab. 2.1 shows some characteristics of these interactions. Gravity is by far the weakest force, which is why it is uniquely challenging to investigate it in high energy physics experiments. The force carrier of gravity has not yet been discovered but it is named the graviton and is expected to have a spin of 2 since it has to couple to the stress-energy tensor of order 2. It is also assumed to be massless since we assume that gravity has infinite range [73].

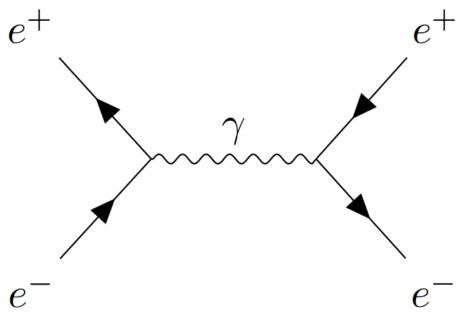


Figure 2.2: A Feynman diagram showing electron-positron scattering (Bhabha scattering). In this case, it is a s-channel scattering.

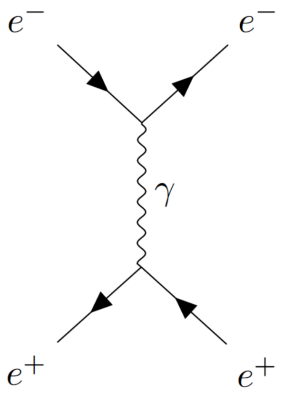


Figure 2.3: A Feynman diagram showing electron-positron scattering (Bhabha scattering). In this case, it is a t-channel scattering.

2.1.1 Feynman diagrams

A clear representation of fundamental interactions is provided by Feynman diagrams. In a Feynman diagram, particles are depicted as lines which meet and interact in vertices. A vertex usually combines three or sometimes four different lines. An example diagram is shown in fig. 2.2. This diagram shows electron-positron scattering, also called Bhabha scattering. Fermions are typically depicted as straight lines. Photons are shown as wavy lines, gluons as helices and other bosons are dashed lines. In this thesis the time axis goes from left to right, indicating that the lines with open ends on the left of the diagram depict initial particles and lines with open ends on the right of the diagram depict final particles of the interaction. If an arrow on the line points in the same direction as the time axis then it is a particle, while it is an antiparticle if it points in the opposite direction. The neutral bosons are their own antiparticles and therefore are drawn without arrows on them.

Feynman diagrams help understand what happens in an interaction and they also tell us which kinds of interactions are possible. One of the most important applications of feynman diagrams is calculating the cross section of certain interactions. At each vertex quantum numbers (such as charge, spin, color charge, etc.) have to be conserved. Energy and momentum do not have to be conserved perfectly as the uncertainty principle allows for temporary energy and momentum violations. Particles that are generated through such violations are called off-shell and can be marked with a star. These

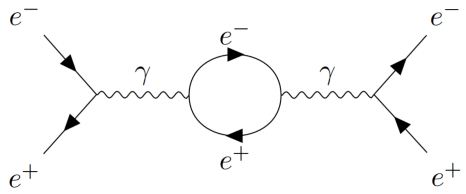


Figure 2.4: A NLO Bhabha scattering feynman diagram. The photon temporarily creates a positron-electron pair that annihilates into a photon again.

particles can only be intermediate particles however, and as a whole the energy and momentum of the initial and final particles of a Feynman diagram do need to be conserved.

To calculate the cross section of a Feynman diagram, it is necessary to integrate over the entire possible phase space of the interaction. For this, every possible 4-momentum distribution in each vertex is considered. When a propagator boson is created off-shell then the term in the integration is modified by a factor $\frac{1}{p^2-m^2}$ where m is the on-shell mass of the propagator and p is the actual 4-momentum of the propagator. One can see that the off-shell contribution quickly falls off for large discrepancies between the on-shell and off-shell mass of the particle.

In addition to the phase space, the matrix elements $\mathcal{M}_i$ of all possible Feynman diagrams that have the same initial and final states have to be calculated. Fig. 2.3 shows another valid Feynman diagram that has the same initial and final states as the diagram in fig. 2.2. In a detector they cannot be told apart, but the processes that generate the final state particles are different. The one on the right is called a s-channel diagram and the one on the right is called a t-channel diagram. Each matrix element $\mathcal{M}_i$ first has to be calculated for each diagram separately before being added up. The square of this quantity is proportional to the cross section:

$$\sigma \propto |\mathcal{M}|^2 \Phi \tag{2.1}$$

where Φ is the phase space of the interaction.

Theoretically there are an infinite number of Feynman diagrams that can contribute because loops can always be added to the diagram. However, each vertex has a coupling strength α attached to it, which is usually $\alpha << 1$. Therefore, for each additional vertex in the Feynman diagram (whereby each loop adds at least two additional vertexes), the cross section contribution drops quickly with α^2. The coupling strengths of the fundamental interactions are also given in tab. 2.1. The simplest Feynman diagrams of a reaction are called the leading-order (LO) terms, the ones with an added loop are called next-to-leading-order (NLO) terms, ones with yet another loop are called next-to-next-to-leading-order (NNLO) terms, etc. One example of an NLO diagram for the $e^+e^- \rightarrow e^+e^-$ s-channel diagram is shown in fig. 2.4. Here the propagator photon temporarily creates an electron positron pair out of the vacuum, which then annihilates back into a photon.

One final effect discussed in this section is initial state and final state radiation. It is possible that an additional particle gets created at the beginning or end of a reaction. These don't loop back into the diagram but radiate away. They are usually massless gluons or photons. An example of initial state radiation is shown in fig. 2.5. Final and initial state radiation play an important role in detector experiments and their effects are further discussed in [27].

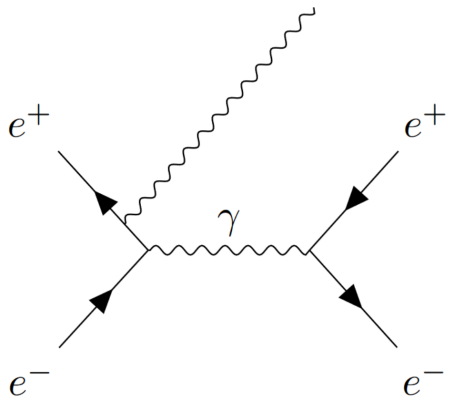


Figure 2.5: A Bhabha scattering Feynman diagram with initial state radiation. A photon is radiated by the positron, which lowers the center of mass energy of the interaction.

2.1.2 Electromagnetism

Electromagnetism (EM) is described by quantum electrodynamics (QED). The interaction is propagated by the photon, a massless spin 1 boson that couples to the electric charge, i.e. to all elementary fermions other than the neutrinos. The force is repulsive for same signed charges and attractive for opposite charges. The photon itself has no charge and therefore doesn't self-couple. The Lagrangian of QED [79] is given by

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\gamma^{\mu}D_{\mu} - m)\psi \tag{2.2}$$

where γ^{μ} are the dirac matrices, ψ is a spinor, the wave function of a fermion, and $\bar{\psi} = \psi^{\dagger}\gamma^{0}$ is the adjoint spinor. $D_{\mu} = \delta_{\mu} - iqA_{\mu}$ is the gauge invariant derivative, where A is the electromagnetic four-potential. $F_{\mu\nu} = \delta_{\mu}A_{\nu} - \delta_{\nu}A_{\mu}$ is the electromagnetic field tensor. Since the photon is massless, the electromagnetic force is a long range force in the Standard Model. The coupling strength of EM is

$$\alpha_{EM} = \frac{1}{4\pi\epsilon_0}\frac{e^2}{\hbar c} \approx \frac{1}{137}. \tag{2.3}$$

The strength of α_{EM} increases for higher energies due to the appearance of virtual particles. These appear more and more for higher energies through the uncertainty relation. This dependence on the energy scale of the interactions is called running coupling. Fig. 2.6 shows the coupling strength as a function of the energy scale, though we do not expect it to follow this pattern for very high energies and instead converge with all other coupling strengths into a unified force. At the mass of the Z boson, the coupling strength is $\alpha_{EM}(m_z = 90\,\text{GeV}) \approx 1/127$.

2.1.3 Weak Interaction

The weak interaction (WI) is the only force in the Standard Model that couples to all left-handed elementary fermions. The lagrangian of the WI [79] is given by

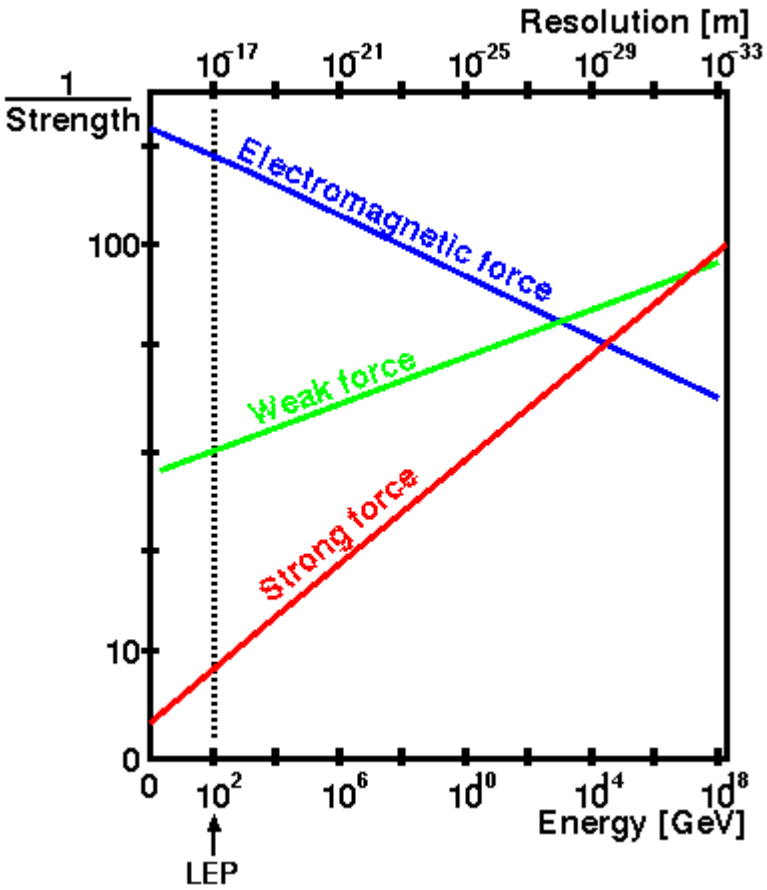


Figure 2.6: A sketch showing the dependence of the coupling constant of the three fundamental interactions described by the SM on the center of mass energy [10].

$$\mathcal{L}_{\text{weak}} = -\frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{\psi} i \gamma^\mu D_\mu \psi \tag{2.4}$$

where ψ describes the left-handed quark or lepton doublets or the right-handed quark or lepton singlets. For the weak interaction, the covariant derivative is defined as

$$D_\mu = \delta_m u + i \frac{g'}{2} Y B_\mu - i \frac{g}{2} T_j W^j_\mu \tag{2.5}$$

where Y is the weak hypercharge, g' is the coupling strength to the weak hypercharge, T_j are the 3-components of weak isospin, and g is the coupling strength to the weak isospin. These are elaborated on later in this section. $W^a_{\mu\nu}$ and $B_{\mu\nu}$ describe the field strength tensors of the weak interaction [69].

The WI is propagated by the Z boson and by the $W^{+/-}$ bosons. The bosons of the WI are very massive ($m_Z = 91.188 \pm 0.002\,\text{GeV}$ and $m_W = 80.38 \pm 0.01\,\text{GeV}$) and have very short half-lives ($\tau_Z = 2.637 \cdot 10^{-25}$ s and $\tau_W = 3.16 \cdot 10^{-25}$ s). The short half-life means that the range of the WI is very short on the range of $0.01 - 0.1$ fm. The WI is much weaker than the other two forces described

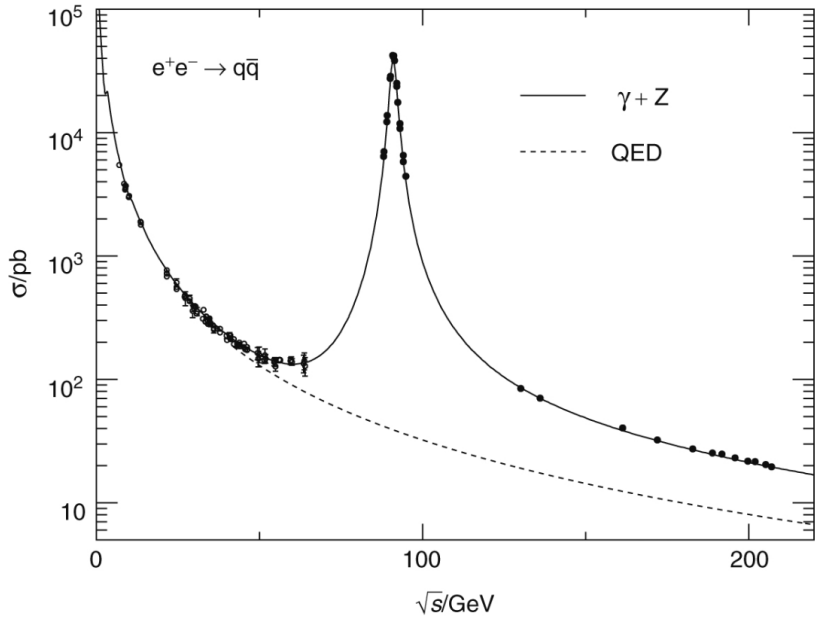


Figure 2.7: An electron-positron annihilation cross section measurement showing a peak in the cross section at the mass of the Z boson (approximately 90 GeV) [60]

by the SM, having coupling strengths on the order of 10^{-6} to 10^{-7}. This is because the propagator in the corresponding Feynman diagrams have the factor $\frac{1}{q^2-m^2}$ where q is the center of mass energy of the event and m is the mass of the propagator. For the very massive WI bosons, this factor becomes very small. At center of mass energies around the heavy boson masses, the cross section reaches significantly higher values. The cross section as a function of the center of mass energy is shown in fig. 2.7. Here two Feynman diagrams contribute to the cross section as the reactions $e^+e^- \to Z \to \mu^+\mu^-$ and $e^+e^- \to \gamma \to \mu^+\mu^-$ mix. For low energies the process with the photon dominates, but for masses around 90 GeV the $e^+e^- \to Z \to \mu^+\mu^-$ process dominates and creates a strong peak.

One special property of the weak force is that it breaks parity symmetry by coupling to left-handed fermions stronger than to their right-handed partners. The WI equivalent of the electric charge that the photon couples to are the weak isospin T_3 and the weak charge Y. Just as with electric charge, the weak isospin and the weak charge are additive quantum numbers that must be conserved in all reactions. A table showing the weak isospin and hypercharge of all elementary fermions is shown in tab. 2.2. The W bosons couple to the weak isospin. It has a magnitude of $T_3 = 1/2$ for all left-handed fermions and a magnitude of $T_3 = 0$ for all right-handed fermions (and the opposite for the antiparticles). This translates to a maximum parity violation in reactions with W bosons. The Z boson coupling to the left- and right-handed chirality states are given by

$$c_L = T_3 - Q \sin^2 \theta_W \tag{2.6}$$

$$c_R = -Q \sin^2 \theta_W \tag{2.7}$$

This also creates a parity violation for the Z boson, though it is not maximal for particles that have electric charge. As neutrinos do not have electric charge, right-handed neutrinos and left-handed

fermion	Q	T_3	Y_L	Y_R	c_L	c_R	c_V	c_A
ν_e,ν_μ,ν_τ	0	$+\frac{1}{2}$	-1	0	$+\frac{1}{2}$	0	$+\frac{1}{2}$	$+\frac{1}{2}$
e,μ,τ	-1	$-\frac{1}{2}$	-1	-2	-0.27	$+0.23$	-0.04	$-\frac{1}{2}$
u,c,t	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$	$+\frac{4}{3}$	$+0.35$	-0.15	$+0.19$	$+\frac{1}{2}$
d,s,b	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$	$-\frac{2}{3}$	-0.42	$+0.08$	-0.35	$-\frac{1}{2}$

Table 2.2: Hypercharge, isospin and couplings of fermions in the Standard Model, with $\theta_W \approx 29^\circ$ [74]

antineutrinos do not interact weakly, effectively making them sterile.

Electroweak Force

For high energies, electromagnetism and the weak force can be combined into the electroweak force. In the electroweak force there are initially four massless bosons, the W^1, W^2, W^3 and B boson. The three W bosons couple to the weak isospin and the B boson couples to the weak hypercharge. In the SM, these four bosons mix and gain mass by spontaneous symmetry breaking through the Higgs mechanism. The electric charge arises through a mixing of the weak hypercharge Y and the weak isospin T_3:

$$Q = T_3 + \frac{1}{2}Y \tag{2.8}$$

This quantity does not couple to the Higgs field, which is why the photon is massless. Any other combination of these quantities must therefore couple to the Higgs field, which is why the realized Z and W bosons are massive. The four realized bosons of the electroweak force are created by mixing:

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B^0 \\ W^0 \end{pmatrix} \tag{2.9}$$

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2), \tag{2.10}$$

where the mixing angle of the electroweak force θ_W is called the Weinberg angle and is experimentally measured to be

$$\sin^2\theta_W = 1 - \left(\frac{m_W}{m_Z}\right)^2 = 0.22290(30). \tag{2.11}$$

The mixing angle also determines the coupling strenghts (e,g,g') of the bosons. Fig. 2.8 shows this relation.

2.1.4 Strong Interaction

The strong interaction is responsible for atoms being held together as well as for nucleus stability. It is therefore, just like the electromagnetic interaction, a necessary feature of our world to exist as it does. The strong interaction also describes the mechanism behind the alpha decay and nuclear power. It is called strong force since it has the strongest coupling strength of all the forces by multiple orders of

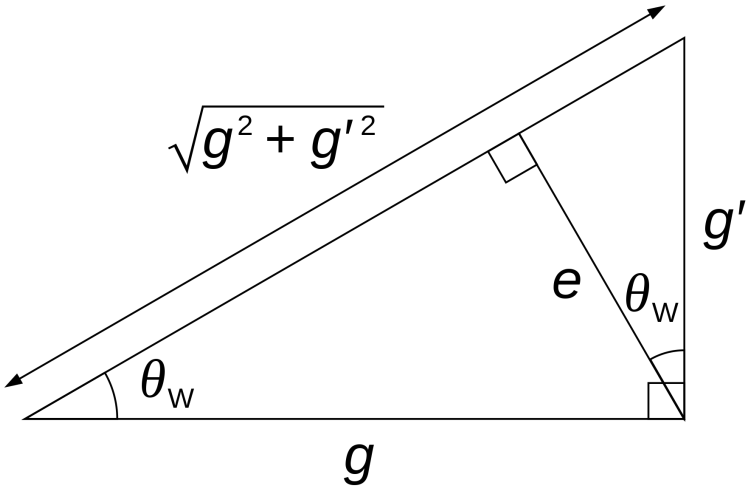


Figure 2.8: A diagram showing the relation between the weinberg angle and the coupling constants of the coupling constants of the electroweak interaction [72].

magnitude:

$$\alpha_S \approx 1 \tag{2.12}$$

The Strong interaction is described by quantum chromodynamics (QCD). The Lagrangian of QCD [79] is

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu a} \tag{2.13}$$

where ψ describes the spinor field of the quarks and $D_\mu = \delta_\mu - igT_aA^a_\mu$ is the gauge invariant derivative. $T_a = \lambda_a/2$ describes the generators of the $SU(3)$ group, which are explained in greater detail later in this section. A^a_μ ($a \in (1, ..., 8)$) describes the gluon fields, which are the strong force equivalent of the electromagnetic four-potential.

The propagator of the Strong Interaction is the massless gluon, which couples to all particles with color charge. The only particles that carry color charge are the quarks and the gluon itself. Since the gluon itself carries color charge it can couple to itself. Color charge is described using three colors, chosen here as red (r), green (g) and blue (b). Every quark carries one of these charges and every antiquark carries an anticolor charge ($\bar{r}$, $\bar{g}$ or $\bar{b}$). The gluon carries one color and one anticolor charge. While this opens up nine possibilities for the gluon, in actuality there are only 8 possible configurations of color charges for gluons since the strong interaction is described through the $SU(3)$ representation which gives an octet of states and a singlet state. The singlet would not be able to couple to any particle and can therefore not be real. The eight generators of the $SU(3)$ group can then be described through the Gell-Mann matrices λ_a. One representation of the Gell-Mann matrices in color space is the following orthogonal superposition of the colors:

$$\lambda_1 = r\bar{g} \qquad \lambda_2 = r\bar{b} \qquad \lambda_3 = g\bar{b} \tag{2.14}$$

$$\lambda_4 = g\bar{r} \qquad \lambda_5 = b\bar{r} \qquad \lambda_6 = b\bar{g} \tag{2.15}$$

$$\lambda_7 = \frac{r\bar{r} - \bar{g}g}{\sqrt{2}} \qquad \lambda_8 = \frac{r\bar{r} + \bar{g}g - \bar{b}b}{\sqrt{6}} \tag{2.16}$$

The color singlet would then be

$$\lambda_9 = \frac{b\bar{g} + \bar{b}g}{\sqrt{2}} \tag{2.17}$$

One effect of the gluon self-interaction is that isolated particles carrying color charge can never exist. Only color neutral composite particles are observed in nature. To be color neutral a particle has to either contain all three colors (or all three anticolors), or each color charge has to be cancelled out by its corresponding anticolor charge. When three quarks (or three antiquarks) combine with the appropriate colors, the first requirement is fulfilled. This composition is called a baryon (or antibaryon). Some examples of baryons are the proton (uud) and the neutron (udd). Since baryons are composed of three fermions, they are half integer spin particles (with $s = 1/2$ or $s = 3/2$, though exited states can have higher spin states). A quark and antiquark can combine into a meson, which fulfills the second requirement. Mesons are full integer spin particles (with either $s = 0$ or $s = 1$, though excited states can have higher spin states here too) [64]. The most well known mesons are the three pions (π^- ($\bar{u}d$), π^0 ($\bar{u}u/\bar{d}d$), or π^+ ($\bar{d}u$)). When an attempt is made to separate two connected quarks to create isolated color charges energy is transferred into the bond between the two quarks. For larger distances $r > 1$ fm the energy of the bond grows at a linear rate to the distance r. The potential between two quarks can be described by

$$V_{q\bar{q}(r)} = -\frac{4}{3}\frac{\alpha_S}{r} + kr \tag{2.18}$$

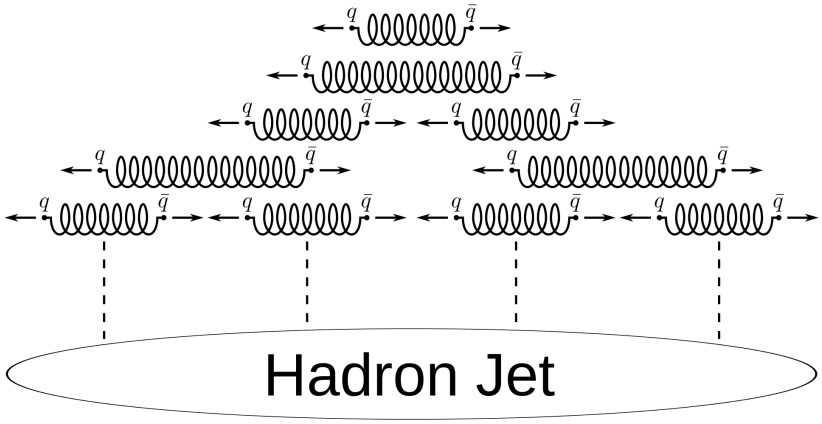


Figure 2.9: A diagram showing the hadronization process. At first there are two quarks separating (top), but as the potential energy in their bond increases, new quark pairs are formed [81].

Another consequence of of the gluon self-interaction is the behavior of coupling constant. Unlike

the EM interaction, the coupling constant of the strong interaction grows for smaller center of mass energies $\sqrt{s}$ and shrinks for higher energies. The dependence of the energy can be mathematically described through

$$\alpha_S(q^2) = \frac{1}{\beta \ln(q^2/\Lambda^2_{QCD})} \tag{2.19}$$

with β being defined as

$$\beta = \frac{\alpha^2}{\pi}\left(-\frac{11N}{6} + \frac{n_f}{3}\right). \tag{2.20}$$

where α is the fine structure constant as defined in eq. 2.3, n_f are the number of quark flavors that can contribute at the energy scale by being pair-produced (so when $q^2 > 2m_f$), and Λ_{QCD} is the so-called QCD scale [58, 21]. For low energies $q < 1.275\,\text{GeV}$, the QCD scale is measured to be $\Lambda = 332 \pm 17\,\text{MeV}$ [46]. It is clear that this value goes to infinity as q^2 approaches Λ^2. This asymptotic freedom makes QCD impossible to calculate perturbatively at low energies as higher-order Feynman diagrams become increasingly relevant. This is also visually depicted in fig. 2.9, as the depiction eventually switches to a bubble labeled "Hadron Jet", which symbolizes that we can not accurately describe what happens in the bubble.

2.2 Tau Lepton

The tau lepton is the heaviest of the six leptons and acts as a heavier partner of the other charged leptons with a mass of $m_\tau = 1776.9 \pm 0.1\,\text{MeV}$ (compared to $m_e = 0.511\,\text{MeV}$ for the electron and $m_\mu = 105.7\,\text{MeV}$ for the muon). It shares it's charge of $q = +1$ and spin of $s = 1/2$ with the other charged leptons. The much higher mass compared to the other leptons give it a much shorter lifetime of $\tau_\tau = 2.903 \pm 0.005 \cdot 10^{-15}\text{s}$ (compared to $\tau_\mu = 2.197\,\mu\text{s}$ for the muon, the electron is stable) and opens up many more decay channels [90]. The muon can only decay through $\mu \rightarrow W^- \nu_\mu \rightarrow e^- \bar{\nu_e} \nu_\mu$. The decay of muons and taus is always propagated through a W boson.

The Feynman diagram of a tau decay is shown in fig. 2.10. The tau first always decays into a tau neutrino and a W^- boson, the latter of which quickly decays into one of three pairs: a lepton pair ($\mu^- \bar{\nu}_\mu$ or $e^- \bar{\nu_\mu}$) or a quark pair ($d\bar{u}$). Similarly, antitau leptons decay into an antitau neutrino and a W^- boson, which then also quickly decays into a $\mu^+ \nu_\mu$ pair, a $e^+ \nu_\mu$ pair, or a $\bar{d}u$ pair. A decay into a $s\bar{u}$ pair or $\bar{s}u$ respectively is also possible through the flavor-breaking weak current described by the Cabibbo-Kobayashi-Maskawa matrix [84], but it is suppressed. Table 2.3 shows some of the possible decays of the tau lepton. While the Feynman diagram only shows three possible decays, the quark pair can decay into multiple particles depending on which specific meson resonance is created in the tau decay. This can be measured in a detector experiment through the amount of particle jets that get created through the tau decay. A nomenclature that gives the amount of charged jets as the number before the p and the number of neutral jets as the number before the n that a hadronic decay corresponds to, is also shown in the table. If a tau lepton decays into a neutrino and one or more hadrons, then it is called a hadronic tau (τ_{had}). If it decays into two neutrinos and a lighter lepton, it is called a leptonic tau (τ_{lep}).

Tau Decay Channel	Branching Ratio / %	XpXn notation
Leptonic	35.21%	NaN
$\tau^- \rightarrow \nu_\tau e^- \bar{\nu}_e$	17.82%	NaN
$\tau^- \rightarrow \nu_\tau \mu^- \bar{\nu}_\mu$	17.39%	NaN
Hadronic	64.79%	XpXn
$\tau^- \rightarrow \nu_\tau \pi^- \pi^0 \nu_\tau$	25.49%	1p1n
$\tau^- \rightarrow \nu_\tau \pi^- \nu_\tau$	10.82%	1p0n
$\tau^- \rightarrow \nu_\tau \pi^- \pi^0 \pi^0 \nu_\tau$	9.26%	1p2n
$\tau^- \rightarrow \nu_\tau \pi^- \pi^+ \pi^- \nu_\tau$	8.99%	3p0n
$\tau^- \rightarrow \nu_\tau \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$	2.74%	3p1n
$\tau^- \rightarrow \nu_\tau \pi^- \pi^0 \pi^0 \pi^0 \nu_\tau$	1.04%	1p3n

Table 2.3: Decay channels of the tau lepton. For hadronic decays the hadronic name classification is also given [1].

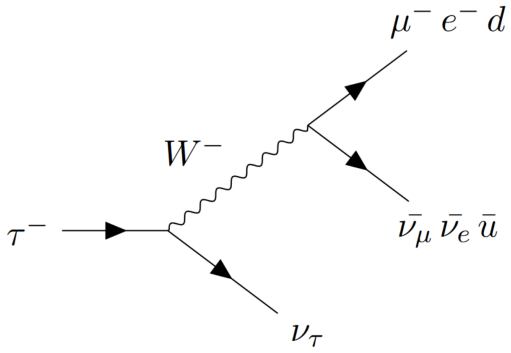


Figure 2.10: Feynman diagram of a τ-decay.

2.3 Higgs Boson

The Higgs boson was originally predicted to describe why the Z and W bosons have such high masses, since their massiveness otherwise breaks the gauge invariance of the Weak Force. This theory has expanded to now describe the masses of all massive elementary particles with the possible exception of the neutrinos.

2.3.1 The Higgs Mechanism

In the SM the Higgs mechanism describes how the coupling of elementary particles to the Higgs field gives those particles their rest mass through the Brout-Englert-Higgs mechanism [62]. The Higgs particle itself is an excitement of the Higgs field. In the $SU(2)_L \times U(1)_Y$ gauge symmetry of the electroweak theory, the potential of the Higgs field is

$$V(\Phi) = \mu^2(\Phi^*\Phi) + \lambda(\Phi^*\Phi)^2 \tag{2.21}$$

The breaking of the $SU(2)_L \times U(1)_Y$ gauge symmetry is achieved through a negative μ^2 in eq.

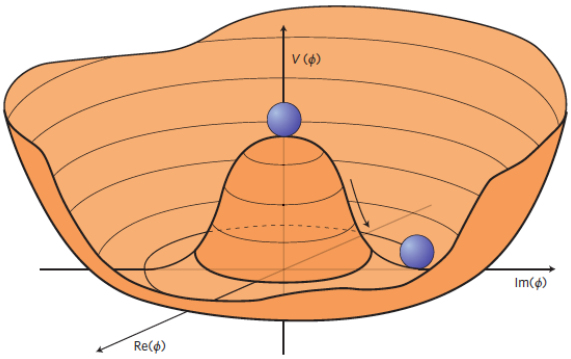


Figure 2.11: Characteristic "Mexican hat" potential of the Higgs boson [51].

2.21. This leads to the characteristic "Mexican hat" potential, which can be seen in fig. 2.11. Since the potential does not have its minimum at zero for negative μ^2, the vacuum expectation value is nonzero at 246 GeV [13]. The Lagrangian that describes the interaction between the Higgs field and the massive gauge bosons is

$$L_h = (D_\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi) \tag{2.22}$$

The mass of the leptons arises through a Yukawa coupling between the Higgs field and massless leptons, described though the a Yukawa Lagrangian:

$$L_y = -y_\psi \epsilon^{ij} h \bar{\Psi} \psi \tag{2.23}$$

where Ψ describes a left-handed fermion doublet and ψ describes a right-handed fermion singlet. y_ψ are the matrices of the Yukawa couplings [58].

2.3.2 Higgs Boson Properties

The Higgs boson is the only scalar elementary particle ($s = 0$) in the SM and it is the second heaviest elementary particle with a mass of $m_H = 125.10 \pm 0.14$ GeV and a half width constrained through measurements to $\Gamma_H < 0.013$ GeV (corresponding to a life time constraint of $\tau_H < 1.6 \cdot 10^{-22}$s). The half width is predicted to be around $\Gamma_H = 4$ MeV but the reconstruction is not accurate enough to measure it. The coupling strength to a massive particle is proportional to it's mass. This means that a Higgs boson is more likely to be produced in collisions of heavy particles and that it is more likely to decay into more heavy particles. The theoretical branching ratios as a function of the Higgs mass are shown in fig. 2.14. For masses $m_H < 120$ GeV the $H \rightarrow b\bar{b}$ branching ratio dominates because it is the channel where the decay products have the highest mass without being higher than the mass of the H boson itself. As the mass approaches $m_H = 2m_W \approx 160$ GeV the $H \rightarrow WW$ channel starts to dominate. As the mass rises further the Z channel opens and therefore dominates. The predicted branching ratios of the Higgs boson for a mass of $m_H = 125.2$ GeV and the ratios of the predicted to the measured branching ratio are listed in tab. 2.4. The channel that is of interest in this thesis is the $H \rightarrow \tau\tau$ channel. For the actual Higgs mass it has a branching ratio of about 6.24%. This decay channel can be split into three different categories: $H \rightarrow \tau_{had}\tau_{had}$, $H \rightarrow \tau_{had}\tau_{lep}$ and $H \rightarrow \tau_{lep}\tau_{lep}$, corresponding to the possible decays of the resulting tau leptons. In this thesis only the $H \rightarrow \tau_{had}\tau_{had}$

channel is considered.

Decay channel	BR_{theo} / %	BR_{theo}/BR_{exp}
$H \rightarrow b\bar{b}$	57.9	0.98 ± 0.12
$H \rightarrow \tau\bar{\tau}$	6.2	$1.15^{+0.16}_{-0.15}$
$H \rightarrow \mu\bar{\mu}$	0.02	1.19 ± 0.34
$H \rightarrow c\bar{c}$	2.9	$37 \pm 17^{+11}_{-9}$
$H \rightarrow gg*$	8.2	NaN
$H \rightarrow \gamma\gamma*$	0.2	1.10 ± 0.07
$H \rightarrow Z\gamma$	0.2	< 3.6
$H \rightarrow WW*$	21.7	1.19 ± 0.12
$H \rightarrow ZZ*$	2.7	1.01 ± 0.07

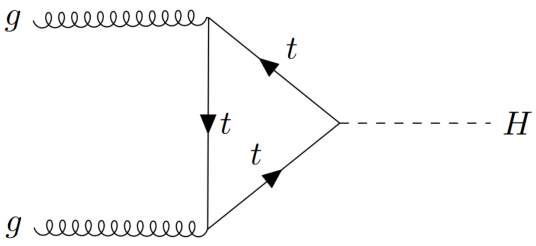
Table 2.4: Theoretical and measured branching rations for the Standard Model Higgs boson.

There are multiple production channels for the Higgs boson in proton antiproton collision. Their Feynman diagrams are given in fig. 2.12. In the gluon fusion production channel (see fig. 2.12(a)) two gluons, one from each proton, fuse through intermediate particles. The intermediate particles are usually top quarks since they are the most massive and therefore have the strongest coupling to the Higgs boson [29]. In the vector boson fusion (see fig. 2.12(b)), a quark from each proton radiates a heavy vector boson (W or Z bosons). They then fuse into a Higgs boson [35]. Vector boson radiation (see fig. 2.12(c)) is characterized by two quarks fusing into a W boson, which then radiates a Higgs boson through final state radiation [59]. The final of the four processes is tt-associated production (see fig. 2.12(d)). Here, two gluons from the protons decay into tt-pairs and two of the top quarks fuse into a Higgs boson [8]. Each of those four production processes has different kinematics and can be detected through different signatures in the detector experiments. The theoretical cross sections of the Standard Model Higgs boson production as a function of the center of mass energy $\sqrt{s}$ for the aforementioned channels are shown in fig. 2.13. gg-Fusion (labeled in the plot as $pp \rightarrow$ H) is by far the most significant contributor to the entire Higgs production cross section [30].

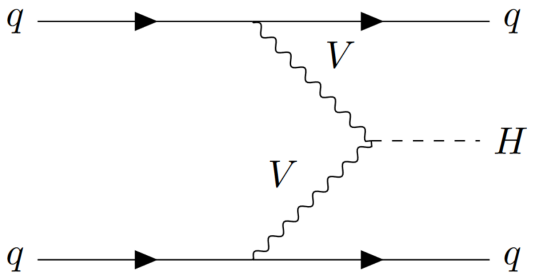
2.4 Z Boson

The Z boson was discovered in 1983 at the UA1 and UA2 experiments at the super proton synchrotron at CERN [48]. It is one of the bosons responsible for propagating the weak force and it has a mass of 91.188 ± 0.002 GeV and a half width of 2.495 ± 0.002 GeV. The Z boson is a vector boson ($s = 1$) with no charge. The decay channels are listed in tab. 2.5. The $Z \rightarrow \tau\tau$ decay has a branching fraction of 3.370 ± 0.008 % [12]. This decay channel can also be split into three categories corresponding to the different decays of the tau leptons: $Z \rightarrow \tau_{had}\tau_{had}$, $Z \rightarrow \tau_{had}\tau_{lep}$ and $Z \rightarrow \tau_{lep}\tau_{lep}$.

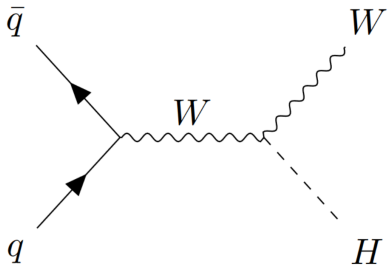
The Higgs and the Z boson share many decay channels and they have similar quantum numbers that need to be conserved. The differences between the two bosons are their masses, their spins and the different branching ratios. However due to the large inaccuracies on the ditau mass reconstruction for both the $Z \rightarrow \tau\tau$ and $H \rightarrow \tau\tau$ decays it is not possible to definitively say if a specific reconstructed ditau event is from a Higgs or a Z boson. This is one of the reasons why it is important to have a good ditau mass reconstruction. In addition, the differing spins of the two bosons lead to different



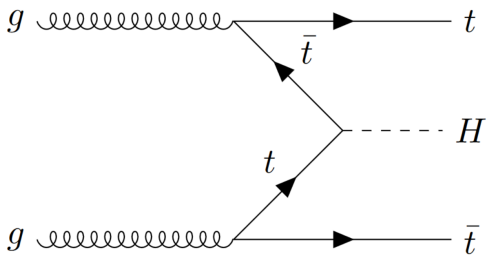
(a) Feynman diagram of gluon fusion. Two gluons fuse through a top quark into a Higgs boson.



(b) Feynman diagram of vector boson fusion. Two quarks radiate a vector boson each (Z or W boson), which fuse into a Higgs boson.



(c) A Feynman diagram showing vector boson radiation. Here two quarks fuse into a W boson, which radiates a Higgs boson.



(d) Feynman diagram of tt-associated production. Two gluons each decay into a tt-pair and two of the top quarks fuse into a Higgs boson.

Figure 2.12: Feynman diagram of the four main Higgs production channels in proton-proton collider experiments.

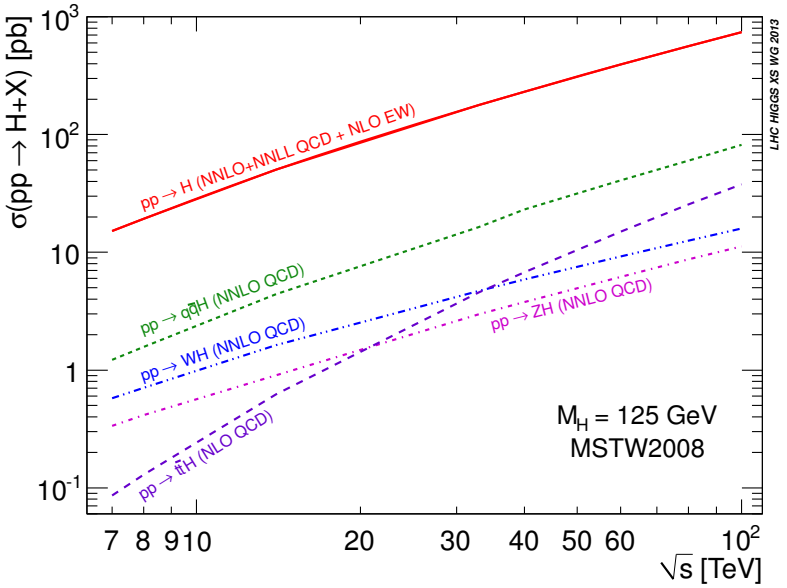


Figure 2.13: The production cross section of the Higgs boson in pp collisions as a function of the center of mass energy $\sqrt{s}$. The cross section of the separate production channels is also shown separately [30].

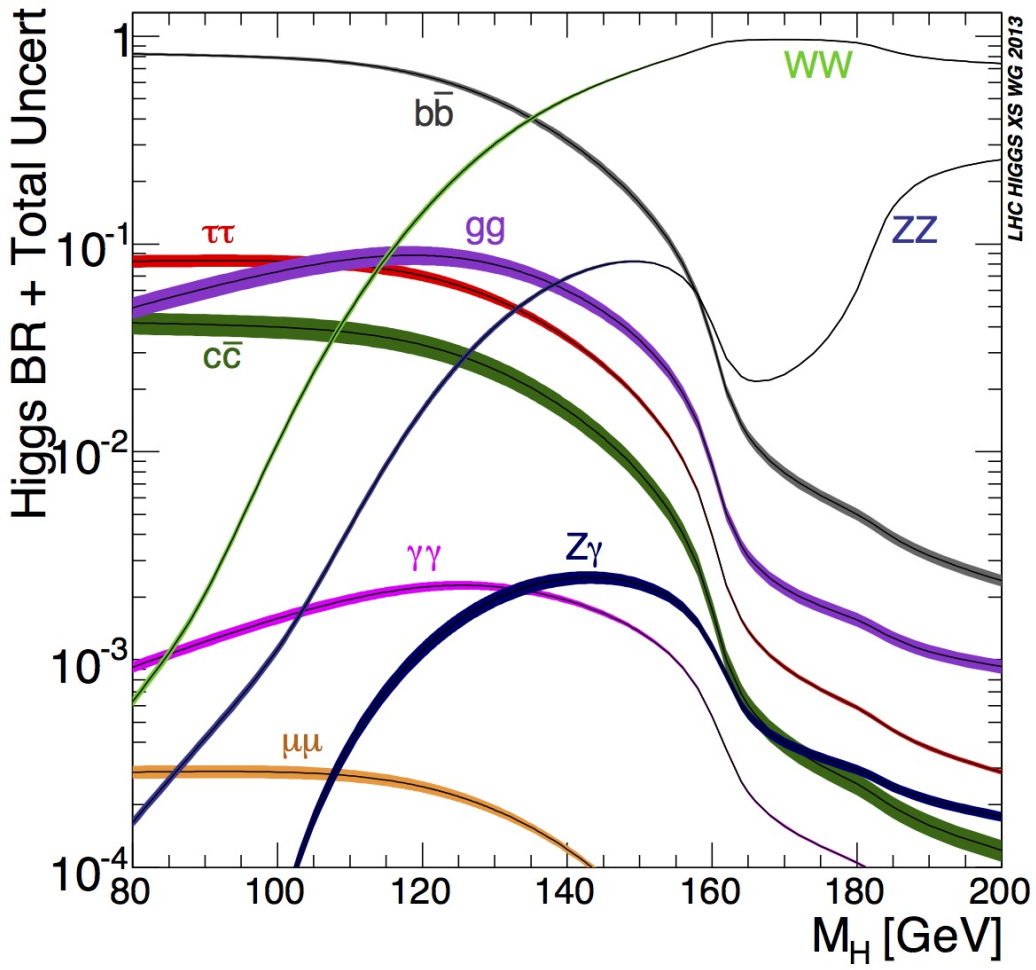


Figure 2.14: The theoretically predicted branching ratio of the Higgs boson as a function of the Higgs mass [30].

Decay channel	BR_{theo} / %
$Z \rightarrow e^+e^-$	3.363 ± 0.004
$Z \rightarrow \mu^+\mu^-$	3.366 ± 0.007
$Z \rightarrow \tau^+\tau^-$	3.370 ± 0.008
$Z \rightarrow$ invisible	20.00 ± 0.06
$Z \rightarrow (d\bar{d} + s\bar{s} + b\bar{b})/3$	15.6 ± 0.4
$Z \rightarrow (u\bar{u} + c\bar{c})/2$	11.6 ± 0.6
$Z \rightarrow b\bar{b}$	15.12 ± 0.05
$Z \rightarrow c\bar{c}$	12.0 ± 0.2

Table 2.5: The decay channels and their branching ratios of the Z boson.

kinematic distributions of the tau decay products. Since tau leptons are spin-1/2 particles, there are only a limited amount of possible helicity configurations for the taus that are created in Higgs or Z

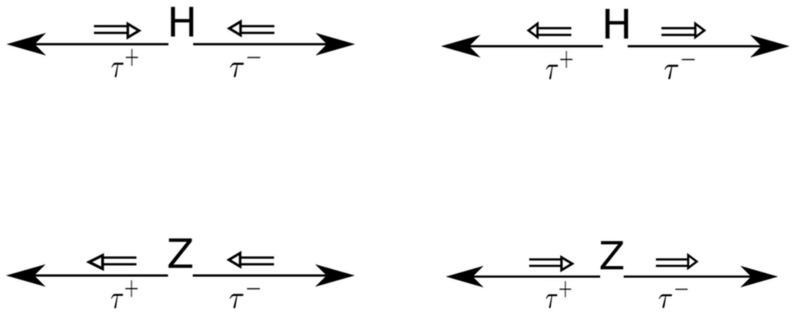


Figure 2.15: The possible tau spin-configurations for Higgs and Z decays [85].

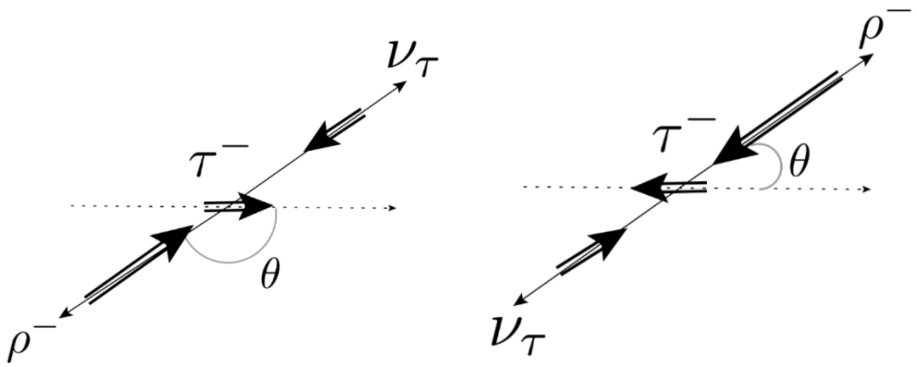


Figure 2.16: Possible spin configurations for 1p1n tau decays [85].

boson decays. They can be seen in fig. 2.15. For the Higgs boson both tau leptons have the same helicities, while for the Z boson the taus have opposite helicities. This is not a perfect relation because it is only true in the rest frame of the boson but gluon jets can push the boson into a boosted frame.

2.5 Charged Energy Asymmetry ϒ

One way in which the different spins of the Higgs and Z boson can manifest themselves in the kinematics of ditau events is through the charged energy asymmetry ϒ of a tau decay. It is defined as

$$\Upsilon = \frac{E_{charged} - E_{neutral}}{E_{charged} + E_{neutral}} \tag{2.24}$$

where $E_{charged}$ is the energy from charged pions and $E_{neutral}$ is the energy from neutral pions in tau decays. The following is a quantitative derivation of the distribution of this asymmetry for the 1p1n decay.

In this decay there is a mediating spin-1 ρ meson resonance: $\tau^- \rightarrow \rho\nu_\tau \rightarrow \pi^+\pi^0\nu_\tau$. Due to spin conservation and the parity violation of the weak interaction only some spin configurations for the decay products are possible. They are shown in fig. 2.16 in the rest system of the tau. The thick arrows indicate the spin and the thin arrows the momentum direction of the individual particles. When the spin and momentum of a particle align then it has positive helicity and is said to be right-handed.

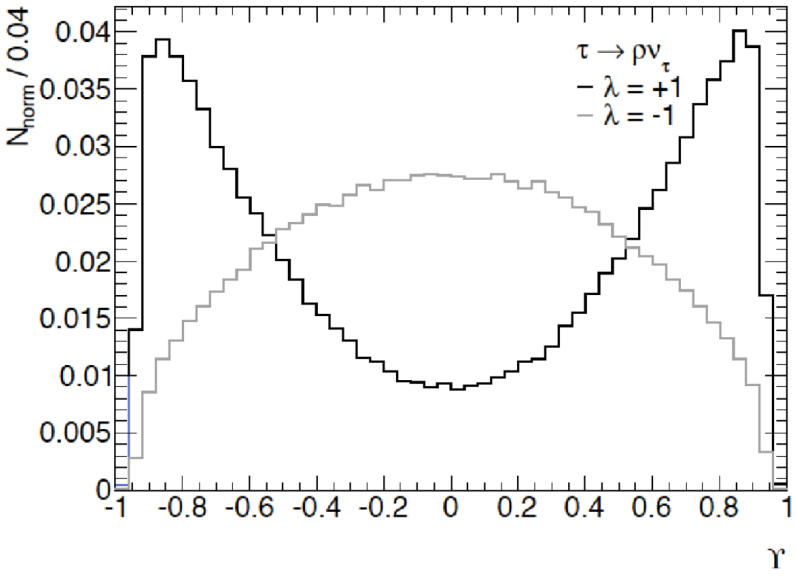


Figure 2.17: The expected Υ distributions of a left-handed and right-handed tau lepton. The expected distributions of an anti tau lepton the exact opposite [45].

If they oppose each other it has negative helicity and is said to be left-handed, where helicity is mathematically described as

$$h = \frac{\vec{s} \cdot \vec{p}}{|\vec{p}|}. \tag{2.25}$$

In the case of a boosted tau this means that the ρ meson, which carries all of the visible energy of the tau decay, carries a larger fraction of the energy for right-handed tau decays and a smaller fraction for left-handed tau decays. Another effect of the kinematics of the tau decay is that the ρ meson is more likely to be longitudinally polarized for right-handed taus and more likely to be transversally polarized for left-handed taus. When the ρ decays into two pions, the spin conservation again constrains the phase-space of the possible decay configurations. Since pions are spin-0 mesons, transversally polarized ρ mesons are more likely to give both pions a similar amount of energy, while a longitudinally polarized ρ is more likely to have one of the pions shoot off into the boost direction of ρ and therefore carry more kinetic energy. This creates a charged enery asymmetry with helicity dependence. The theoretically expected distribution for Υ can be seen in fig. 2.17. This quantity can be used to verify the spin information of Monte-Carlo datasets and can also be used as a spin sensitive quantity in event reconstruction techniques. For anti tau leptons the same argumentation can be used for opposite spins [85].

CHAPTER 3

The Experiment

This thesis is based on Monte Carlo (MC) simulated events at the ATLAS detector [7] at the Large Hadron Collider (LHC) [32] in CERN. To understand the data and the problems with reconstructing $H \rightarrow \tau\tau$ and $Z \rightarrow \tau\tau$ events better, this chapter explains how the LHC and the ATLAS detector work. First, the actual collider complex and how the particles are accelerated to their high energies is described. After that the ATLAS detector and its subdetectors are described.

3.1 the Large Hadron Collider

The LHC is the largest and most powerful man-made particle accelerator in the world. It is part of the larger LHC complex which is located at CERN in Geneva. The largest part of the complex lies in France and the rest is in Switzerland. Some parts of the LHC are over 110 meters underground and it has a circumference of 27 km [49]. It is used to accelerate and then collide protons or sometimes lead ions. Before being used for experiments in the LHC, the particles need to be accelerated in multiple steps in circular and linear accelerators (see fig. 3.1). First, the protons (or lead) are accelerated in a linear accelerator, from which they are led into three different synchrotrons (The Proton Synchrotron Booster, the Proton Synchrotron and the Super Proton Synchrotron) where they are accelerated further. After those steps they finally have enough energy to enter the LHC. There they reach energies of up to 6.5 TeV each for a total center of mass energy of 13 TeV (or 2.3 TeV and 4.6 TeV respectively for lead nuclei) [33].

At the LHC there are four main detector experiments which all specialize in different aspects of event detection and complement each other. The ALICE (A Large Ion Collider Experiment) detector (seen on the left in fig. 3.1) aims to explore the quark-gluon plasma resulting from Pb-Pb nuclei collisions. The CMS (Compact Muon Solenoid) has a very sophisticated muon detection system and aims to explore TeV-scale physics and investigate the properties of the Higgs boson. The LHCb (Large Hadron Collider beauty) experiment focuses on b-physics involving hadrons containing bottom quarks. The last of the four main detectors is the ATLAS detector, which will be more thoroughly described in the next section. In addition to these four there are also other experiments collecting data at the LHC complex. They are listed in 3.1.

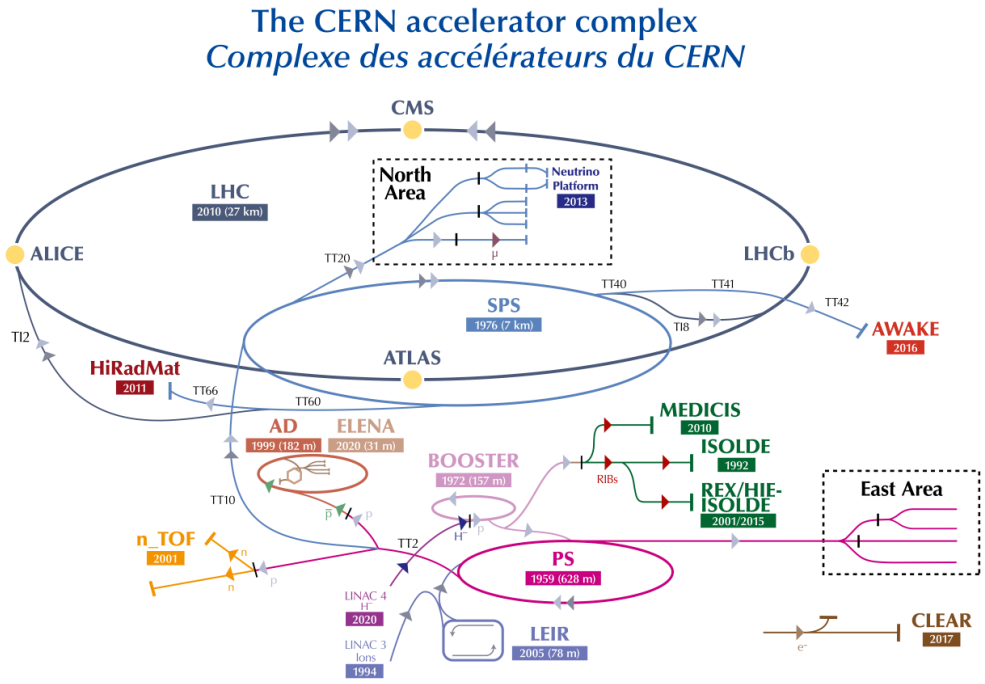


Figure 3.1: A sketch of the CERN complex, showing the different accelerator systems and detectors within the complex [28].

Experiment	Full Name	Type of Detector	Mass / Tonnes
ALICE	A Large Ion Collider Experiment	lead ion collisions	10000
ATLAS	A Toroidal LHC ApparatuS	pp collisions	7000
CMS	Compact Muon Solenoid	pp collisions	14000
LHCb	Large Hadron Collider beauty	pp collisions	5600
TOTEM	Total, elastic and diffractive cross section measurement	pp collisions	3
LHCf	Large Hadron Collider forward	pp collisions	0.08
MoEDAL	Monopole and Exotics Detector at the LHC	pp collisions	1 [80]
FASER	Forward Search Experiment	pp collisions	1.1
SND@LHC	Scattering and Neutrino Detector at the LHC	pp collisions	0.83

Table 3.1: The experiments at the LHC [33].

3.2 The ATLAS detector

The ATLAS (A Toroidal LHC Apparatus) detector is the largest detector at the LHC. It is 46 meters long, has a 25-meter diameter and weighs 7 000 tonnes. As with the other detectors, it is comprised of multiple sub-detectors that work together to track the particles that are created when the beams collide

at the collision point in the center of the detecter, as well as identify their particle type and measure their energy and momentum. Fig. 3.2 shows a cross section view of how different particles interact with the detector subcomponents. First, particles pass through the inner detector where charged particles leave bent tracks. After the inner detector they pass to the electromagnetic and then the hadron calorimeters. There most particles create particle showers and deposit all their energy for measurement. Only muons and neutrinos remain. After the calorimeters comes the muon spectrometer, which is used to track the momentum of muons. Both the muons and neutrinos leave the detector. Neutrinos can not be measured in any way in the detector. Two superconducting magnet systems are used to bend charged particles for charge and momentum measurements in multiple of the sub-detectors, one solenoid magnet for the inner detector and a toroid magnet for the muon spectrometer system [7, 3].

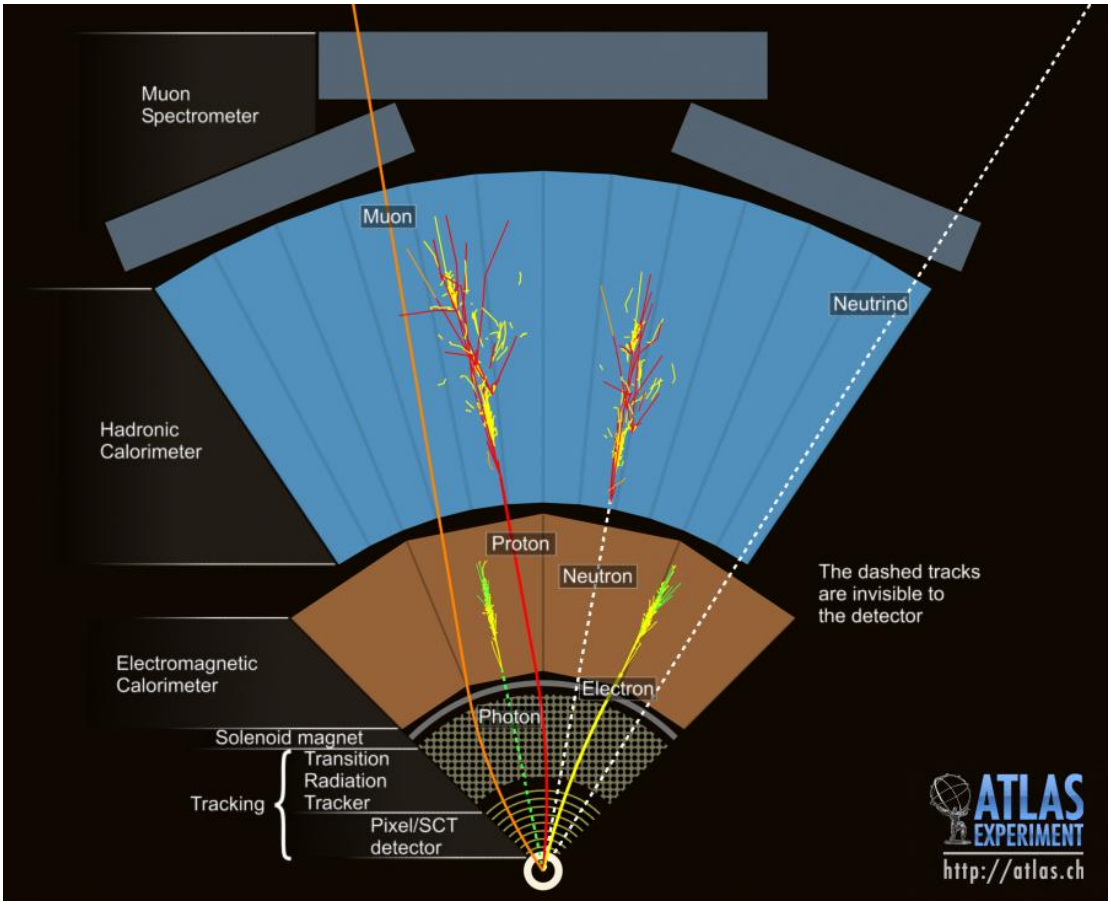


Figure 3.2: A cross section view of how different particles interact with the subdetectors of the ATLAS detector [78].

3.2.1 ATLAS coordinate system

The z-axis of the ATLAS coordinate system lies along the beam line, crossing through the center of the cylindrical detector. The x-axis points from the interaction point towards the center of the LHC and the y-axis points upwards. The plane formed by the x- and y-axis is called the transverse plane.

The azimutal angle ϕ is measured around the beam axis and the polar angle θ is the angle from the beam axis, although the so-called pseudorapidity η is generally used instead of the polar angle. It is defined as

$$\eta = -\ln\tan(\frac{\theta}{2}) \tag{3.26}$$

and it is the high energy or low mass approximation for the rapidity:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right), \tag{3.27}$$

where p_z is the longitudinal momentum in z-direction. This high energy approximation $\eta \approx y$ can be used with the high beam energies of the LHC. The advantage of using the pseudorapidity η over the polar angle θ is that the particle generation cross section is constant as a function of y since differences in rapidity Δy are invariant under lorentz boosts along the z-axis.

Some variables such as the transverse energy are defined as the component of the vector that lies on the transverse plane: $E_T^2 = E_x^2 + E_y^2$. Angular distances are measured in relation to the azimutal angle and the pseudorapidity: $\Delta R^2 = \Delta\eta^2 + \Delta\phi^2$ [7].

When two protons collide in the center of the ATLAS detector, the high energies lead to deep inelastic collisions. This means that the partons inside the protons interact directly with each other. Since a parton only carries a fraction x of the entire proton energy, the actual center-of-mass energy $\sqrt{s}$ of the event as well as the rest frame of the event are unknown. The only certain thing is that the transverse momentum of the interaction is zero. This is why the actual measured energies themselves are not relevant and the transverse quantities are often used instead [7, 71].

3.2.2 Inner detector

The inner detector of ATLAS is the closest sub-detector to the beam-line. It completely surrounds it from a radius of 3.3 cm to 110.6 cm and has a length of 6.2m. The inner detector itself is made up of three different parts. Their arrangement can be seen in fig. 3.3. The particles first pass through the Pixel Detector (PD). It surrounds the beam line from 3.3 cm to 19 cm and has a length of 1.85 m. The PD consists of 4 barrel layers radially surrounding the collision point and 3 disks at both ends of the detector along the beam axis. The layers consist of a total of 92 million silicon pixels, each of which has an electronic channel that can be read out. The size of the pixels is $50 \times 499\,\mu m^2$ for the outermost layer and $50 \times 250\,\mu m^2$ for the innermost layer. In total the silicon covers an area of $1.9\,m^2$ and consumes 15 kW of power. As the particles pass through the silicon they are bent by a magnetic field, which ionizes the silicon, thus allowing for momentum reconstruction using the bending radius [26]:

$$\frac{1}{\rho} = \frac{eB}{p} \Leftrightarrow \frac{1}{\rho}[m^{-1}] = 0.3\frac{B[T]}{p[GeV/c]}. \tag{3.28}$$

The PD is also used to reconstruct the primary and possible secondary vertices of the event.

After the Pixel Detector particles enter the semiconductor tracker (SCT). It consists of 4 088 modules and a total of over 6 million micro-strip silicon sensors that are distributed over 4 cylindrical barrel layers and 18 planar end-cap discs. The semiconductor tracker has a total area of $60\,m^2$. Each particle passes through at least four silicon layers of the semiconductor tracker. The SCT has readout strips every 80 µm. The system allows for a reconstruction precision of up to 25 µs [44].

The last part of the inner detector is the transition radiation tracker (TRT). The TRT is made up of 300 000 4-mm gas filled straws that have a 30 µm gold-plated tungston wire in their center. 50 000 of the straws are in the barrel and are 144 cm long and the other 250 000 straws are 39 cm long are in the end-caps of ATLAS. Charged particles passing through these straws ionize the gas, generating a detectable signal on the wires in their center. The signal is read out through 350 000 read-out channels. The total volume of the TRT is 12 m^3. In addition to being able to reconstruct the flight path of the particle, the TRT also gives information on the type of particle based on the transition radiation that is generated when particles pass through fibers interleaved between the straws in the barrel region and foil between the layers in the end-cap region [5, 42].

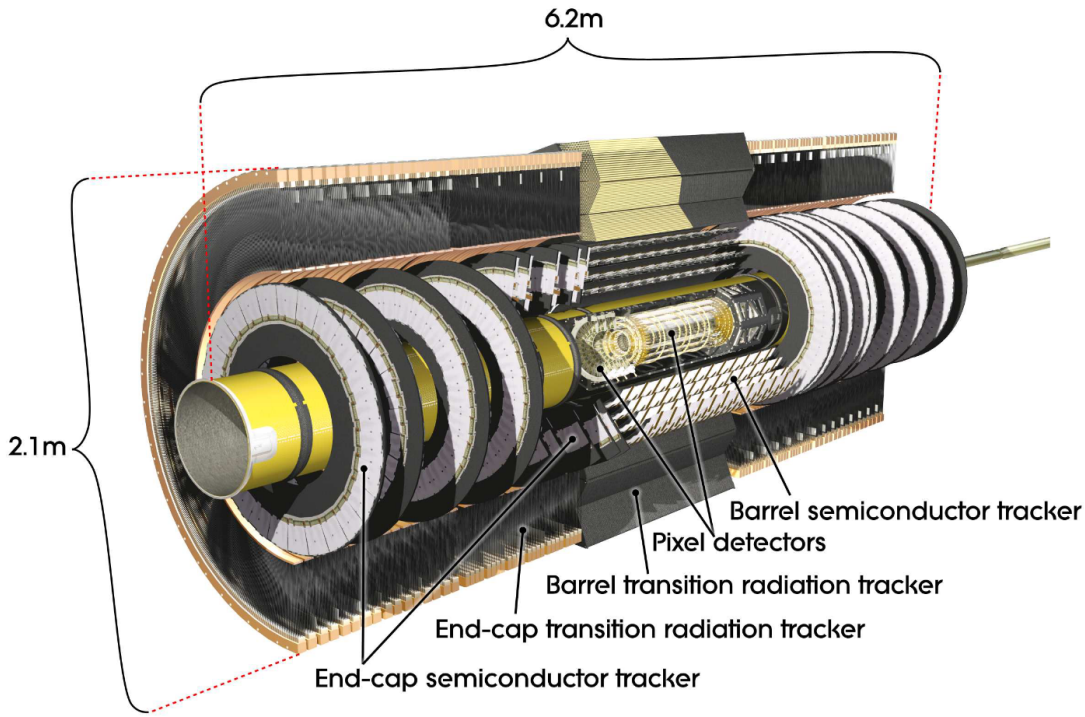


Figure 3.3: A schematic view of the inner detector of ATLAS [4].

3.2.3 Calorimeters

After passing through the inner detector, particles enter the calorimeters of ATLAS. The calorimeters are designed to make all particles, except for neutrinos and muons, deposit their entire energy in them. This makes it possible to reconstruct the particle energy and also gives us insight into the particle type. The calorimeter system of ATLAS consists of a liquid argon (LAr) electromagnetic calorimeter and a tile hadron calorimeter [39].

The LAr calorimeter is 6.4 m long and 53 cm thick. The end-caps of the LAr electromagnetic calorimater are 2.08 m long and 63 cm thick [23]. It is designed to make electrons and photons create electromagnetic showers and deposit their entire energy. Hadrons also interact with the LAr calorimeter but generally do not deposit their entire energy in it. The calorimeter has layers of lead

and liquid argon arranged in an accordion shape to maximize the amount of layer crossings that particles go through. In the metal layers particles interact with the lead atoms, creating photons and electrons, which then deposit some of their energy in the LAr gas through ionization. This creates a measurable electric current that is read out through 110 000 channels and makes it possible to measure the energyof the particle. To keep the argon in a liquid state, it is kept at -184° [6].

The tile hadron calorimeter is designed to make hadrons deposit their entire energy in them and it consists of layers of steel and 420 000 scintillating plastic tiles. It is composed of 192 wedges, each of which is 2.6 m − 5.6 m long. In the steel layers hadrons produce hadron showers which then create photons as they pass through the scintillating tiles. The photons get converted into a measurable electric current through 9 500 photomultiplier tubes. The hadron calorimeter itself weights 2 900 tonnes [54, 15].

3.2.4 Muon Spectrometer

The only particles that can pass through all sections of the calorimeter without decaying or depositing their entire energy or decaying are muons and neutrinos. While neutrinos only interact weakly and therefore can escape the detector without being measured in any way, muons leave tracks in all the detector sub-components. Due to their weight, they do not deposit their entire energy in the calorimeters. In order to get better momentum measurements of muons the muon spectrometer has a stronger magnetic field covering a much larger area than the inner detector. Since the only particles that leave tracks in the muon spectrometer are muons, it also leads to a higher muon identification efficiency. The spectrometer consists of two parts, the precision detectors and the fast-response detectors. The precision detectors consist of monitored drift tubes (MTDs), which are composed of 3 cm wide aluminum gas filled tubes with a thin wire in the center. These create electric signals when muons pass through them and are able measure the position of muons with an accuracy of 100 µs.

The fast-response detectors consist of two parts, the resistive plate chambers (RPCs) and the thing gap chambers (TGCs). The RPCs surround the central region of the ATLAS detector. They are pairs of parallel plastic plates that are separated by a gas volume and have a potential difference between them. When muons pass through the gas they generate an electric signal. The TGCs lay at the ends of the ATLAS detector and consist of 30 µm wires in a gas mixture, which also generate electrical signals through ionization when muons travel through them. The fast-response detectors allow for rapid event selection and can discard events that aren't relevant in 2.5 µs [41].

3.2.5 Magnet System

ATLAS uses two types of superconducting magnets. The inner detector is surrounded by a solenoid magnet [94]. It is 5.6 m long, 2.56 m in diameter and creates a 2 T magnetic field in a 4.5 cm thick region. In order to achieve the strong magnetic field, 9 km of niobium-titanium superconducting wires are used. The wires are embedded in aluminum strips to minimize interactions between the magnet and the event particles. The other magnets are toroid magnets that use eight coils to generate a magnetic field up to 3.5 T in the muon spectrometer [91]. There are two at the end caps and one massive one surrounding the central section [40].

3.2.6 Trigger System

In addition to the detectors themselves, the data system is also of vital importance. Each event creates approximately 1.6 MB, worth of data and the event frequency of beam crossings in the detector is 40 MHz. This creates a total of 64 TB of raw data per second. As this is too much for the detector to read out and store, a trigger system is used to sort out most events. The Level 1 (L1) trigger is based on custom hardware, making it possible to accept or reject events in less than 2.5 µs. This lowers the frequency of read-outs to 100 kHz. The Level 2 (L2) is a software based selection system. It uses a large computer system to decide which events to store in 200 µs. This step further reduces the frequency of relevant events from 100 kHz to 1 kHz. In total this two step trigger process reduces the data from 64 TB to 1.6 GB per second. After the raw data passes the triggers, it is turned into physical objects such as jets or particles in offline reconstruction steps. After that the data can be used in analysis [43, 77].

CHAPTER 4

Event Reconstruction

Reconstructing events is a very complicated task involving multiple methods of combining raw data from all relevant sub-detectors. Different particles show very different signals in the detector. A π_0, for example, decays almost immediately and is only detected as two photons flying in opposite directions with similar energies in a boosted frame of reference, while a muon is reconstructed by pairing a track in the muon spectrometer to a track in the inner tracker. This chapter focuses on tau reconstruction which also shows a lot of variety since hadronic tau decays often involve both π_0 and hadron jets while leptonic tau decays involve either a muon or an electron.

4.1 Jet reconstruction

Since gluons and quarks produced in LHC collisions cannot exist in isolation, they hadronize (as described in chapter 2.1.4) and create a jet of collimated particles that deposit their energy in the calorimeter systems. Jets themselves are only abstract concepts and are actually made up of many particles. One significant factor in jet reconstruction is the choice of the opening angle R of jets. Here different choices lead to different interpretations of the same physical objects.

Jet reconstruction consists of three stages. The input to the process consists of TopoClusters, which are noise-suppressed clusters of calorimeter cells. A TopoCluster is constructed around a cell which has a 4σ above noise energy value. The cluster then expands by recursively including all neighboring cells with over 2σ energy in them.

In the second step the jets are formed. For this step, the anti-k_t algorithm [22] is used. Two different angular distances R are used for the reconstruction of different types of jets. Quark and gluon jets, as well as jets from hadronically decaying tau leptons, are typically more collimated and therefore use $R = 0.4$. Jets forming through the decay of hadronically decaying massive particles are typically less collimated [38] and therefore use $R = 1.0$. The necessity for this can be seen in fig. 4.1. It shows the angular distance between a pair of quarks that is created in a W boson decay as a function of the W boson transverse momentum. An approximate relation for the angular distance of two particles that are created through the decay of a massive particles χ with transverse momentum p_T^χ is

$$\Delta R \gtrsim \frac{2m^\chi}{p_T^\chi}. \tag{4.29}$$

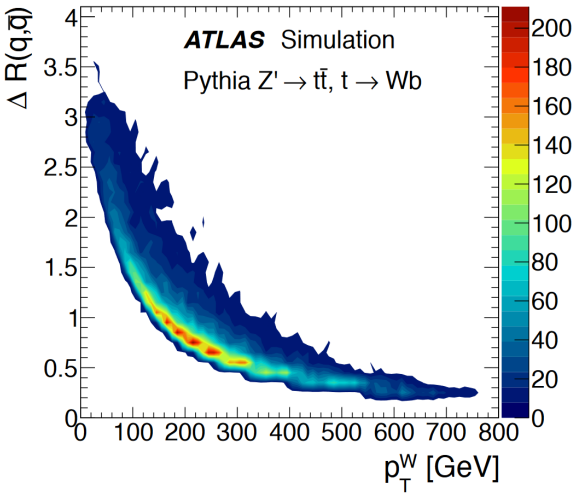


Figure 4.1: The angular distance ΔR between a pair of quarks created in a W boson decay as a function of the transverse momentum of the W boson [87].

In addition to the reconstruction the jets are also calibrated. This also happens in three steps. First they are corrected for pile-up effects. Then they are calibrated to Monte-Carlo data. In the final step the difference between MC and real data is corrected. The final jet reconstruction has a sub-1% energy uncertainty for 100 GeV to 1 TeV [87].

4.2 Tau Reconstruction

Since tau leptons usually decay before reaching the ID, tau reconstruction is split into two categories depending on their decaymode. Hadronic taus are reconstructed through jets. Their reconstruction is further explained in this section. Leptonic taus are measured by matching them to a measured muon or electron.

4.2.1 Hadronic Taus

Hadronic taus make up 65% of all tau decays. They are generally detected as one (1p) or three (3p) tracks in the inner detector and as energy deposits in the electromagnetic and hadron calorimeters that match the expected signals of charged or neutral hadrons. Most of the time tau leptons decay into charged and neutral pions. The largest background contributions for hadronic tau leptons are quark or gluonic jets, as they leave similar deposits in the calorimeters and have a large production cross section. In the case of 3p tau decays, tau candidates can be differentiated from background jets by using the measured secondary vertex of the tau, which can decay relatively far away from the primary interaction point. The additional energy deposits in the EM calorimeter from π^0 decays, which appear more often in tau decays than in gluonic jets, can also be used to differentiate hadronic tau jets from the background. Other background sources are light lepton candidates, which might look like 1p hadronic tau decays. If a tau candidate is also reconstructed as an electron or muon candidate, then it is no longer considered a tau candidate [56].

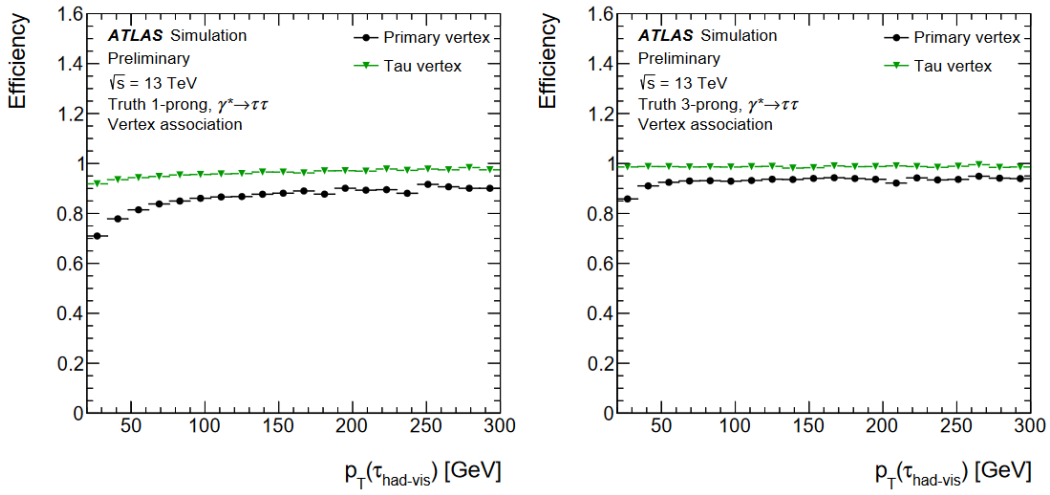


Figure 4.2: The efficiencies to select the correct tau production vertex and primare vertex as a function of the visible transverse momentum $p_T^{had-vis}$ of the tau decay for truth 1p (left) and 3p (right) tau decays [83].

Tau Identification

Tau candidates are seeded by jets reconstructed through the anti-k_t algorithm using a distance parameter $R < 0.4$. They are required to have $p_T > 5\,\text{GeV}$ and $|\eta| < 2.5$. The primary vertex that is assigned to the hadronic tau decay is defined as the vertex with the highest sum of the squared transverse momenta of the tracks. The efficiencies of choosing the correct tau production vertex and the correct primary vertex as a function of the visible transverse momentum $p_T^{had-vis}$ of the tau decay are shown in fig. 4.2. Tracks are assigned to the tau lepton if they have $\Delta R < 0.25$ compared to the reconstructed jet and also have at least $p_T > 1\,\text{GeV}$. The tracks are also required to have at least two associated hits in the pixel layers of the inner detector and at least a combined seven hits in the pixel and SCT layers. Requirements on the distance of closest approach to the tau production vertex are also used [83].

The final track association is done with a recurrent neural network. The variables used in the neural network are related to the kinematics of the jet as well as of the associated assigned tracks. They are listed in [83]. Four track categories are used. Tau Tracks (TT) originate from the tau lepton decay products. Conversion Tracks (CT) originate from electrons and positrons that are created by photon conversion in the detector. Isolation Tracks (IT) originate from quark and gluon jets of hard scattering interactions. Fake Tracks (FK) are mostly misreconstructed and pile-up tracks. The probabilities of reconstructing a given track as a certain track type is shown in fig. 4.3 for truth 1p and 3p tau decays [83].

Tau Energy Calibration

For the energy calibration, the sum of the energies of all TopoClusters with $\Delta R < 0.2$ to the hadronic tau lepton candidate is used. The pile-up energy contribution of this sum is then subtracted. Furthermore, the calorimeter deposits that can be assigned to a charged track are subtracted and give a measurement for the charged component of the hadronic tau decay. The resulting energy scale is parametrised in relation to the transverse momentum p_T, the pseudorapidity η, and the five decaymodes of the

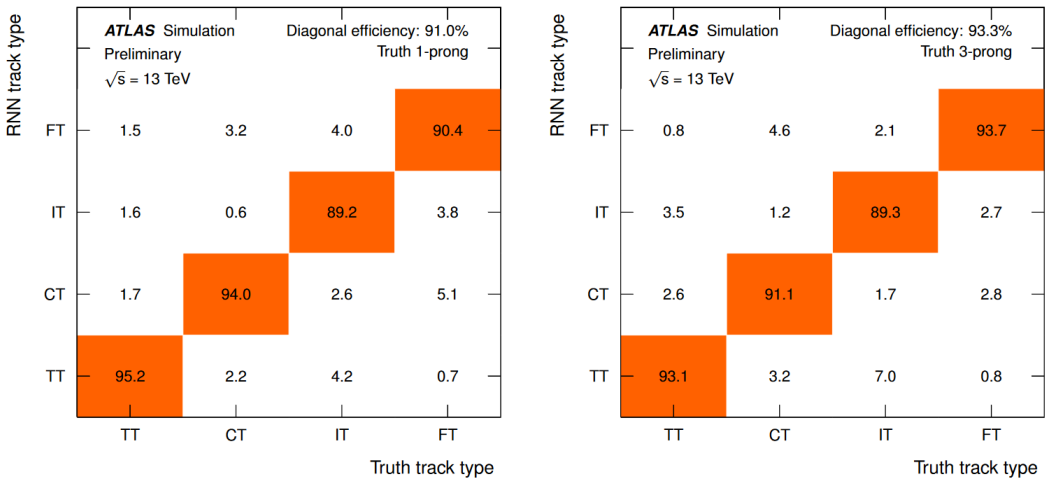


Figure 4.3: The probabilites to reconstruct a given track as a certian track type for truth 1p (left) and 3p (right) tau decays [83].

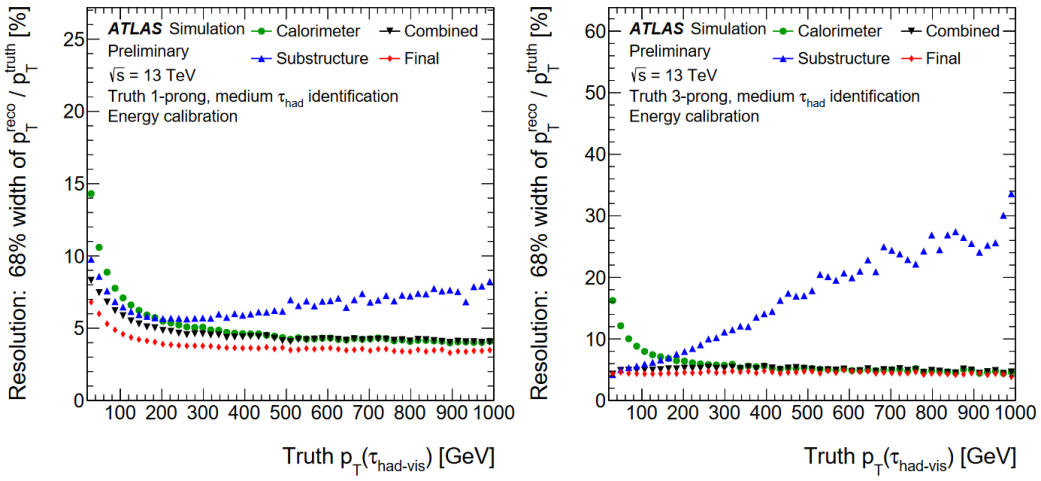


Figure 4.4: The energy 68%-quantile width of the energy resolution of the hadronic tau decay energy calibration for 1p (left) and 3p (right) candidates [83].

hadronic tau lepton. In the final step, a Boosted Regression Tree (BRT) is used. The input variables of the BRT are listed in [83]. The resulting energy resolutions as a function of p_T for truth 1p and 3p candidates are shown in fig. 4.4.

CHAPTER 5

Ditau Mass Reconstruction

Since $H \rightarrow \tau\tau$ and $Z \rightarrow \tau\tau$ events often look very similar in detector experiments and the Higgs mass lies on the shoulder of the Z boson resonance, a good ditau mass reconstruction is necessary. Good ditau mass reconstructions are also important for other analyses [2, 14, 37]. Currently the resolution is not good enough to detect excesses in the ditau channel with high precision as it is, for example, in the $Z/H \rightarrow \gamma\gamma$ channel. The goal of this thesis is to optimize the ditau mass reconstruction over a large mass spectrum that encompasses the masses of the Higgs and Z bosons. In essence, it is a simple problem where the energy of the ditau resonance is comprised of the energy of the two taus:

$$m^2_{\tau\tau} = (p_{\tau_1} + p_{\tau_2})^2. \tag{5.30}$$

However, the reconstruction also has to consider the neutrinos that elude the detector completely:

$$m^2_{\tau\tau} = (p_{\tau_1^{vis}} + p_{\tau_1^{\nu}} + p_{\tau_2^{vis}} + p_{\tau_2^{\nu}})^2. \tag{5.31}$$

Each ditau event typically contains two to four neutrinos. As mentioned in chapter 3.2.1, the only information we have about them is the reconstructed missing transverse momentum which is very sensitive to the detector resolution. There are multiple approaches to this problem, some of which will be covered in this chapter [52].

5.1 Visible Mass

The simplest ditau mass reconstruction is to use the visible reconstructed mass of the event and to ignore the neutrinos completely. In this approach, eq. 5.31 simplifies to

$$m^2_{\tau\tau,vis} \approx (p_{\tau_1^{vis}} + p_{\tau_2^{vis}})^2. \tag{5.32}$$

This leads to a significant underestimation of the ditau mass and, since the fraction of the tau momenta that is carried by the neutrino is randomly distributed, it also leads to a smearing of the reconstructed mass distribution. Fig. 5.1 (left) shows a density plot of the visible reconstructed mass as well as the true mass of a $Z \rightarrow \tau\tau$ and $H \rightarrow \tau\tau$ samples. The samples used are introduced in chapter 7.5. As expected, the mass is significantly underestimated (μ^H_m : 124.4 GeV $\rightarrow$ 83.9 GeV , μ^Z_m : 93.6 GeV $\rightarrow$ 66.2 GeV), and the standard deviation of both distributions is much higher for the visible mass than it is for

the true mass distributions (σ_m^H : 0.2 GeV $\rightarrow$ 16.3 GeV , σ_m^Z : 13.8 GeV $\rightarrow$ 14.2 GeV). Note that for the Z sample, even though the standard deviation has only increased slightly, it is much more significant since the mean value is about 30% lower. Fig. 5.1 (right) shows the relative resolution of the reconstruction. Almost all events are reconstructed with a too low mass, which results in a large bias. On average the reconstruction underestimates the mass by 32.5% for the Higgs sample and 29.0% for the Z sample. Only 0.38% of all Higgs events and 0.49% of all Z events are overestimated. This can only happen in events where the neutrino carries almost no momentum and the detector overestimates the visible part of the tau decay.

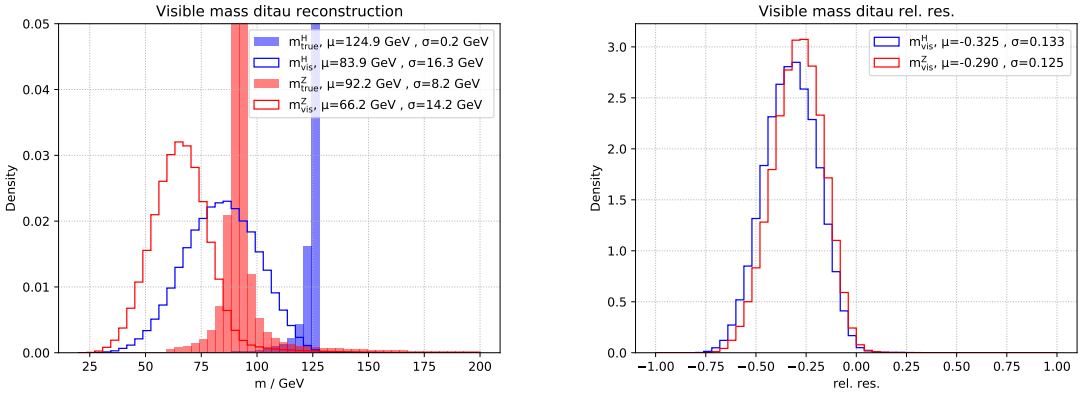


Figure 5.1: The reconstructed visible mass of $H \rightarrow \tau\tau$ and $Z \rightarrow \tau\tau$ events (left) and their relative mass resolution (right).

5.2 Collinear Approximation

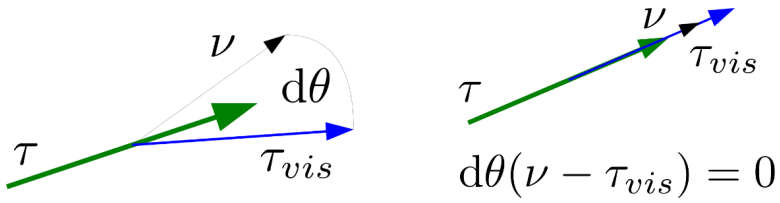


Figure 5.2: The angular assumption made in the collinear approach. It is assumed that the neutrino and the visible part of the tau decay have no angular distance $\theta = 0$.

The collinear approximation is another method of reconstructing the ditau mass.It assumes that the neutrino's four-vectors point in the same direction as the visible part of the tau decay (i.e. $\Delta R = 0$), as shown in fig. 5.2. This approximation can be very accurate under certain conditions and in principle makes an analytic solution possible. This approximation does, however, make some assumptions. It assumes that the entire missing transverse energy E_T^{miss} measured is produced by the neutrinos. This is generally not true since hard jets and pile-up also contribute to the MET. If any particles fly into the blind regions of the detector, they can also lead to misleading MET values. The collinear

approximation also only holds up under the assumption that the mass of the mother particle is much higher than the mass of the tau leptons that it decays into ($m_{\text{H/Z}} \gg m_\tau$). In addition, the detector resolution is also neglected in this approach. Under these assumptions the problem can be solved analytically with

$$E_{T,x}^{mis} = p_{mis,\tau_1} \sin\theta_{vis,\tau_1} \cos\phi_{vis,\tau_1} + p_{mis,\tau_2} \sin\theta_{vis,\tau_2} \cos\phi_{vis,\tau_2} \tag{5.33}$$

$$E_{T,y}^{mis} = p_{mis,\tau_1} \sin\theta_{vis,\tau_1} \sin\phi_{vis,\tau_1} + p_{mis,\tau_2} \sin\theta_{vis,\tau_2} \sin\phi_{vis,\tau_2} \tag{5.34}$$

$$x_i = \frac{p_{vis,i}}{p_{vis,i} + p_{mis,i}} \tag{5.35}$$

$$m_{\tau\tau}^{coll} = \frac{m_{\tau\tau}^{vis}}{\sqrt{x_1 x_2}} \tag{5.36}$$

where x_i is the visible fraction of the tau momentum.

If the transverse momentum of the resonance particle is low, then the two tau leptons fly in opposite directions, which degenerates the possible solutions for the equations. Events where the resonance is boosted significantly in the transverse plane by hadronic jets don't have this problem.

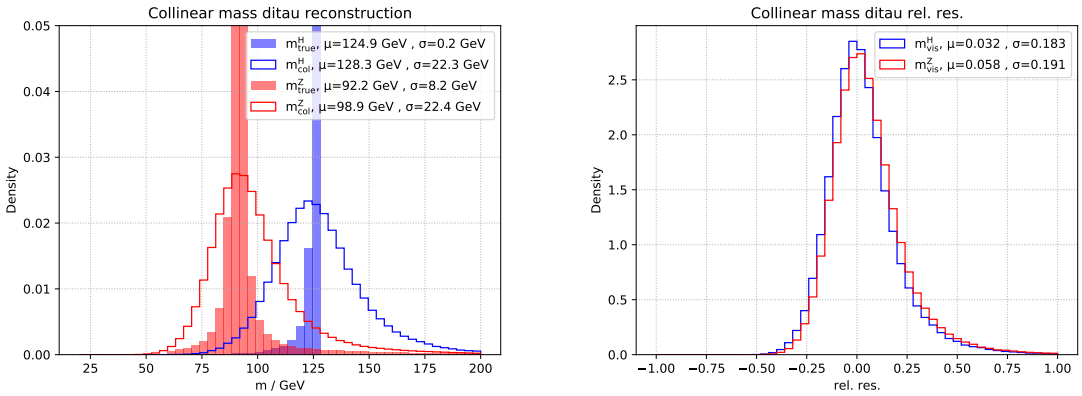


Figure 5.3: The collinear mass reconstruction of $H \rightarrow \tau\tau$ and $Z \rightarrow \tau\tau$ events (left) and their relative mass resolutions (right).

Fig. 5.3 (left) shows the true mass and the collinear approximation. The reconstruction is already significantly improved. The mean collinear mass for the Higgs sample is 128.3 GeV and for the Z sample 98.9 GeV. The standard deviation is 22.3 GeV for the Higgs sample and 22.4 GeV for the Z sample. The overestimation comes from the visible tail towards very high masses. This tail comes from the aforementioned degeneration for events when the boson has low transverse momentum of the boson. The relative resolution of the approximation is shown in fig. 5.3 (right). The mean relative resolution for the Higgs sample is 0.032 and for the Z sample is 0.058. The distribution looks very similar for both samples, though the Z sample has a slightly higher bias and width.

5.3 Missing Mass Calculator

The missing mass calculator (MMC) [50] is the approach currently used in ATLAS. The principle behind the MMC is to calculate kinematic probability density functions from simulations and then use those to calculate the probability for all viable kinematic distributions of the event.

In order to reconstruct the ditau mass without using the collinear approximation, six to eight variables need to be solved for. These are the three-momentum $p_{mis,0}$ and $p_{mis,1}$ of the neutrinos and, if a tau decays leptonically, then the invariant mass of the two neutrino system as well. However, we only have four equations available to solve for the variables:

$$E_{T,x}^{mis} = p_{mis,\tau_1} \sin\theta_{vis,\tau_1} \cos\phi_{vis,\tau_1} + p_{mis,\tau_2} \sin\theta_{vis,\tau_2} \cos\phi_{vis,\tau_2} \tag{5.37}$$

$$E_{T,y}^{mis} = p_{mis,\tau_1} \sin\theta_{vis,\tau_1} \sin\phi_{vis,\tau_1} + p_{mis,\tau_2} \sin\theta_{vis,\tau_2} \sin\phi_{vis,\tau_2} \tag{5.38}$$

$$M_{\tau_0}^2 = m_{mis_0}^2 + m_{vis_0}^2 + 2 \cdot \sqrt{p_{vis_0}^2 + m_{vis_0}^2}\sqrt{p_{mis_0}^2 + m_{mis_0}^2} - 2p_{vis_0}p_{mis_0}\cos(\Delta\theta_{vm_0}) \tag{5.39}$$

$$M_{\tau_1}^2 = m_{mis_1}^2 + m_{vis_1}^2 + 2 \cdot \sqrt{p_{vis_1}^2 + m_{vis_1}^2}\sqrt{p_{mis_1}^2 + m_{mis_1}^2} - 2p_{vis_1}p_{mis_1}\cos(\Delta\theta_{vm_1}) \tag{5.40}$$

$M_{\tau_{0/1}}^2$ is the square of the invariant tau mass and corresponds to 1 776.86 GeV. $\Delta\theta_{vm_{0/1}}$ is the angular difference between the visible and invisible decay products of the tau lepton. m_{inv} is the invariant mass of the neutrino system and is set to zero for hadronically decaying taus. The two missing equations needed to solve for the six variables in the case of hadronically decaying tau leptons are due to the fact that it is impossible to reconstruct the pseudorapidity of the missing energy in hadronic colliders. As there are infinite solutions to these equations, the MMC uses probability density functions to calculate the probability with a Markov-Chain-Monte-Carlo process for a physically realistic subset of these solutions. The probability density functions that are used in the MMC are $\Delta\theta_{vm_{0/1}}$, the ratio of visible to invisible momentum $R_{0/1}$, and the resolution of the transverse missing energy $\Delta E_{x/y}^{MET}$. They are determined separately for $\tau_{had}\tau_{had}$, $\tau_{had}\tau_{lep}$ and $\tau_{lep}\tau_{lep}$ events. The total probability for a certain kinematic distribution is then computed by using the product of the probability density functions [63]:

$$P_{tot} = P(\Delta\theta_{vm_0}) \cdot P(\Delta\theta_{vm_1}) \cdot P(R_0) \cdot P(R_1) \cdot P\left(\Delta E_x^{MET}\right) \cdot P\left(\Delta E_y^{MET}\right) \tag{5.41}$$

There are three different options for the output of the MMC:

- 1. MAXW: solution of the MMC with the highest probability.
- 2. MLM: the ditau-mass with the highest probability. This is the output used for the comparison in this thesis.
- 3. MLNU3P: the most likely 3-momentum of neutrino.

The MMC output for the Z and Higgs sample are shown in fig. 5.4 (left). The mean MMC mass for the Higgs sample is 120.2 GeV with a 16.0 GeV standard deviation. For the Z sample the mean is 92.8 GeV with a 17.4 GeV standard deviation. The MMC produces a better reconstruction than the collinear approximation for both samples. It is noteworthy that the tail to large masses is no longer present and the mean Z reconstruction is actually very close to the true Z mass. Fig. 5.4 (right) shows the relative resolution of the MMC reconstruction. The mean relative resolution of the H sample is

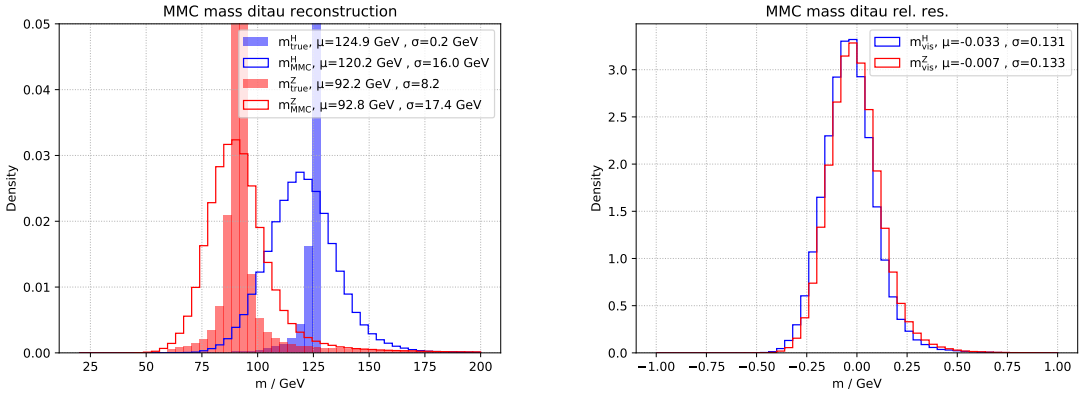


Figure 5.4: The MMC mass reconstruction of H $\rightarrow \tau\tau$ and Z $\rightarrow \tau\tau$ events (left) and their relative mass resolutions (right).

-0.033 and of the Z sample it is -0.007. While the collinear approximation overestimates the masses on average because of the tails, the MMC slightly underestimates the masses.

While the MMC does have advantages in comparison to the collinear approximation, a new parametrisation for the MMC has to be applied every time it has to be retuned, which requires a lot of work. The MMC reconstruction also takes a significant amount of time to calculate the probabilities of events. In addition, it is also currently only trained on a Standard Model Monte-Carlo Z sample and is therefore not calibrated for other mass ranges.

CHAPTER 6

Artificial Neural Networks

Artificial neural networks (ANNs), usually simply referred to as neural networks (NNs), are a method of machine learning that has a large range of applications like pattern recognition, data processing, and solving non-linear problems. It is based on the idea of how biological neurons are connected in the brain. The theory behind NNs started in 1943 when Warren McCulloch and Walter Pitts created the first computational model for neural networks. A large step in advancing research into NNs was taken in 1975 when Paul Werbos invented backpropagation used to train multi-layered NNs. This led to a great advancement in research and applications. Unless otherwise specified, the source for this chapter is [19].

6.1 Basic Structure

A neural network typically consists of multiple layers L_i of artificial neurons v_n, also called nodes, that are connected to each other. The first layer is called the input layer, in which each node represents one of the input variables of the NN. The last layer is called the output layer. It can consist of multiple nodes but for the purpose of this thesis there is only one output. The layers inbetween are called hidden layers. An example diagram of a NN with four input variables and one hidden layer with five nodes can be seen in fig. 6.1.

6.1.1 Nodes

The output of every node is connected to all nodes of the following layer. Each input of a node is multiplied by a weight w_i. The inputs are summed up and an additional bias b is also included. This value, labeled z, is then taken as input of an activation function f_{act} to give the output of the node:

$$output = f_{act}\left(\sum_{i=1}^{n} x_i w_i + b\right) = f_{act}(z). \tag{6.42}$$

The weights w_i and biases b make up the variables θ that are varied during the training to optimize the neural network. At first the weights and bias are all initialized pseudorandomly through a seed.

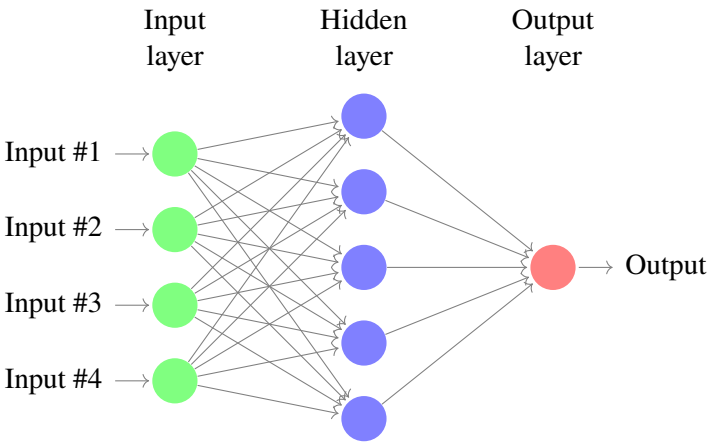


Figure 6.1: Example neural network diagram with four input variables, one hidden layer, and one output [53].

6.1.2 Activation functions

The activation function is the final step in defining the output of a node, which is then sent to the next layer. Activation functions are necessary to make NNs non-linear. Without them, NNs would not be able to solve non-linear problems like the ditau mass reconstruction [57]. Some of the most common activation functions are as followed [88]:

1. Sigmoid: $f_{act}(x) = \frac{1}{1+e^{-x}}$
2. Hyperbolic tangent: $f_{act}(x) = \tanh(x)$
3. Rectified linear unit (ReLU): $f_{act}(x) = \max(0, x)$
4. Scaled Exponential Linear Unit (SELU): $f_{act} = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$

6.2 Training

In order to train NNs, first a training sample made up of m events is needed. Each event must consist of an input array $\vec{x}$ and a target y. NNs often aim to classify events into either 0 or 1, but the target can also be a continuous variable. The NNs in the latter case are called Regression Neural Networks (RNN). When feeding an event into the input layer of the neural network it produces an output $\hat{y}$ at the output layer. For the training, a loss function L must be defined. A typical loss function is the squared error loss:

$$L(y;\theta) = \frac{1}{m}\sum_{0}^{m}(\hat{y} - y)^2. \tag{6.43}$$

The goal of the training process is to minimize this loss function. This is achieved through recursive backpropagation from the output layer to the input layer [86]. The derivate of the loss function as a function of each weight is computed as

$$\frac{\delta L}{\delta w_{ij}} = \frac{\delta L}{\delta o_j}\frac{\delta o_j}{\delta z_j}\frac{\delta z_j}{\delta w_{ij}} = \frac{\delta L}{\delta o_j} f'_{act}(z_j)\frac{\delta z_j}{\delta w_{ij}}. \tag{6.44}$$

where w_{ij} describes the weight of the connection between node i to node j, and o_j is the output of the node j as calculated in eq. 6.42. The backpropagation is relevant to calculate the $\frac{\delta L}{\delta o_j}$ part. The partial derivative must be expanded to include every node n that uses o_j as an input:

$$\frac{\delta L}{\delta o_j} = \sum_{n \in N} \left(\frac{\delta L}{\delta o_n} \frac{\delta o_n}{\delta z_n} \frac{\delta z_n}{\delta o_j} \right) = \sum_{n \in N} \left(\frac{\delta L}{\delta o_n} f'_{act}(z_n) w z_{jn} \right). \tag{6.45}$$

This allows for the calculation of $\frac{\delta L}{\delta w_{ij}}$. The recursive term with the f'_{act} term can be combined into a computable variable:

$$\delta_j = \frac{\delta L}{\delta o_j} \frac{\delta o_j}{\delta z_j} = \frac{\delta L}{\delta o_j} f_{act}(z_j). \tag{6.46}$$

The weights of the nodes can now be updated through the stochastic gradient descent (SGD) optimizer with a learning rate α, where usually $\alpha << 1$ [70]:

$$w_{ij} = w_{ij} - \alpha \frac{\delta L}{\delta w_{ij}}. \tag{6.47}$$

All weights are updated simultaneously. This process is not conducted on each event individually but instead is done in batches. Both the learning rate α and the batch size are variables that are set in the neural network setup before the training starts.

6.3 Further NN optimization

While the above explanations are sufficient to create NNs, there are many more methods to improve the performance of neural networks. They are mostly modifications to the training algorithm, usually done by modifying the optimizer of the node weights.

To create a faster optimization, the input data is often normalized to a Gaussian distribution, with a standard deviation which is equal for all input variables. This is visually depicted in fig. 6.2.

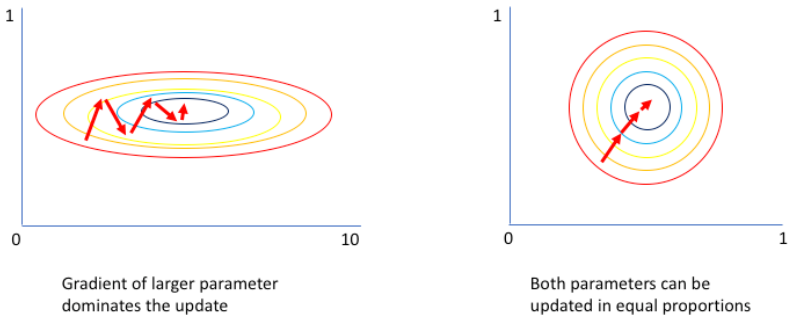


Figure 6.2: visual depiction of SGD before and after input data normalization [65].

During the SGD the gradient might be too large for the weights to fall into a minimum, and instead oscillates around it. To circumvent this problem the learning rate α can be decreased during the learning process if the loss plateaus.

A common alternative to the SGD algorithm is the Adam optimizer [67]. It creates individual learning rates for each parameter. The learning rates are updated during the training process through the inclusion of momentum terms:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\alpha \left(\frac{\delta L}{\delta w_{ij}}\right)_t \tag{6.48}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\alpha \left(\frac{\delta L}{\delta w_{ij}}\right)_t^2. \tag{6.49}$$

These variables represent the first and second momentum of the gradient. β_1 and β_2 are constants that are set during the neural network initialization. An additional correction to m_t and v_t is applied to avoid biases:

$$\hat{m_t} = \frac{m_t}{1-\beta_1} \tag{6.50}$$

$$\hat{v_t} = \frac{m_t}{1-\beta_2}. \tag{6.51}$$

The final update to the weights is then

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{v_t}} + \epsilon}\hat{m_t} \tag{6.52}$$

where ϵ is some very small number. The Adam optimizer can often increase the training speed since it allows for faster optimization of the individual weights.

CHAPTER 7

Monte Carlo Samples

7.1 Event Simulation

Simulated data is often used to calibrate detectors and verify experiment data. They are also used to train neural networks since they require true level data. The data samples used in this thesis are Monte-Carlo-generated samples based on pp collisions. For this, random events are generated which follow probability density functions derived from theoretical calculations and experimental data. During the event generation, multiple processes must be simulated in matrix calculations. First, hard inelastic processes are generated. Initial state radiation and final state radiation are added on to these processes. In additional, gluonic jets and hadronic fragmentation are generated. There are integrated frameworks which simulate these effects. *SHERPA* [89, 20, 17] is used for parton showers and QED radiation as well as for the generation of tau leptons. *PYTHIA8* [9, 66] is used for hadronization processes and also generates tau leptons. *POWHEG* [76, 11, 55] is used for NLO matrix element generation. Tab. 7.1 lists the samples used in this thesis and some of their characteristics. The spin-0 samples are generated with *POWHEG* and are based on the *PYTHIA8* event generator. The spin-1 samples are generated with *SHERPA*.

Sample	Size			Spin of ditau resonance
	total	60 GeV<m<220 GeV	m<60 GeV	
γ*	4 176 070	2 142 022	2 454	1
Tauspinner	941 026	384 731	130	1
Spin0hh	297 683	151 359	342	0
H-Sample	195 862	195 862	0	0
Z-Sample	873 241	873 241	0	1

Table 7.1: The used MC samples in this thesis.

7.2 γ^* Sample

The first sample is a spin-1 γ^* sample based on proton-antiproton interactions that involve an off-shell photon decaying into two hadronic taus. The mass spectrum of the γ^* sample can be seen in fig. 7.1

(left). There are a total of 4 176 070 events in the sample. The mass distribution is not flat. Instead, there is a sharp decline for smaller masses and there are only 2 454 events (0.06% of all events in the sample) under 60 GeV. This is because the tau reconstruction for lower energies becomes a lot more difficult and the reconstruction steps cut deeply into the phase space of low energy γ^* events. The distribution reaches has a mean at 263.0 GeV. 2 142 022 events (51.29% of all events in the sample) have a mass over 220 GeV. This sample is by far the largest out of all samples used, but it emulates neither the spin of the Higgs boson nor the parity violation of the Z boson.

Since a RNN trained on this sample should not be biased to high energies, the sample is reweighted to have a flat distribution for the mass range of 60 – 220 GeV. There are not sufficient statistics to extend this range to lower masses and higher masses are not relevant for $H \longrightarrow \tau\tau$ and $Z \longrightarrow \tau\tau$ analysis. The mass spectrum after reweighting can be seen in fig. 7.1 (right).

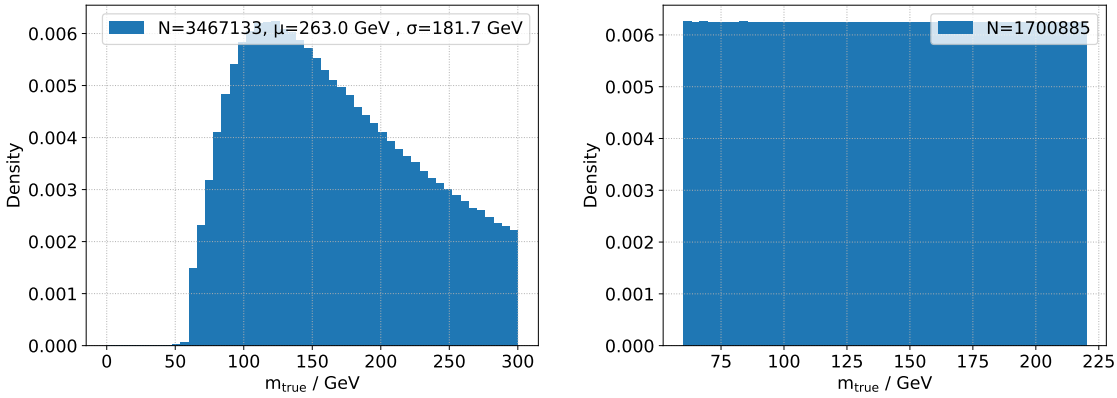


Figure 7.1: Mass distribution of the γ^* sample (left) and the reweighted sample (right)

The cuts that are applied during the creation of the sample in the derivation steps are listed in tab. 7.2. The relative and absolute efficiency of the cuts is also listed. DxAOD (Derived Analysis Object Data) describes the derivation of the events. Only events with two reconstructed taus, which have to have $p_T > 33$ GeV and $p_T > 23$ GeV, as well at least one tau lepton with a loose tau-ID, pass through the derivation step. This is the biggest cut of the sample. The NTaus cut demands there to be exactly two taus that are reconstructed in the event, and the Ditau common Vx cut demands that both tau leptons are from approximately the same vertex, as reconstructed by the ATLAS detector. In total, only 18.26% of all events get accepted into the sample. In addition to the cuts applied during the creation of the sample, a tau identification cut is applied to make sure that the two reconstructed taus in the event are actually tau leptons (a check on the PDGID = 15 or = −15 is done), and a final cut is carried out to make sure there is a valid MMC reconstruction for the event.

Since the γ^* sample is a spin-1 sample we expect both tau leptons in the decay to share similar Υ distributions. Fig. 7.2 shows a 2D histogram of Υ_0 and Υ_1. The figure shows accumulations of events in the corners, which is an indication for a spin-1 sample.

The distribution of the input variables used in the neural network training are shown in the appendix A.

Cut	N	Efficiency	Total Efficiency
All events	29 781 000	100%	100%
DxAOD	9 501 676	31.91%	31.91%
Processed events	9 501 676	100%	31.91%
NTaus	6 352 713	66.86%	21.33%
Ditau common Vx cut	5 436 998	85.59%	18.26%

Table 7.2: Cut efficiencies of the derivation of the γ^* sample. Both the individual efficiencies and the total efficiency after each step are listed.

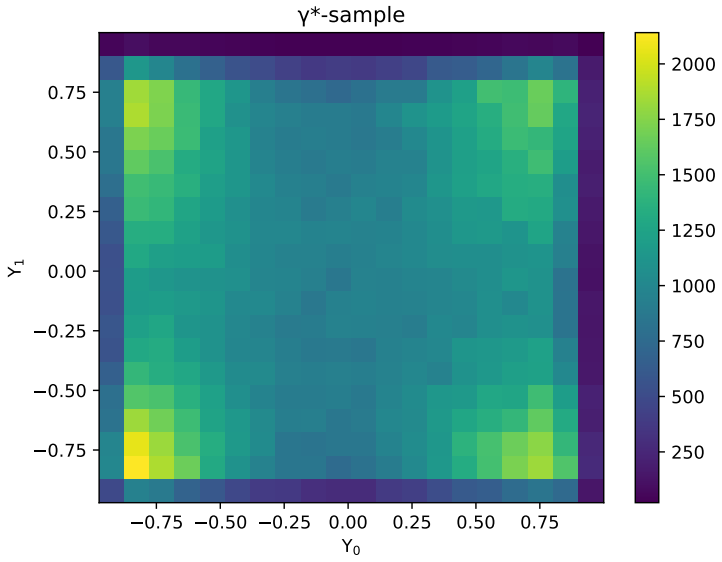


Figure 7.2: 2D histogram showing the relation between Υ_0 and Υ_1 for the γ^* sample

7.3 Tauspinner Sample

The goal of the tauspinner sample is to imitate the parity violation effects of the Z decay. For this, the *TauSpinner* [82] program is used on a γ^* sample to generate weights that simulate the parity violation of the Z boson on each event. The mass spectrum of this sample can be seen in fig. 7.3 (left). The mass distribution here is also not flat and drops off sharply for low energies. Only 130 events (0.01% of all events in the sample) have a mass below 60 GeV. As with the γ^* sample, the distribution drops off slowly again for higher masses. The mean of the distribution is 286.6 GeV. For this sample 556 165 events (59.10% of all events in the sample) have a mass above 220 GeV. This sample is also reweighted to a flat distribution for neural network training. The results of the reweighting is shown in fig. 7.3 (right). Due to the lower statistics, the reweighting is not able to create a perfectly flat distribution (especially for low masses).

The derivation cuts that are applied to this sample and their relative and absolute efficiencies are listed in tab. 7.3. This sample has some addition cuts applied during its creation that are not present in the γ^* sample. The lepton veto cut removes all events in which a lighter lepton is reconstructed and the

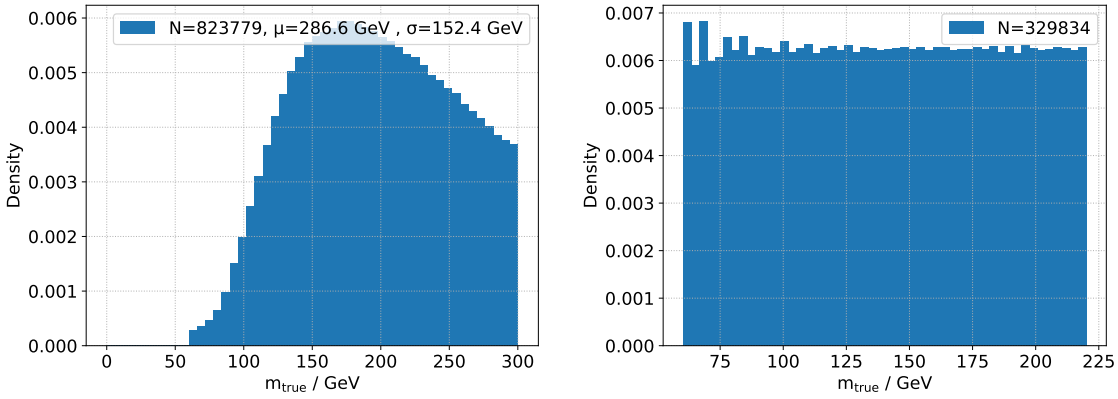


Figure 7.3: Mass distribution of the tauspinner sample (left) and the reweighted sample (right)

MET cut is only accepts events where MET $> 20\,\text{GeV}$. As with the γ^* sample, the only cuts that are applied after the creation of the sample are the PDGID and MMC cuts. The LooseBad EventCleaning filter removes events which are marked as having a bad jet in their reconstruction and the dEta cut removes events where $|\Delta\eta| < 1.5$.

Cut	N	Efficiency	Total Efficiency
All events	27 459 000	100%	100%
DxAOD	9 059 046	32.99%	32.99%
Lepton veto	2 767 096	30.55%	10.08%
LooseBad EventCleaning	2 752 715	99.48%	10.02%
Tau pT cut	2 457 835	89.29%	8.95%
Ditau common Vx cut	2 294 569	93.36%	8.36%
MET cut	1 649 848	71.90%	6.01%
Ditau dEta cut	1 429 324	86.63%	5.21%

Table 7.3: Cut efficiencies of the derivation of the tauspinner sample. Both the individual efficiencies and the total efficiency after each step are listed.

The $\Upsilon_{0,1}$ distribution of this sample is shown in fig. 7.4. As with the γ^* sample, the figure shows accumulations of events in the corners, which is indicative of a spin-1 sample.

The distribution of the variables used in the neural network training are shown in the appendix A.

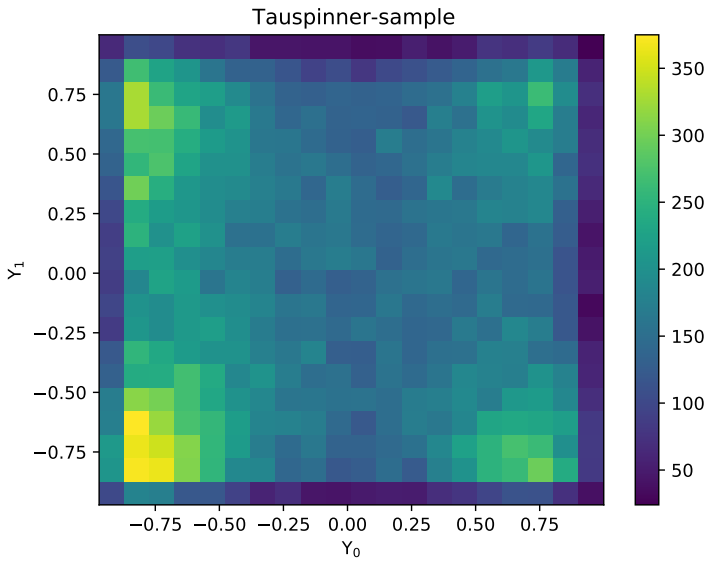


Figure 7.4: 2D histogram showing the relation between Υ_0 and Υ_1 for the tauspinner sample

7.4 Spin0hh Sample

Since the RNN should be optimized in both $Z \rightarrow \tau\tau$ and $H \rightarrow \tau\tau$ events an additional sample is used for the training. For this, a spin-0 sample is used to emulate the kinematics of Higgs boson decays across a large mass range. The sample consists of multiple beyond the Standard Model (BSM) Higgs samples with different masses and broad widths which are combined to create a large mass distribution. The mass distribution of the sample is shown in fig. 7.5 (left). The only cuts applied to this sample are shown in tab. 7.4. As with the other samples, additional PDGID and MMC cuts are also applied to the sample.

The distribution of the spin0hh sample is even more biased to higher masses than the other two training samples. It has only 463 950 events, making it the smallest out of the three training samples. Only 0.11% of all events are under 60 GeV. The mean of the distribution lies at 228.4 GeV and 51% of the events have masses over 220 GeV. The reweighting to a flat mass distribution for the $60\,\text{GeV} < m_{true} < 220\,\text{GeV}$ events is shown in 7.5 (right). Similarly to the tauspinner sample, it shows indications of low statistics for low masses.

The derivation cuts that are applied to this sample and their relative and absolute efficiencies are listed in tab. 7.4.

A 2D-histogram showing the correlation between Υ_0 and Υ_1 is shown in fig. 7.6. When the energy asymmetry Υ of one tau is near zero, the Υ of the other tau is more likely to have a higher absolute value. This leads to the ring structure that can be seen in the figure, and therefore confirms that the spin0hh sample is a spin-0 sample.

The distribution of the variables used in the neural network training are shown in the appendix A.

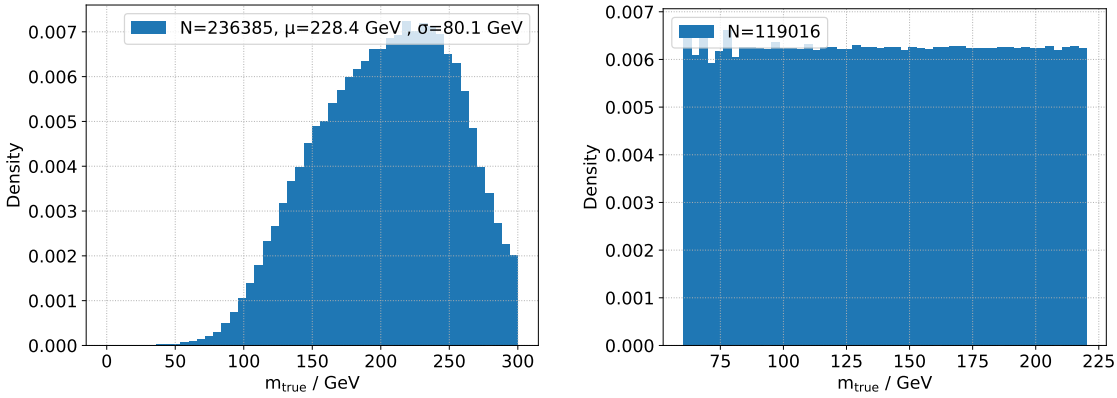


Figure 7.5: Mass distribution of the spin0hh sample (left) and the reweighted sample (right)

Cut	N	Efficiency	Total Efficiency
All events	2 366 000	100%	100%
DxAOD	1 136 291	48.03%	48.03%
Passed through channel events	680 659	59.90%	28.77%
LooseBad EventCleaning	676 645	99.41%	28.60%
Tau pT cut	491 562	72.65%	20.78%
Ditau common Vx cut	463 950	94.38%	19.61%

Table 7.4: Cut efficiencies of the derivation of the spin0hh sample. Both the individual efficiencies and the total efficiency after each step are listed.

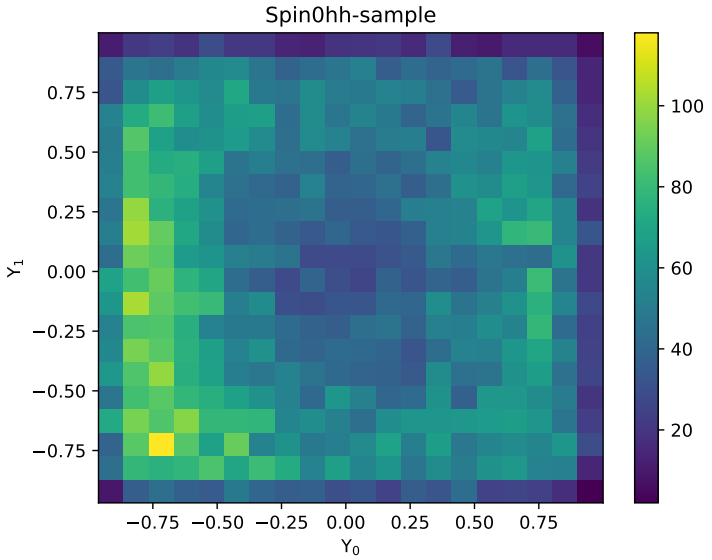


Figure 7.6: 2D histogram showing the relation between Υ_0 and Υ_1 for the spin0hh sample

7.5 H Sample and Z Sample

The H and Z samples are also Monte-Carlo generated samples. They are not used for the training but instead for the evaluation of RNNs. The H sample has a total of 197 240 events and the Z sample has a total of 878 838 events. Their respective ditau mass distributions are shown in fig. 7.7. The tail to lower masses is from final state radiation. The H sample has a mean mass of $m_H = 124.9\,\text{GeV}$. The mean mass of the Z sample is $m_Z = 92.0\,\text{GeV}$. This is approximately 1% higher than the theoretical value, because the cuts to the sample primarily remove low-mass events.

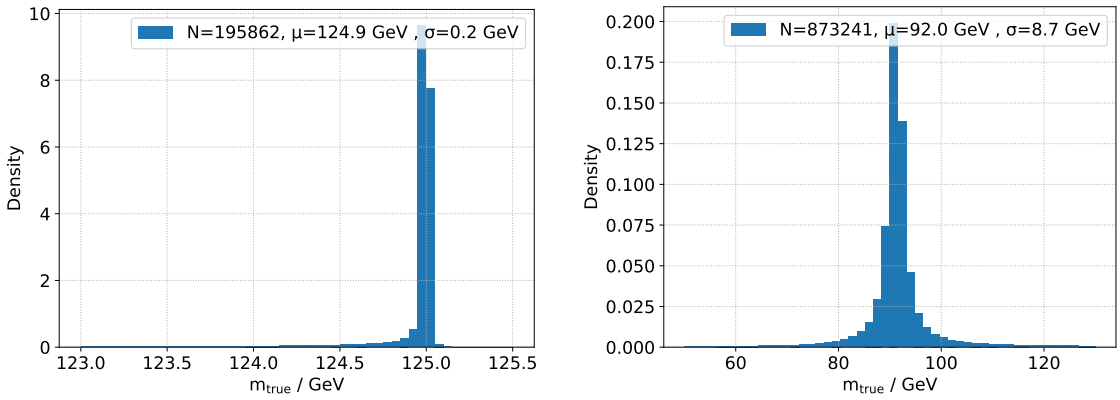


Figure 7.7: Mass distribution of the true ditau mass including FSR effects of the H sample (left) and the Z sample (right).

The derivation cuts that are applied to this sample and their relative as well as their absolute efficiencies are listed in tab. 7.5. The cuts that are applied to these samples are stricter than for the three training samples, since these samples should be closer to the conditions of an analysis which uses the ditau reconstruction. To ensure that the tau leptons reconstructed during the online trigger level and the offline tau reconstruction match, a trigger matching cut is applied. The tau p_T cuts are $p_{T_{\tau_0}} > 40\,\text{GeV}$ and $p_{T_{\tau_1}} > 30\,\text{GeV}$. A Leading jet cut of $p_T > 70\,\text{GeV}$ with an additional pseudorapidity cut of $|\eta| < 3.2$ is applied. The angular cuts between the two tau leptons are $0.6 < \Delta R < 2.5$ and $|\Delta\eta| < 1.5$. To ensure the ditau resonance is sufficiently boosted for the reconstruction, a MET $> 20\,\text{GeV}$ selection is also used. In addition, a collinear approximation cut of $0.1 < x < 1.4$ is applied on both tau leptons, where x stands for the visible fraction of the tau decay as defined in eq. 5.35. The listed cuts primarily serve to increase the background rejection in analysis and filter out tau leptons for which a ditau mass reconstruction becomes more difficult. Only 3.22% of the H sample and 2.04% of the Z sample passes all cuts.

The 2D histograms of the charged energy asymmetries for the H and Z samples are shown in fig. 7.8. The H sample shows the distinct ring structure while the Z sample has an X structure with accumulations of events in the corners. This verifies the spins of the samples and also shows that the energy asymmetry contains information that can be used to reconstruct the spin of the resonance.

Cut	H-sample			Z-sample		
	N	Eff.	Tot. eff.	N	Eff.	Tot. eff.
All events	6 129 000	100%	100%	42 980 400	100%	100%
DxAOD	5 042 028	82.27%	82.27%	31 703 877	73.76%	73.76%
Passed through channel	954 410	18.93%	15.57%	4 446 256	14.02%	10.34%
EventCleaning	947 936	99.32%	15.47%	4 415 750	99.31%	10.27%
Trigger matching	947 901	>99.99%	15.47%	4 415 555	>99.99%	10.27%
Tau pT cut	715 802	75.51%	11.68%	2 666 447	60.39%	6.20%
common Vx cut	710 725	99.29%	11.60%	2 647 764	99.30%	6.20%
Leading jet cut	322 081	45.32%	5.26%	1 519 770	57.40%	3.54%
MET cut	273 943	85.05%	4.47%	1 211 731	79.73%	2.82%
Ditau dR cut	252 407	92.14%	4.12%	1 154 237	95.25%	2.69%
Ditau dEta cut	240 761	95.39%	3.93%	1 133 797	98.22%	2.64%
Coll approx cut	197 240	81.92%	3.22%	878 838	77.51%	2.04%

Table 7.5: Cut efficiencies of the derivation of the H sample and Z sample. Both the individual efficiencies and the total efficiency after each step are listed.

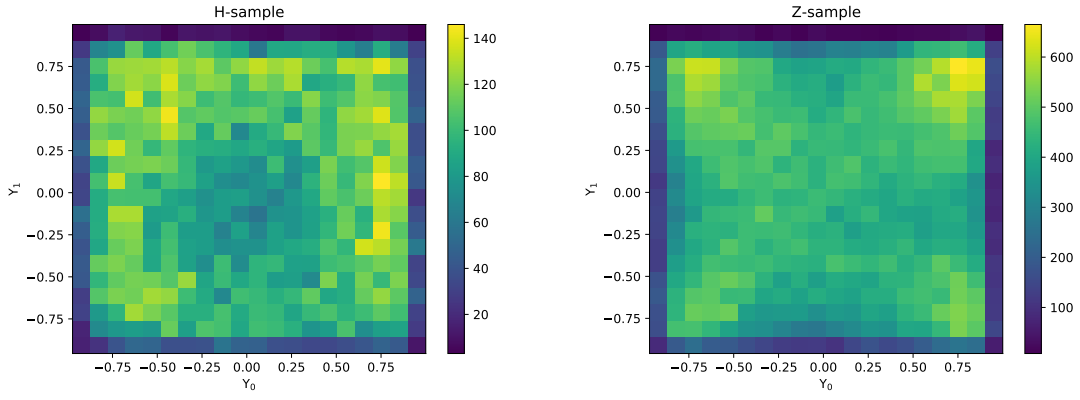


Figure 7.8: 2D histogram showing the relation between Υ_0 and Υ_1 for the H sample (left) and the Z sample (right).

CHAPTER 8

Neural Network Analysis

This section discusses the setups and performances of multiple neural networks. The neural networks are trained on the samples introuced in chapter 7. First, the general architecture of the neural networks used in this thesis are presented. Then, the methods of evaluating the neural networks are introduced. In the next section, the performance of neural networks trained on the different Monte Carlo samples are discussed and compared to the MMC performance. Some modifications to the neural network are in the next sections. The final section includes a summary of all neural networks presented in this chapter.

8.1 Architecture

The general architecture of the RNN used in this thesis is optimized as determined in [93]. Unless otherwise stated, all neural networks in this thesis use this architecture. Four layers are used with 16 nodes each. TensorFlow is used for the neural network creation and training. Using fewer layers or nodes leads to a sharp decline in NN performance, while adding more does not significantly alter the performance for this use case and leads to a large increase in training time. The SELU activation function is used for all layers. The batch size is 32 and a learning rate of $\alpha_0 = 0.001$ is used. Every time there is no improvement in the loss detected for 15 epochs, the learning rate is halved. No dropout or regularization is used. The loss function is the mean square error loss as defined in 6.43 and the Adam optimizer is used with a first momentum of $\beta_1 = 0.9$ and a second momentum of $\beta_2 = 0.99$. The NN always runs for 200 epochs unless there is a NaN in the loss calculation, in which case the training terminates. The output of the neural network is defined as the reconstructed m_{reco}.

The standard input variables that are used in the RNN are listed in tab. 8.1. They include the transverse momentum p_T, the azimuthal angle ϕ and the pseudorapidity η of the leading tau τ_0 as well as that of the subleading tau τ_1. With these quantities the visible mass of the event can be reconstructed. Some information on the neutrino momentum can be extracted from the missing transverse energy vector, from which the resulting transverse momentum p_T^{MET} and the pseudorapity ϕ^{MET} are used. As discussed chapter 3.2.1, the pseudorapidity of the MET can not be determined. In addition to those four-vector quantities, the number of reconstructed charged tracks $n_{\tau 0}^{charged}$ and $n_{\tau 1}^{charged}$ that are assigned to the two taus are included. They are relevant since the kinematics of the tau decay depend on the decaymode of the tau lepton. In total, there are 10 variables in the input variable set. The distributions of the input variables are shown in the appendix A for the five samples.

τ_0				τ_1				MET	
$\tau_0^{p_T}$	τ_0^{η}	τ_0^{ϕ}	n_{τ_0}	$\tau_1^{p_T}$	τ_1^{η}	τ_1^{ϕ}	n_{τ_1}	MET	ϕ_{MET}

Table 8.1: Input variables of the RNN.

8.2 Sample Splitting

The three samples used for the training of the NNs are split into three parts: the training set, the testing set and the validation set, all of which cover the entire mass range of the sample. The 160 GeV range of the samples is split into 1600 slices, each with a size of 0.1 GeV. The slice with the fewest events n_{min} is used as a bottleneck for the splitting. The first $0.1 \times n_{min}$ (rounded down) events of each slice are assigned to the testing set. The following $0.1 \times n_{min}$ events are assigned to the validation set and all other events are part of the training sample. This method has a flat mass distribution for the testing and validation sets, so they don't have to be reweighted. For the spin0hh sample there are not enough events in each slice for this method, so the sample is divided into 160 slices instead. The Z sample and H sample are not split and instead only used for validation purposes.

8.3 Methods of NN evaluation

Evaluating the performance of a neural network at reconstructing the ditau mass is not trivial. There are multiple metrics that give insight into the performance of a neural network. They can not be considered independently but instead complement each other and must all be taken into account. In this thesis three different methods of evaluation are used. First, the loss is considered. A low loss means that the neural network is good at what it was explicitly told to do, which is to minimize the loss function. It is important to consider both the loss of the training set and the loss of the testing set, as a large disparity between the two can reveal overtraining or underlying issues in the sample.

In the second method used to evaluate the performance of a NN, the relative resolution of the mass reconstruction is considered. For this the quantity

$$\Delta m_{rel} = \frac{m_{true} - m_{recon}}{m_{recon}} \tag{8.53}$$

is evaluated and shown in bins of the true mass. A relatively smooth distribution of the relative mass resolution for the entire mass range of the sample is a satisfactory result of the NN reconstruction.

In the third method discussed in this thesis, the performance of the neural network is evaluated by applying it to the Higgs sample and Z sample in order to create a ROC curve (receiver operating characteristic curve). The aim of this method is to have good H and Z boson reconstructions and to maximize the area under the curve (AUC) of the ROC curve.

8.4 RNN Stability

In this section the general stability of the learning process is studied. Prior work has shown that the loss often spikes and has inconsistent final values. To study the randomness of this, the seed of TensorFlow is varied. In addition, the possibility for a bias in the sample when splitting into testing,

training and validation is also studied. For this, the sample is split differently by assigning different parts of the slice to the three categories.

Some loss graphs for the γ^* training sample are shown in fig. 8.1. The first plot is a very common loss graph of the training process. The training loss quickly drops to under 25, where it slowly drops further and then stabilizes at about 24.6. While the training loss is rather smooth, with only some bumps, the testing loss is very erratic and has small spikes every few epochs, before also slowly stabilizing in the course of the training. The other three plots shown in fig. 8.1 are handpicked to show problematic irregularities. In the second plot there is a sudden sharp increase of the training loss at about 126 epochs into the training. This behavior continues for about 25 epochs, doubling the loss of the training, before it slowly starts falling off again. Since the training rate has already been halved multiple times by that point, the remaining epochs are not enough to lower the loss to its minimum and the training ends suboptimally. The testing loss remains unaffected during the entire process. While the cause of this type of phenomena cannot be determined with certainty, it probably stems from a very small phase space of events for which the derivative of some weights in the network grow to extreme values. In this specific case, the spike can be attributed to one single event. The third plot shows a somewhat typical loss plot in which the training takes a long time and both the training and the testing loss show erratic behavior. After about 130 epochs the loss eventually stabilizes and doesn't peak anymore. In the fourth plot the training seems to get stuck at an elevated value, before suddenly dropping sharply at about 23 epochs into the training to a lower value where it then slowly stabilizes. The cause for falling into an early plateau can not be determined.

The effects of halving the learning rate α can be seen as little steps in the training loss of the first plot from the epochs 25-125 and in the fourth plot from epochs 50-125. Here, the training loss starts growing very slowly after reaching a local minimum, while the high learning rate keeps the loss becoming smaller. Once the learning rate is halved, the loss manages to continue to drop again.

In general the loss is usually stable by the time it reaches 200 epochs, which is why a training length of 200 is used in this thesis. Early stopping conditions are also not used since the training sometimes plateaus temporarily, during which the learning would end early if early stopping conditions are used.

Similar example plots for the tauspinner sample are shown in fig. 8.2. For this sample the training process seems to be more stable. The loss value converges to a range of 22-23 every time and no large peaks or other significant irregularities are visible.

The loss plots for the spin0hh sample are shown in 8.3. Here, the training is also significantly more stable than the γ^* sample but it does show more spiking than the tauspinner sample. The training loss consistently converges to values 25-27 and the testing loss reaches values of 27-29. This divergence between the testing and training sets might be a result of the fact that the spin0hh sample is the smallest of the three samples and it also has the largest bias to larger masses.

To analyze the stability of the training process further, the testing loss of 180 neural networks after 40 epochs is shown in fig. 8.4. The x-axis shows the random seed of TensorFlow and the y-axis shows the input permutation used for the training, testing and xvalidation split. In this plot, the γ^* sample is used. While the loss is stable and reaches values between 24 and 26 for most neural networks, there are three outliers above 27. The same plot after 80 epochs is shown in fig. 8.5. Here, there are only two outliers left and most losses have dropped slightly. Fig. 8.6 shows the losses after 200 epochs. The overall loss of most neural networks has continued to drop slightly, but the two outliers remain at a high loss and a new outlier spikes and reaches a loss of over 29. The same plot after 200 epochs for the training sample is shown in fig. 8.7. Here, five losses over 26 can be found. The outliers overlap with the outliers of the testing sample, but also include two new outliers that are not visible in the testing

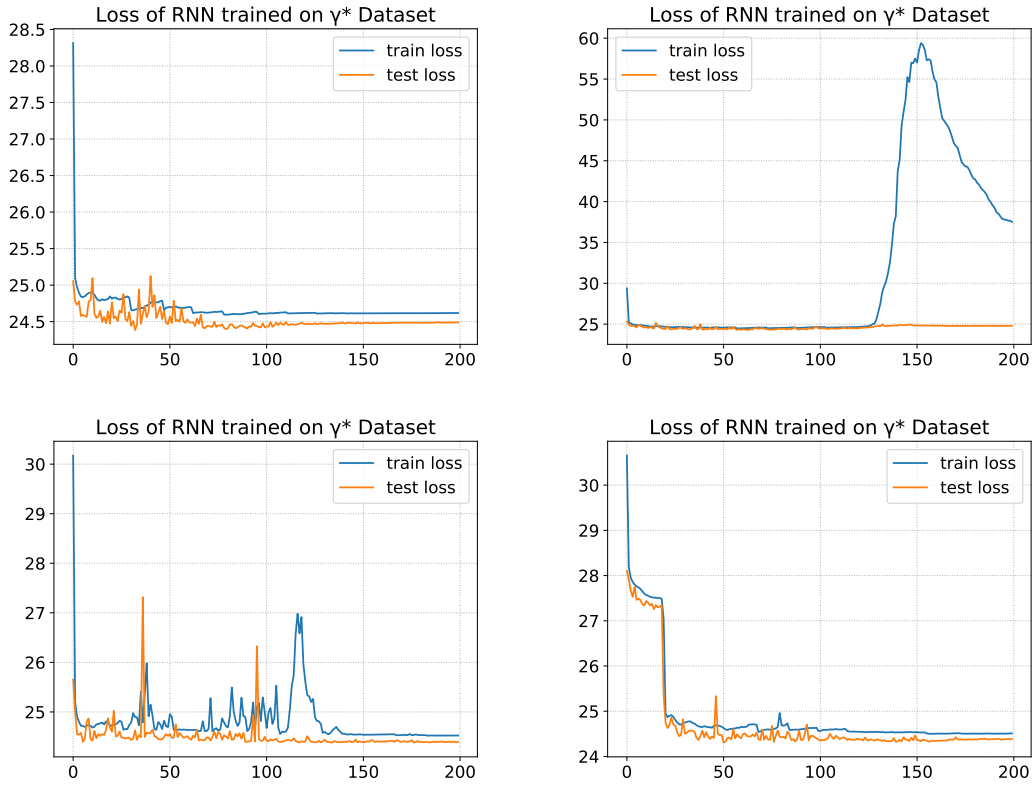


Figure 8.1: Four selected loss graphs of a neural network trained on the γ^* sample. The training and testing loss is shown.

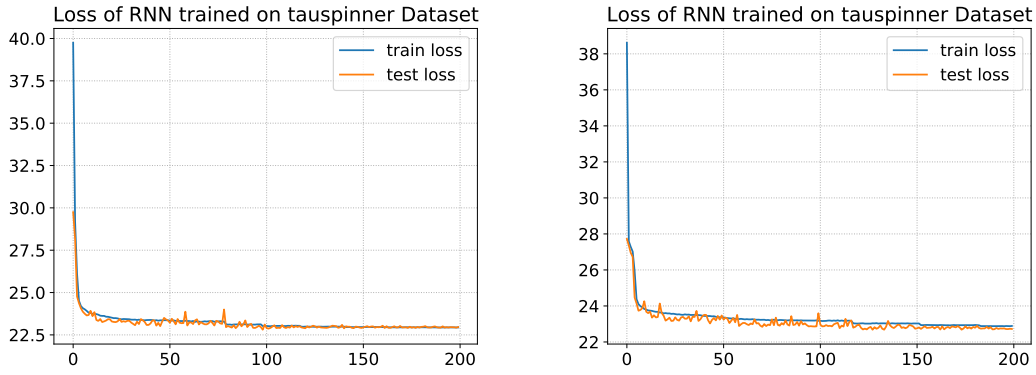


Figure 8.2: Two selected loss graphs of a neural network trained on the tauspinner sample. Training and testing loss is shown.

loss. Neither the permutation of the input sample split nor the random TensorFlow seed seem to be particularly relevant to the optimization of the loss function. Instead, it seems to be an effect based on the combined random nature of all aspects of the neural network with no specific underlying structure.

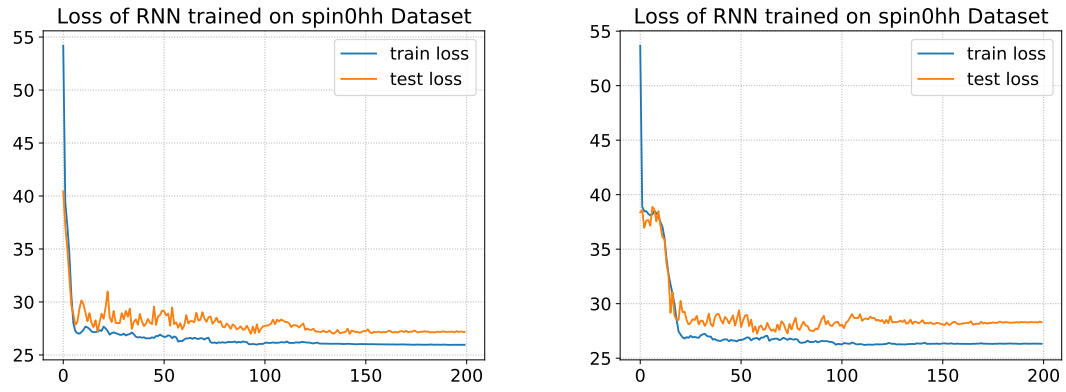


Figure 8.3: Two selected loss graphs of a neural network trained on the spin0hh sample. Training and testing loss is shown.

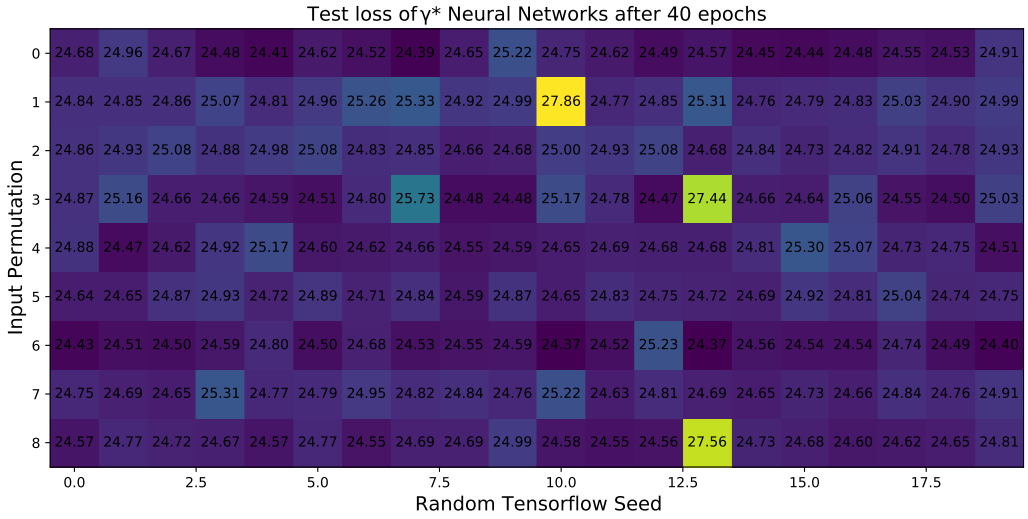


Figure 8.4: Test loss after 40 epochs of 180 neural networks trained on the γ^* sample with varied random TensorFlow seed and test-train split.

A histogram of the testing losses for the γ^* after 200 epochs is shown in fig. 8.8. The mean of the loss is $\mu = 24.63$ with a standard deviation of $\sigma = 0.49$. The same plot for the tauspinner sample is shown in fig. 8.9 and for the spin0hh sample in fig. 8.10. For the tauspinner sample the mean loss is $\mu = 22.70$ with a standard deviation of $\sigma = 0.50$. The mean loss of the spin0hh sample is $\mu = 26.40$ with a standard deviation of $\sigma = 0.78$. As already discussed, the training process seems more stable for the tauspinner and spin0hh samples. While the loss for tauspinner sample seems to be generally lower than for the spin0hh sample and γ^* sample, this does not mean that these neural networks are better. Due to the different shapes of the Monte Carlo samples, the loss value itself cannot be compared between different setups.

In general the training seems to stabilize after 200 epochs around the same value. Therefore the

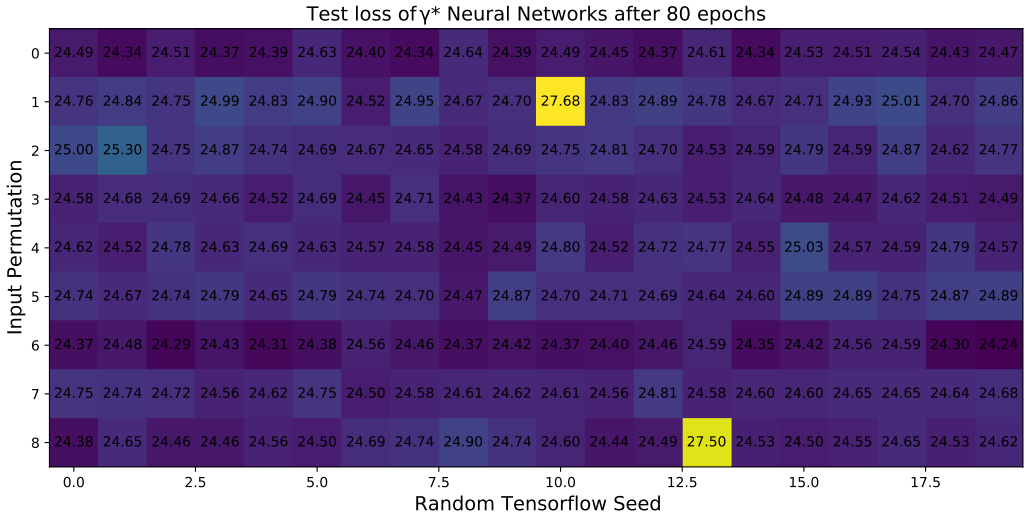


Figure 8.5: Test loss after 80 epochs of 180 neural networks trained on the γ^* sample with varied random TensorFlow seed and test-train split.

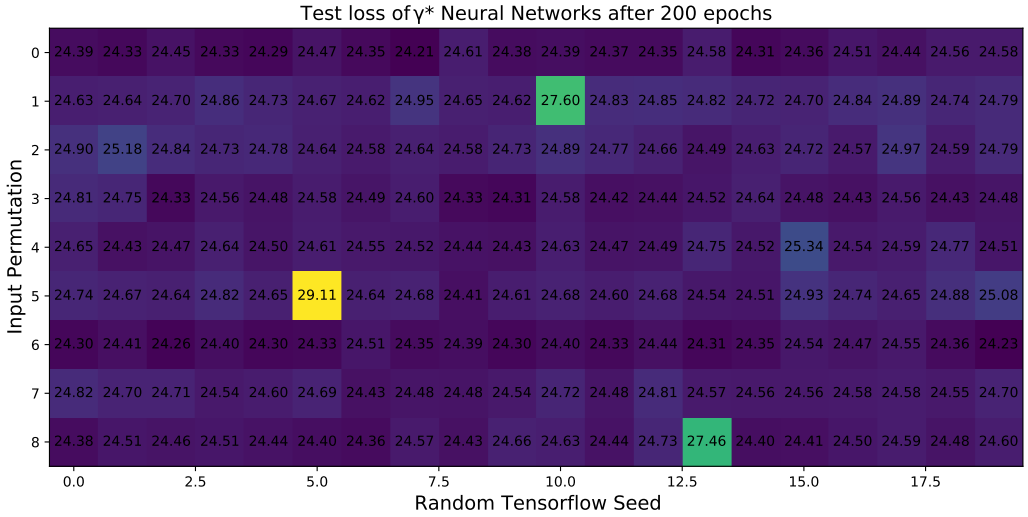


Figure 8.6: Test loss after 200 epochs of 180 neural networks trained on the γ^* sample with varied random TensorFlow seed and test-train split.

neural networks are trained for 200 epochs in this thesis. Since there is still a chance that the training is still unstable after 200 epochs or that it stabilizes at a high value, three or more neural networks are always trained simultaneously. They are then checked for consistency.

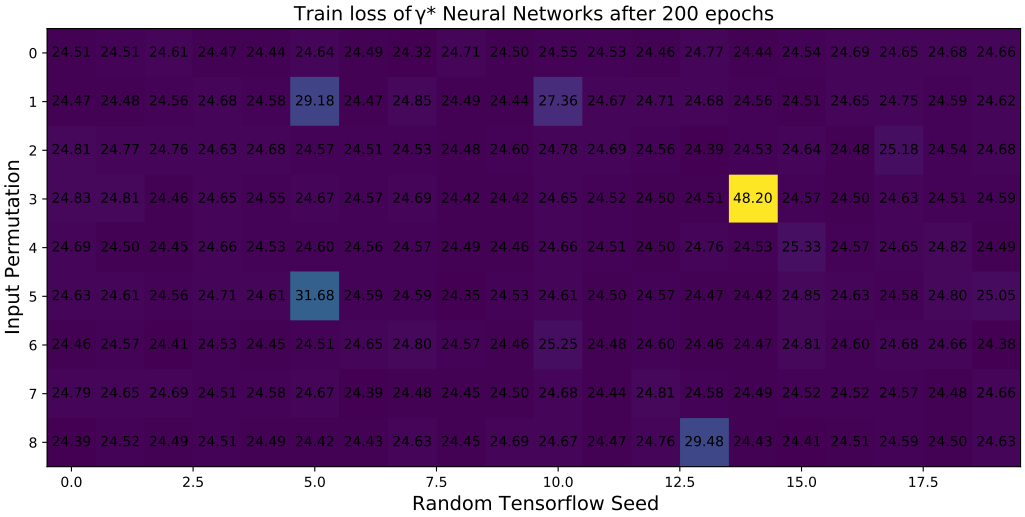


Figure 8.7: Train loss after 200 epochs of 180 neural networks trained on the γ^* sample with varied random TensorFlow seed and test-train split.

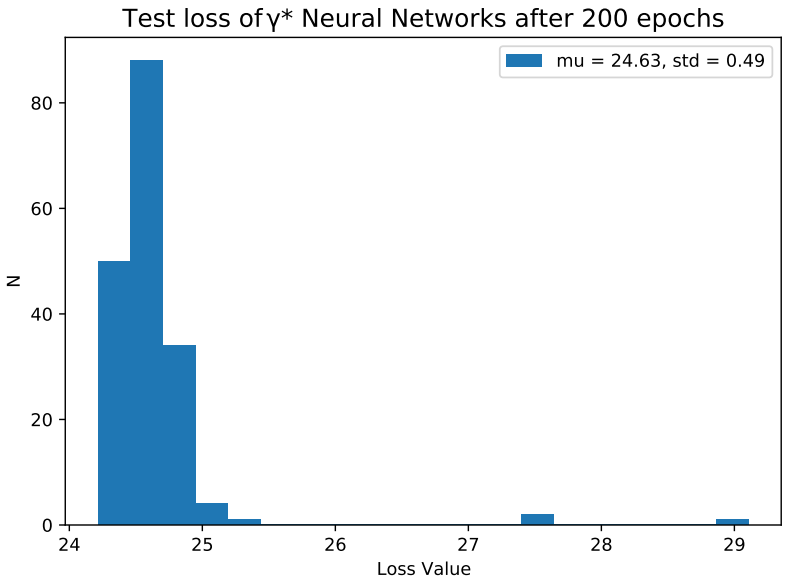


Figure 8.8: Histogram of test loss after 200 epochs of 180 neural networks trained on the γ^* sample with varied random TensorFlow seed and test-train split.

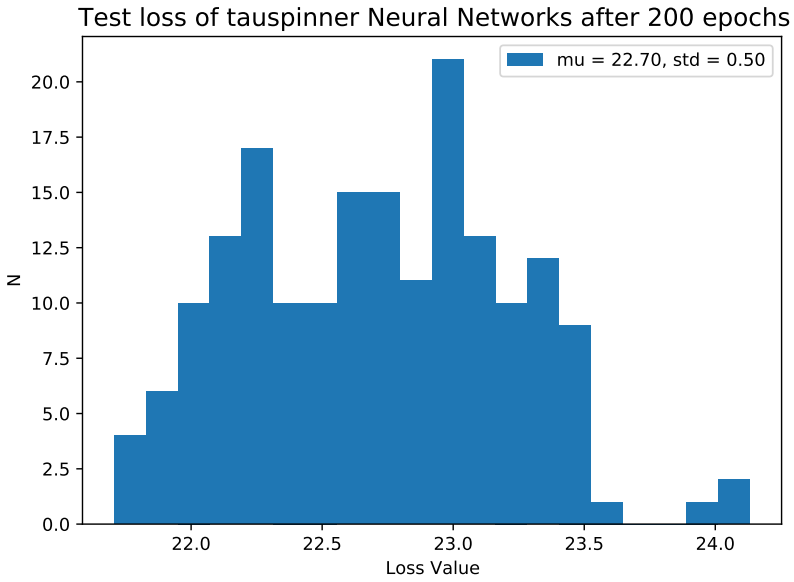


Figure 8.9: Histogram of test loss after 200 epochs of 180 neural networks trained on the tauspinner sample with varied random TensorFlow seed and test-train split.

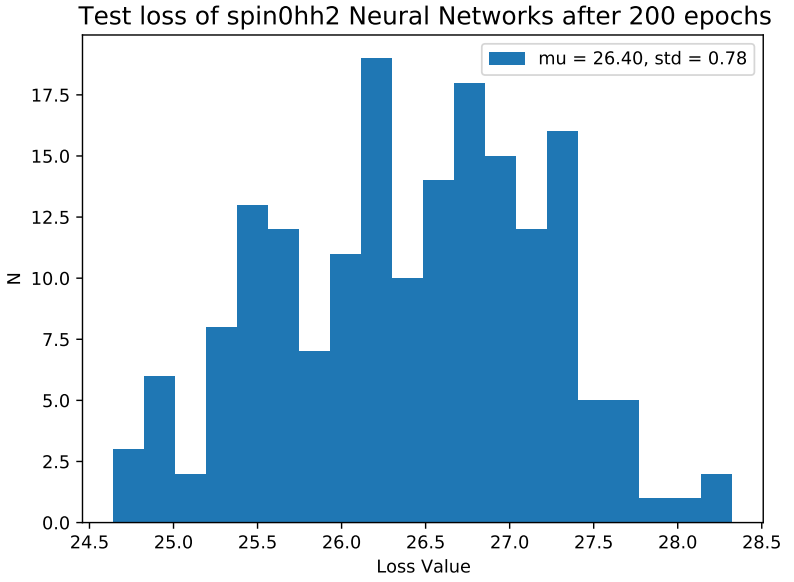


Figure 8.10: Histogram of test loss after 200 epochs of 180 neural networks trained on the spin0hh sample with varied random TensorFlow seed and test-train split.

8.5 Reconstruction Performance

In this section the performance of the reconstruction of neural networks trained on the three training samples are analyzed. For this, the relative resolution of the reconstructed masses is used. In addition, a calibration method is introduced and implemented. The neural network reconstruction of the Higgs sample and the Z sample and their separation power between those two samples are also presented. The target of the RNN reconstruction in this section is the true ditau mass m_{true} of the event.

8.5.1 γ^* Sample Reconstruction

The relative mass resolution of a neural network trained on the γ^* sample is shown in fig. 8.11 (left). It has a mean value of $\mu = -0.070$ and a standard deviation of $\sigma = 0.188$. 8.11 (right) shows the entire mass distribution of the RNN reconstruction. Despite the large standard deviation of the relative resolution of the reconstruction, the reconstruction itself is constrained in the $60 - 220\,\text{GeV}$ mass range of the training sample. Fig. 8.12 shows the relative mass resolution (left) and the mass distribution (right) of the MMC ditau mass reconstruction. The distribution has a mean value of $\mu = -0.070$ and a standard deviation of $\sigma = 0.188$. Therefore, the RNN has a lower mean value standard deviation for the relative resolution than the MMC reconstruction on this sample. The mean of both the RNN and MMC reconstruction in bins of the true mass are shown in fig. 8.13. The error bars show the 68% quantiles of each bin. The total bias b is calculated as

$$b = \frac{\sqrt{\sum_{i=1}^{n_b} b_i^2}}{n_b} \tag{8.54}$$

where b_i is the individual bias of the i-th bin and is equivalent to the mean of the relative mass resolution of that bin. $n_b = 16$ stands for the number of bins. For a RNN trained on the γ^*-sample the total bias is $b = 0.170$ and for the MMC reconstruction on the γ^*-sample the total bias is $b = 0.143$. The mean 68% quantile μ_b is calculated similarly as

$$\mu_b = \frac{\sqrt{\sum_{i=1}^{n_b} \left(\frac{l_{err,i}+h_{err,i}}{2}\right)^2}}{n_b} \tag{8.55}$$

where $l_{err,i}$ is the width between the mean and the lower 16% quantile of each bin content and $h_{err,i}$ is the width between the mean and the upper 16% quantile of each bin. For this RNN the mean 68% quantile width is calculated to be $\mu_b = 0.195$ and for the MMC reconstruction $\mu_b = 0.210$.

The 68% and 95% quantile widths of the relative mass resolution are shown in fig. 8.14. At very low masses the RNN and the MMC have very similar 68% quantile widths: Both are at $\sigma \approx 0.15$. The 68% quantile width of the neural network then quickly rises with higher masses before reaching a maximum at $90 - 100\,\text{GeV}$ with $\sigma \approx 0.18$. It then shrinks almost linearly and reaches a minimum at $210 - 220\,\text{GeV}$ with $\sigma \approx 0.09$. The MMC follows a different trend. At first the 68% quantile width stays almost constant is at its minimum value of $\sigma \approx 0.15$ at $70 - 80\,\text{GeV}$, before then continuously rising and reaching a maximum of $\sigma \approx 0.19$ at $210 - 220\,\text{GeV}$. The RNN starts having a smaller width than the MMC between $120 - 130\,\text{GeV}$. The 95% quantiles follow a similar trend.

There are two noteworthy features of these distributions. First , there is a very clear downward trend

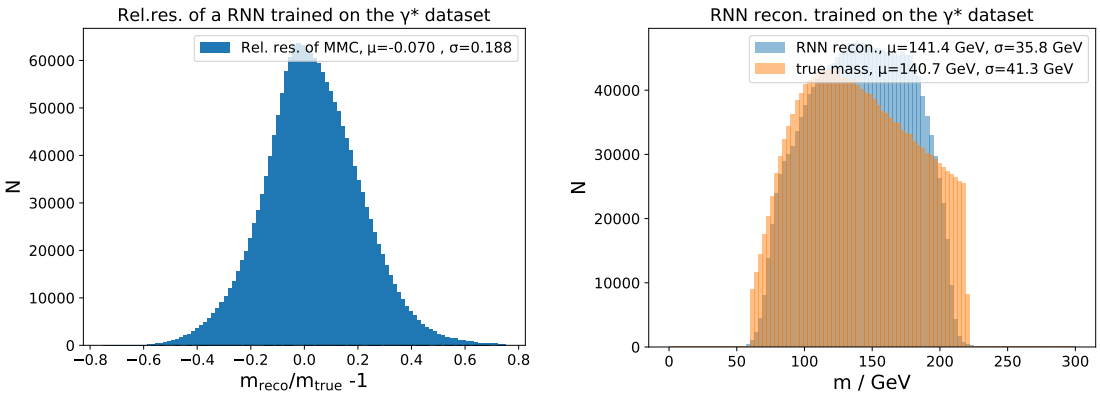


Figure 8.11: The relative mass resolution (left) and mass reconstruction as well as the true mass (right) of a neural network trained on the γ^* sample

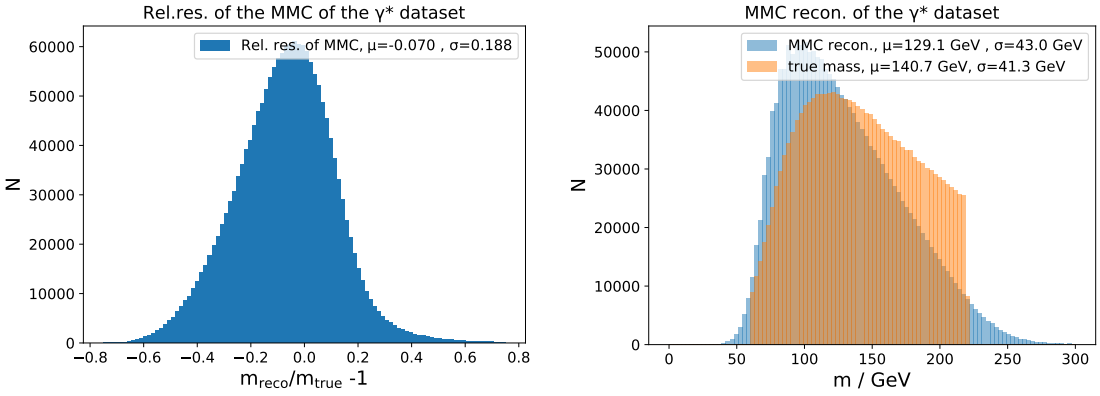


Figure 8.12: The relative mass resolution (left) and mass reconstruction (right) of the MMC reconstruction of the γ^* sample.

of the bias visible in the plot for the relative mass resolution of the neural network. There are very high positive biases for low masses, reaching $b_{max} \approx +0.22$, and very high negative biases for high masses, reaching $b_{min} \approx -0.14$. The MMC reconstruction also shows a downward trend in the bias though it is not as strong. For low masses it reaches up to $b_{max} \approx +0.05$ and it goes down to $b_{min} \approx -0.10$. For the RNN this is, in fact, to be expected since the network learns that the sample is constrained to $60 - 220\,\text{GeV}$ and therefore is biased to overestimate low mass events and to underestimate high mass events. This can be seen in fig. 8.15 and fig. 8.16, where the relative mass resolution for the lowest bin ($60 - 70\,\text{GeV}$) and highest bin ($210 - 220\,\text{GeV}$) are shown respectively. 8.15 shows that almost no events have their mass underestimated. The mean of the distribution is $\mu = 0.241$ with a standard deviation of $\sigma = 0.189$. In a similar manner, 8.16 shows that almost no events have their mass overestimated. The mean is $\mu = 0.138$ with a standard deviation of $\sigma = 0.103$. These distributions are also highly un-gaussian, which is not desired. The reconstructed mass distributions of those mass slices are also shown in fig. 8.15 (right) and fig. 8.16 (right). The mean of the $60\,\text{GeV} < m_{truth} < 70\,\text{GeV}$ slice is $\mu = 81.4\,\text{GeV}$ with a standard deviation of $\sigma = 12.7\,\text{GeV}$. This

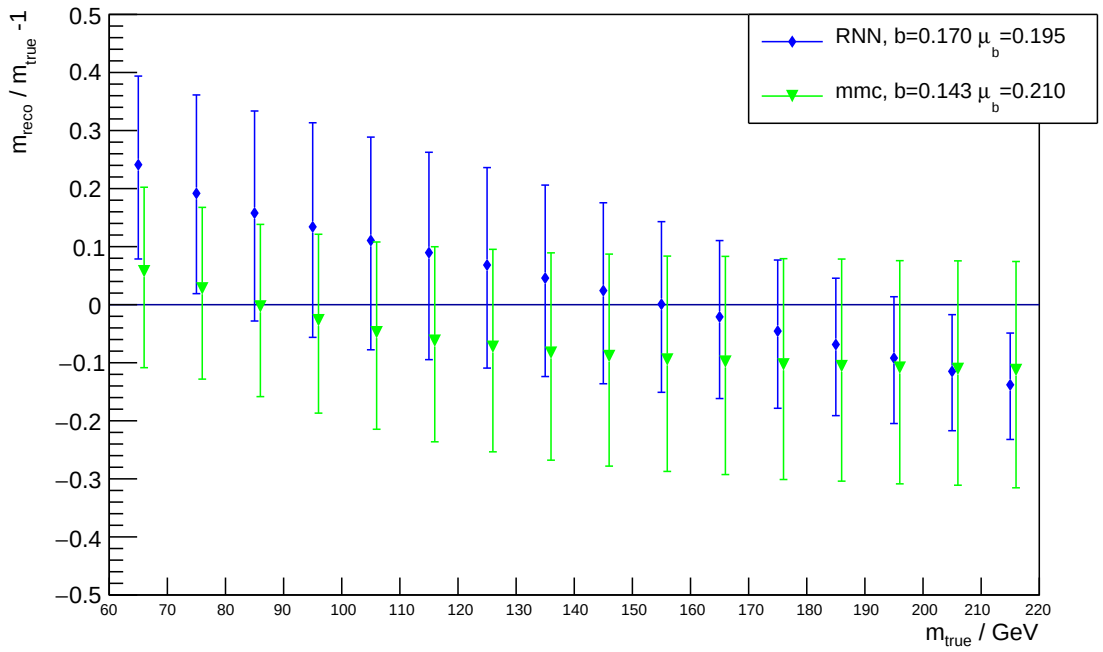


Figure 8.13: The mean relative mass resolution and the 68% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the γ^* and the MMC reconstruction of the γ^* sample.

distribution is highly overestimated with the mean lying above the actual mass slice by almost one standard deviation. For the 210 GeV $< m_{truth} <$ 220 GeV slice the mean lies at $\mu = 185.3$ GeV and has a standard deviation of $\sigma = 22.1$ GeV. This distribution is highly underestimated by over one standard deviation to the edge of the slice.

The same plots for the MMC reconstruction are shown in fig. 8.17 and fig. 8.18. The mass resolution distribution for 60 GeV $< m_{truth} <$ 70 GeV has a mean of $\mu = 0.057$ with a standard deviation of $\sigma = 0.191$. The distribution is also not Gaussian. There is a more pronounced tail to positive resolutions than to lower resolutions. The mass distribution has a mean of $\mu = 69.4$ GeV with a standard deviation of $\sigma = 13.1$ GeV. For the 210 GeV $< m_{truth} <$ 220 GeV slice, the mean of the relative mass resolution distribution is $\mu = -0.113$ with a standard deviation of $\sigma = 0.197$. This distribution is also not Gaussian. It has a sharp cutoff towards positive relative resolutions and a larger tail towards negative relative resolutions. The mass distribution of this slice has a mean of $\mu = 190.0$ GeV with a standard deviation of $\sigma = 41.0$ GeV. This is a significantly larger standard deviation than for the same mass slice for the RNN reconstruction and is consistent with the values plotted in fig. 8.13. Overall both the RNN and the MMC seem to be biased by overestimating low mass events and underestimating high mass events.

Also noteworthy is that the results shown in the plots do not allow for a direct comparison between the MMC and the neural network. The MMC has a larger μ-value but a smaller total bias b than the neural network. This is expected, since the RNN is inclined to develop a bias by learning the shape of the mass distribution of the training sample. A way to reduce the large bias and its slope

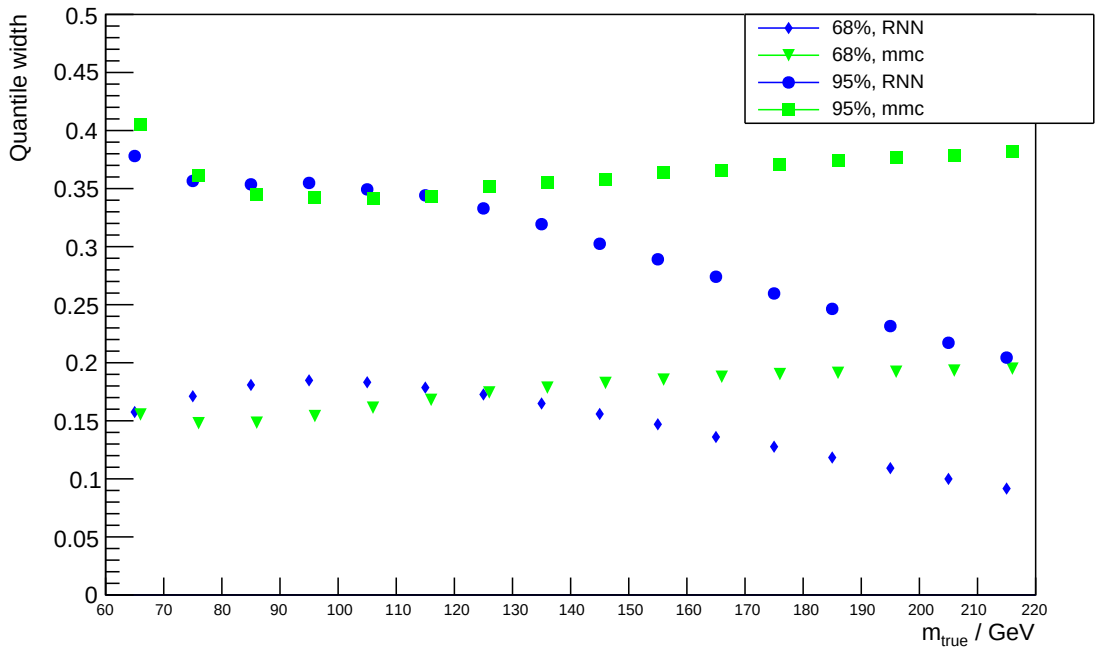


Figure 8.14: The 68% and 95% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the γ^* sample and the MMC reconstruction of the γ^* sample.

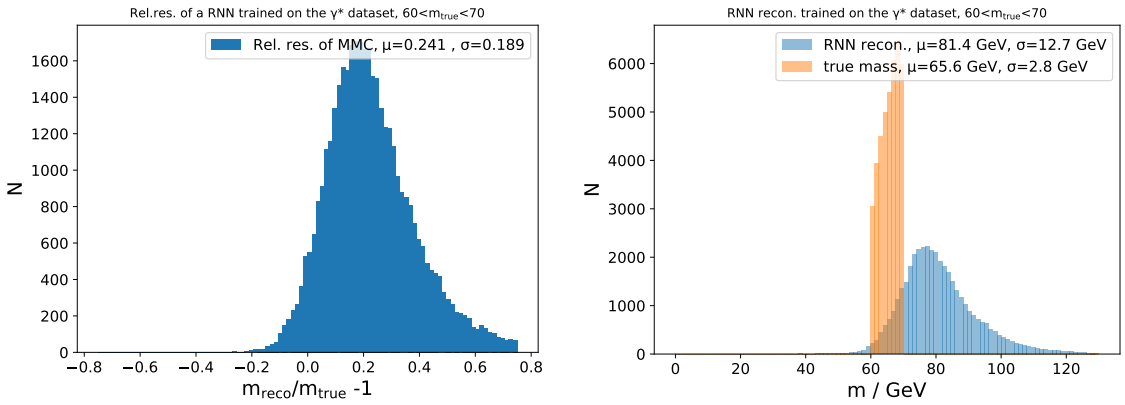


Figure 8.15: The relative mass resolution (left) and mass reconstruction as well as the true mass (right) of a neural network trained on the γ^* sample for true masses 60 GeV $< m_{true} <$ 70 GeV

visible in fig. 8.13 is by calibrating the reconstructed masses to the real masses. To make the results displayed in the plots comparable, an approach developed in [61] is applied. With this method, the mass distribution in bins of 10 GeV across the 60 – 220 GeV range is fitted to their correct value. Due to the very asymmetrical distributions, the mean of each bin is used for the fitting since it provides a compromise between the peak position and the large outliers. The mass distribution with the 68%

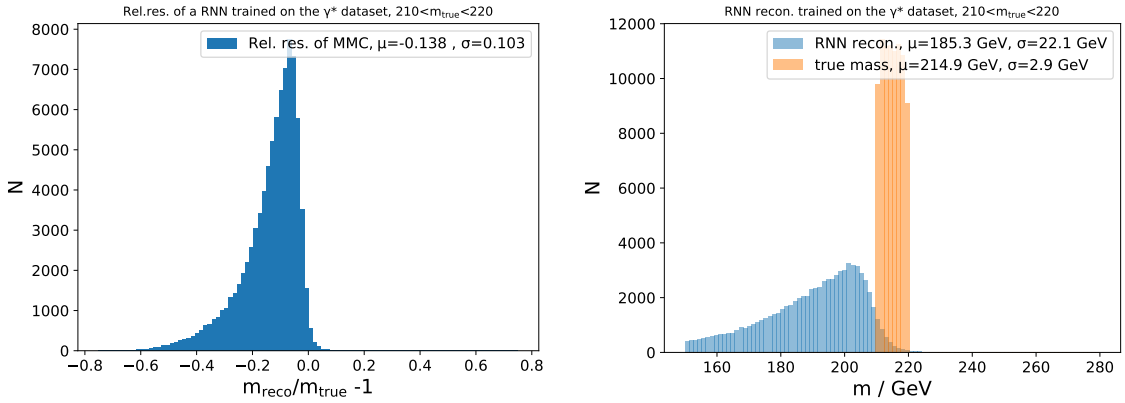


Figure 8.16: The relative mass resolution (left) and mass reconstruction as well as the true mass (right) of a neural network trained on the γ^* sample for true masses 210 GeV $< m_{true} <$ 220 GeV

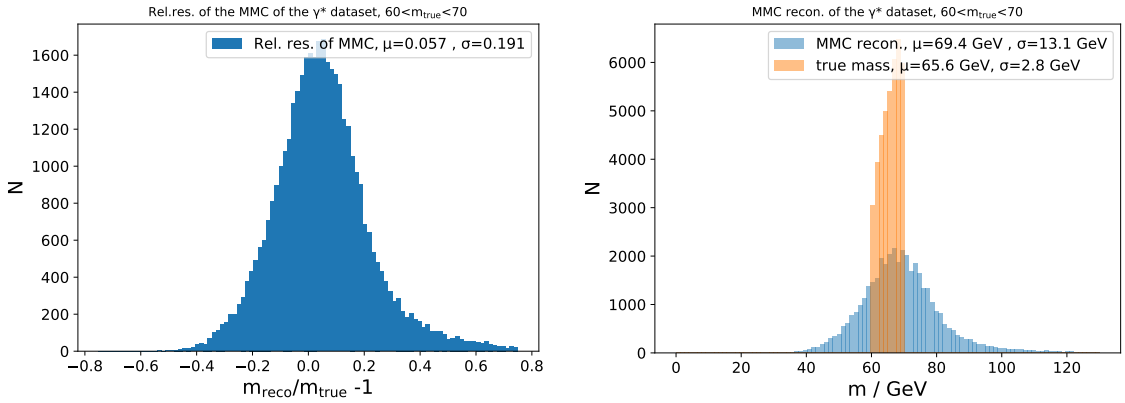


Figure 8.17: The relative mass resolution (left) and mass reconstruction (right) of the MMC reconstruction of the γ^* sample for true masses 60 GeV $< m_{true} <$ 70 GeV

quantiles in bins of 10 GeV is shown in fig. 8.19 (left: RNN, right: MMC). The blue line shows the ideal $m_{reco} = m_{truth}$ line that is fitted to. To fit a function to the medians of the bins, a polynomial of 4th degree is used:

$$m_{true} = a \cdot (m_{reco})^4 + b \cdot (m_{reco})^3 + c \cdot (m_{reco})^2 + d \cdot (m_{reco}) + e \tag{8.56}$$

The fit is also shown in the plot as the red line.

Applying the fit to all events leads to fig. 8.20. The The medians of the bins now lie approximately on the $m_{reco} = m_{truth}$ line, so that the bias of the plot is almost completely removed. However, the goal of the calibration is to remove the bias in bins of the true mass and not in bins of the reconstructed mass. Those two requirements are not equal because the asymmetrical mass distributions of each bin lead to bin-to-bin migration effects. The median values of the corrected values in bins of true masses is shown in fig. 8.21. A new bias from the correction is now visible, but it can be approximated with a

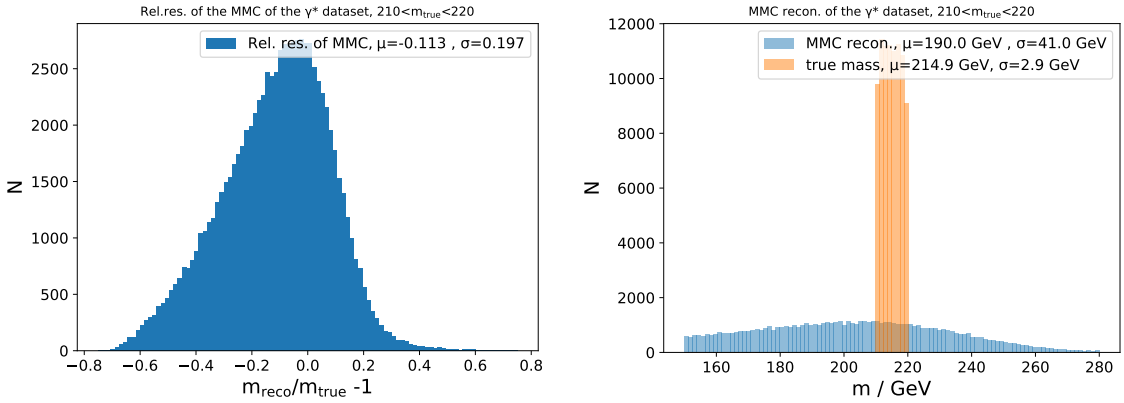


Figure 8.18: The relative mass resolution (left) and mass reconstruction (right) of the MMC reconstruction of the γ^* sample for true masses 210 GeV $< m_{true} <$ 220 GeV

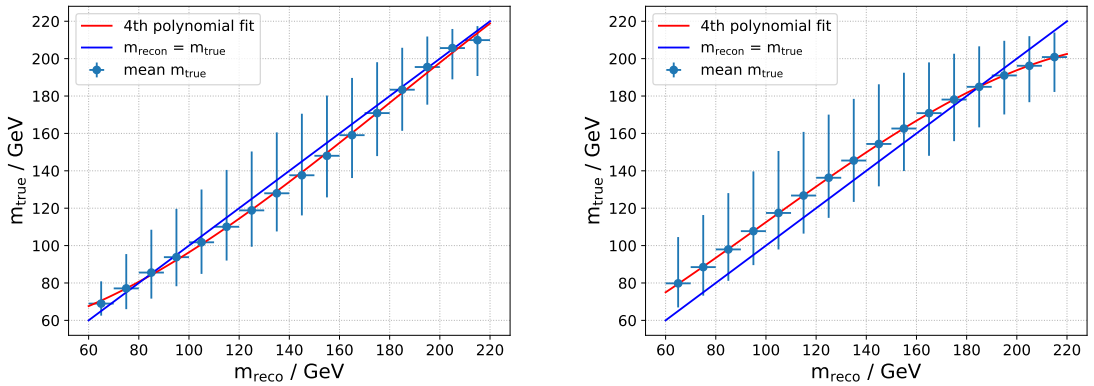


Figure 8.19: The median and 68% quantiles of the reconstructed mass in bins of the true mass for the RNN (left) and MMC(right) reconstruction.

linear function of the form:

$$m_{reco} = p \cdot m_{true} + t \tag{8.57}$$

$$\Leftrightarrow m_{true} = \frac{m_{reco} - t}{p}. \tag{8.58}$$

Since the true masses are usually not known, the function has to be inverted so that the function can be applied to the corrected mass $m_{\tau\tau}^{corrected}$. The final result is shown in fig. 8.22. The bias is mostly gone. As shown, this method can be applied to both the RNN reconstruction and the MMC reconstructions. The relative resolution plot of the calibrated reconstruction in bins of m_{truth} is shown in fig. 8.23 and the 68% and 95% quantiles are shown fig. 8.24. The new plot shows that the bias is mostly gone, but the 68% quantile widths have increased. This is an expected effect since the entire mass reconstruction has been stretched out to remove the bias of the mean of each bin. The new plot also allows for a comparison between the RNN and MMC reconstruction. The total bias is $b_{RNN} = 0.086$ and $b_{MMC} = 0.085$ for the RNN and MMC reconstruction respectively and the

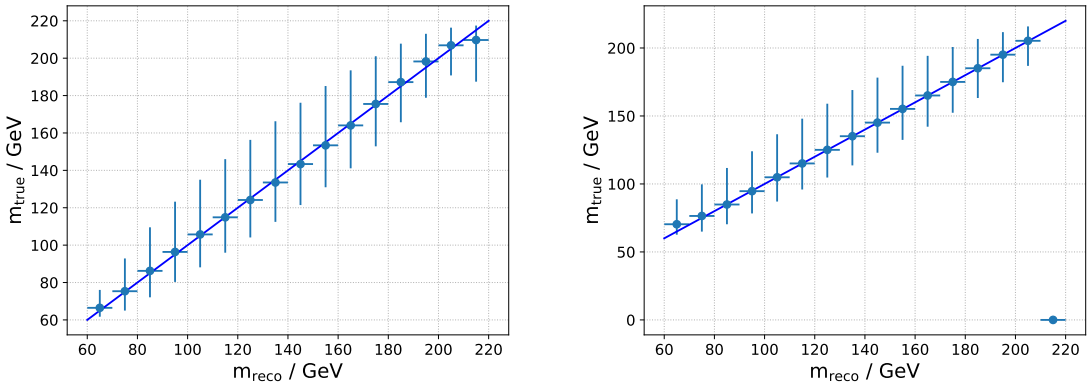


Figure 8.20: The median and 68% quantiles of the reconstructed mass in bins of the true mass for the RNN (left) and MMC (right) reconstruction after the first calibration step.

mean 68% quantile width is $\mu_b^{RNN} = 0.219$ and $\mu_b^{MMC} = 0.225$ respectively. While the bias is nearly zero for the bins of $m_{true} < 180\,\text{GeV}$, for higher masses the bias becomes significant again. The 68% quantile width of the RNN is slightly higher for the bins in the mass range $80\,\text{GeV} < m_{true} < 130\,\text{GeV}$ than for the MMC and lower in the other mass bins. The 95% quantile widths are lower for the RNN than for the MMC across the entire mass range. The difference in the 95% quantile width is the most significant for the lowest bins.

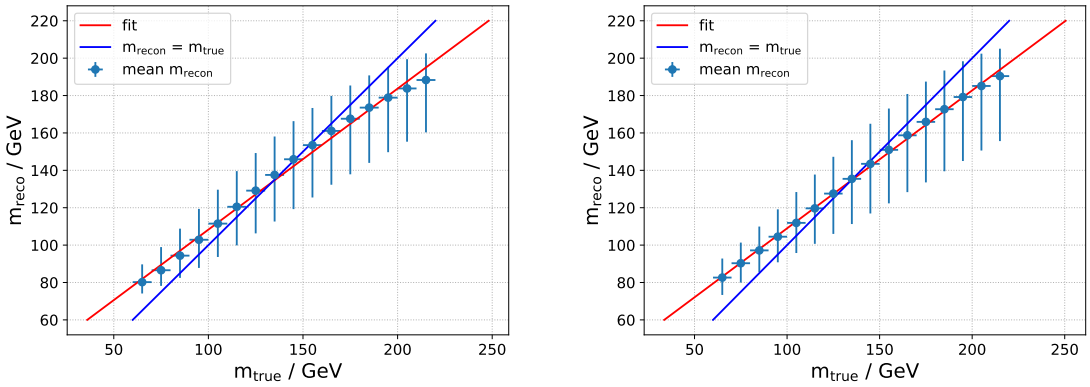


Figure 8.21: The median and 68% quantiles of the true mass in bins of the reconstructed mass for the RNN (left) and MMC (right) reconstruction after the first calibration step.

The final method of evaluating the neural networks presented in this thesis is a receiver operating characteristic (ROC) curve. The Higgs and Z samples are used to measure the separation power of the RNN and MMC. Since the samples generated by Monte Carlo are based on Standard Model Higgs and Z bosons, and since the cuts applied to these samples are generally used in analysis, this method gives a useful comparison between the RNN and the MMC reconstruction. The ROC curve for an RNN trained on the γ^* sample is shown in fig. 8.25. The area under the curve (AUC) is 0.875 for the RNN and 0.908 for the MMC, which demonstrates that the MMC is better able to separate Higgs

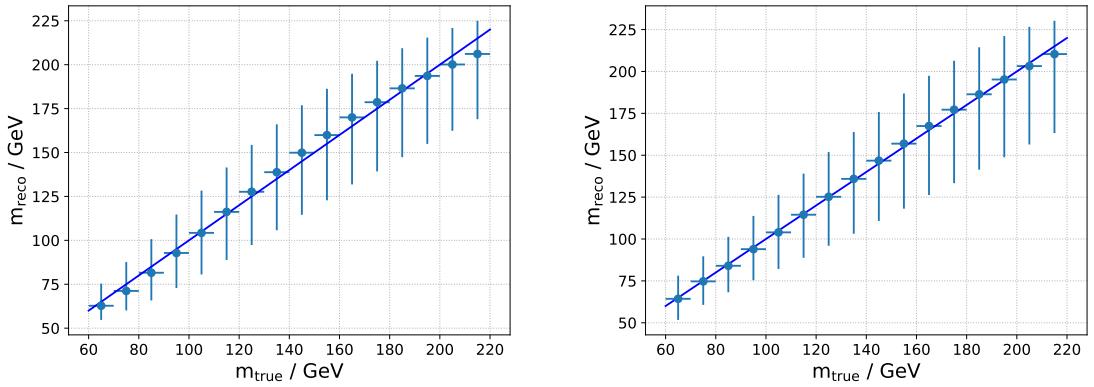


Figure 8.22: The median and 68% quantiles of the true mass in bins of the reconstructed mass for the RNN (left) and MMC (right) reconstruction after the second calibration step.

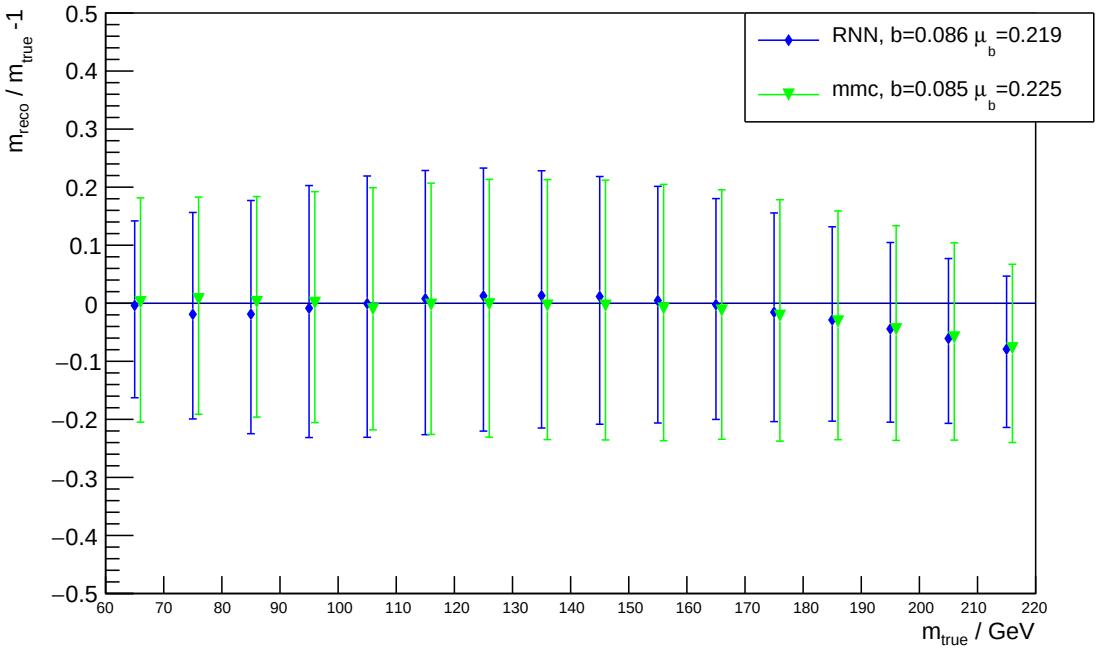


Figure 8.23: The mean relative mass resolution and the 68% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the γ^* sample and the MMC reconstruction of the γ^* sample after the calibration.

and Z event. This is because the cuts applied to the Higgs and Z samples appear to benefit the MMC reconstruction more than the RNN reconstruction.

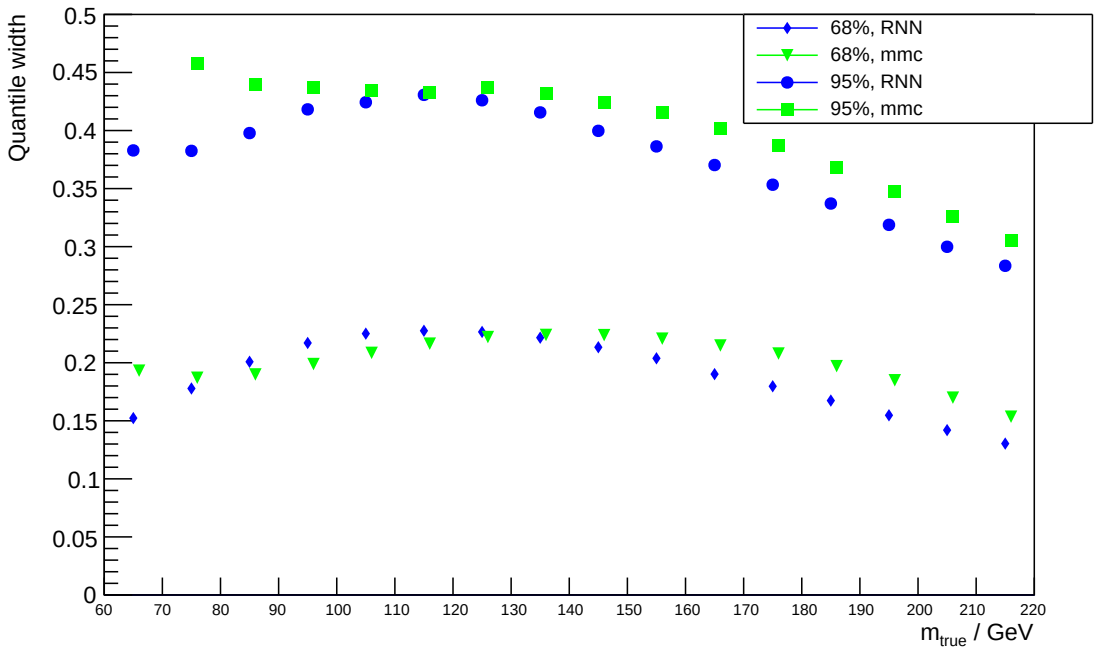


Figure 8.24: The 68% and 95% quantile width of the relative mass resolution in bins of the true ditau-mass of a RNN trained on the γ^* sample and the MMC reconstruction of the γ^* sample after the calibration.

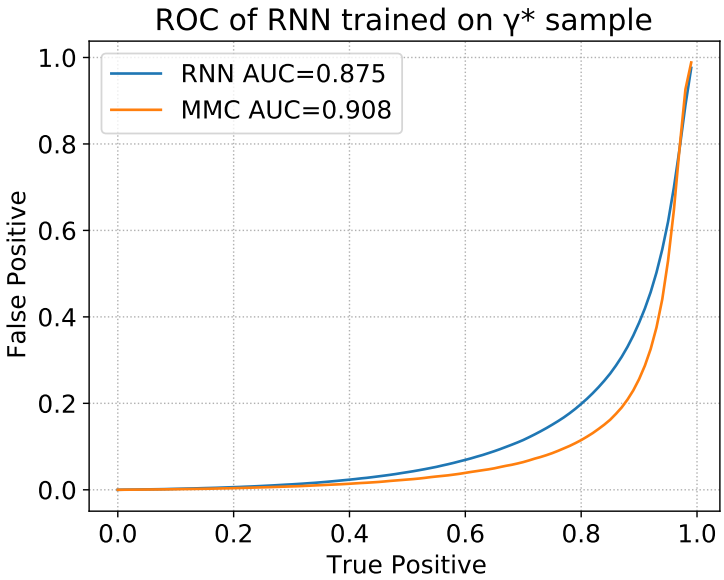


Figure 8.25: A ROC curve of the separation power between the H sample and the Z sample for the RNN trained on the γ^* sample and the MMC reconstruction.

8.5.2 Tauspinner Sample Reconstruction

A similar analysis is done for the tauspinner sample. The relative mass resolution of an RNN trained on the tauspinner sample sample is shown in fig. 8.26 (left). The distribution has a mean value of $\mu = -0.027$ with a standard deviation of $\sigma = 0.161$. This is a significant improvement to the same mass resolution distribution of the γ^* sample. The mass distribution of the tauspinner sample reconstruction is shown in fig. 8.26 (right). It has a thick short tail for small masses (50 GeV – 100 GeV) and reaches a sharp peak at $m_{recon} \approx 190$ GeV. This is significantly different from the mass distribution of the γ^* sample, which looks much more symmetrical. The underlying reason for this is could be the original mass distribution of the sample, which is much more biased to higher masses for the tauspinner sample compared to the γ^* sample (see fig. 7.1 and fig. 7.3).

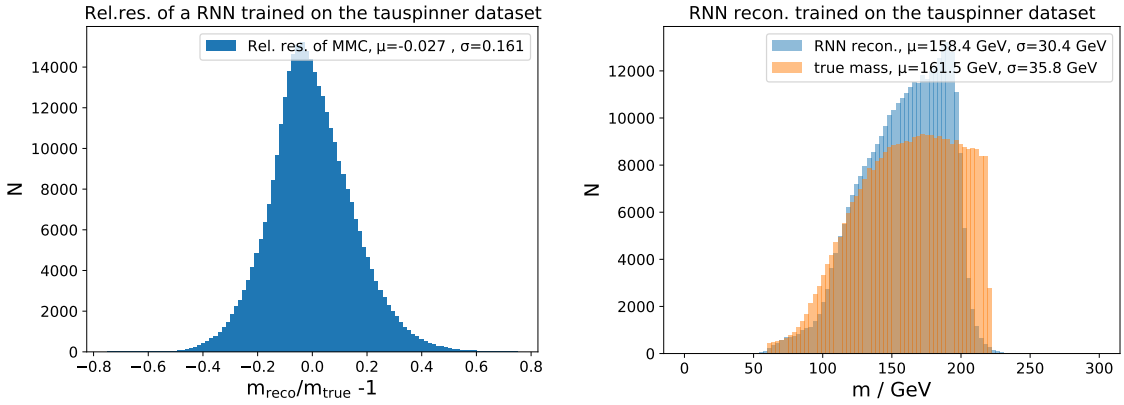


Figure 8.26: The relative mass resolution (left) and mass reconstruction (right) of a neural network trained on the tauspinner sample.

The relative mass resolution and mass distribution of the MMC reconstruction of the tauspinner sample is shown in fig. 8.27. The relative resolution of the MMC has a mean of $\mu = -0.027$ and a standard deviation of $\sigma = 0.161$. This is also significantly lower compared to the MMC reconstruction of the γ^* sample. More cuts are applied to the tauspinner sample than to the γ^* sample. These cuts lead to an improved reconstruction for both the RNN and the MMC, which is why the relative mass resolution distribution of the MMC reconstruction is narrower for the tauspinner sample than for the γ sample.

The relative mass resolution for both the RNN and MMC reconstruction in bins of the true mass are shown in fig. 8.28. The total bias for the RNN reconstruction is $b = 0.156$ with a mean 68% quantile width of $\mu_b = 0.187$. The MMC reconstruction has a total bias of $b = 0.133$ and a mean 68% quantile width of $\mu_b = 0.192$. A significant difference to the MMC and RNN reconstruction of the γ^*-sample is that the bin bias b_i starts rising in the lower bins for both the MMC and RNN reconstruction. The bias starts at $b_i = 0.10$ for the RNN and rises to a maximum of $b_i = 0.15$ at 90 – 100 GeV. It then continuously falls and reaches $b_i \approx -0.13$ in the highest bin. The MMC reconstruction follows a similar trend. It starts with $b_i \approx 0.04$ and reaches a maximum of $b_i = 0.12$ in the 80 – 90 GeV and 90 – 100 GeV bins. Then it falls and reaches negative values, reaching a minimum value of $b_i \approx -0.09$ at the highest bin.

Fig. 8.29 and fig. 8.30 show the relative mass resolution (left) and the mass distribution (right) of

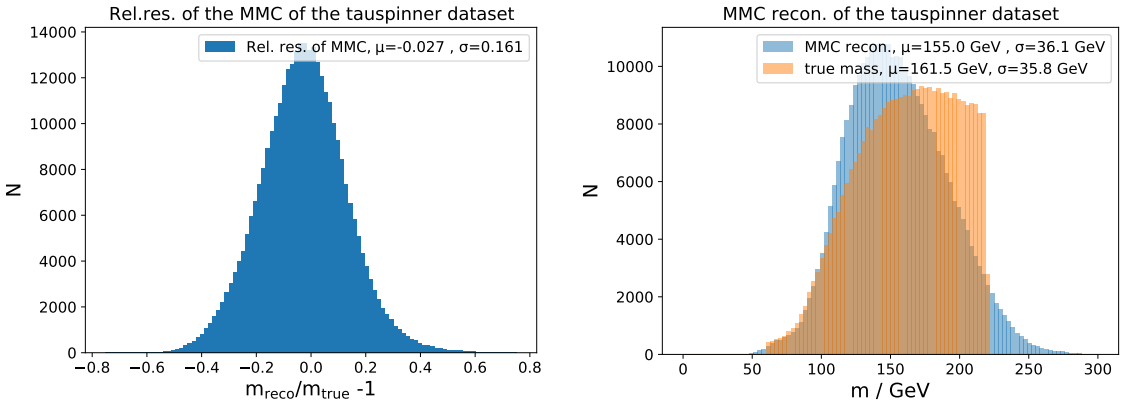


Figure 8.27: The relative mass resolution (left) and mass reconstruction (right) of the MMC reconstruction of the tauspinner sample.

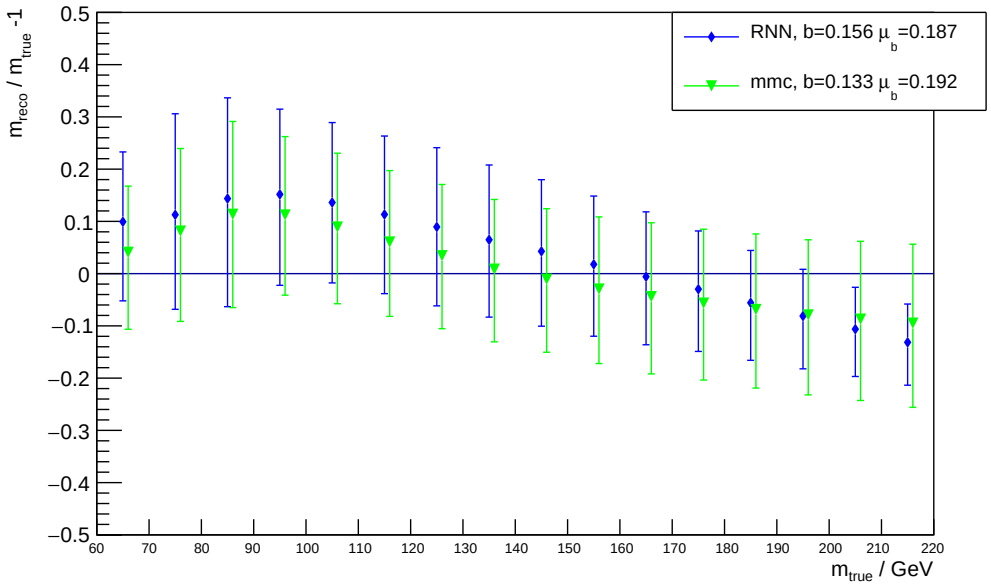


Figure 8.28: The mean relative mass resolution and the 68% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the tauspinner sample and the MMC reconstruction of the tauspinner sample.

the lowest and highest mass bins for the RNN reconstruction. The low statistics of the lower mass regions of the tauspinner sample create a very uneven distribution for the 60 GeV $< m_{truth} <$ 70 GeV slice. The mean value of the relative mass resolution for that slice is $\mu = 0.099$ ($\sigma = 0.204$) and the mean value of the mass distribution is $\mu = 71.8$ GeV ($\sigma = 13.8$ GeV). For the highest mass slice (210 GeV $< m_{truth} <$ 220 GeV) the mean relative resolution is $\mu = -0.131$ ($\sigma = 0.089$), and the mean of the mass distribution is $\mu = 186.7$ GeV ($\sigma = 19.1$ GeV). The distributions are again highly skewed

into one direction and represent that the RNN learns the underlying shape of the mass distribution of the tauspinner sample.

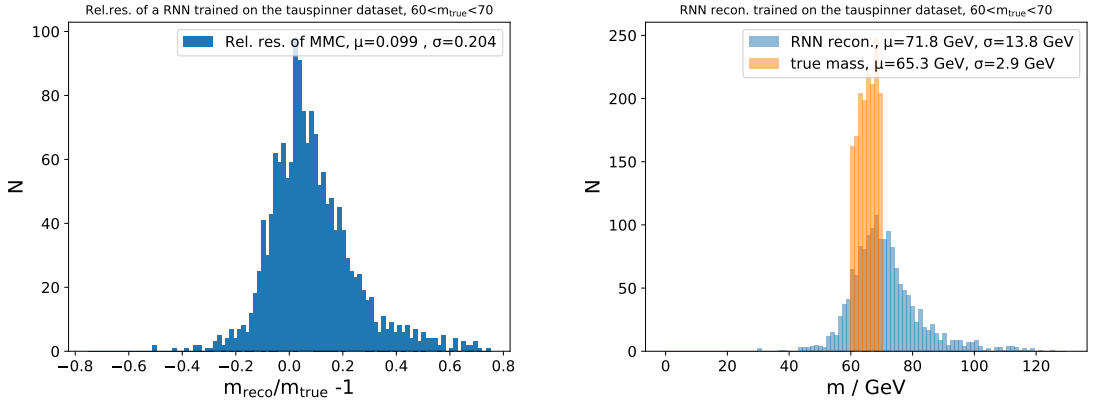


Figure 8.29: The relative mass resolution (left) and mass reconstruction (right) of a neural network trained on the γ^* sample for true masses 60 GeV $< m_{true} <$ 70 GeV

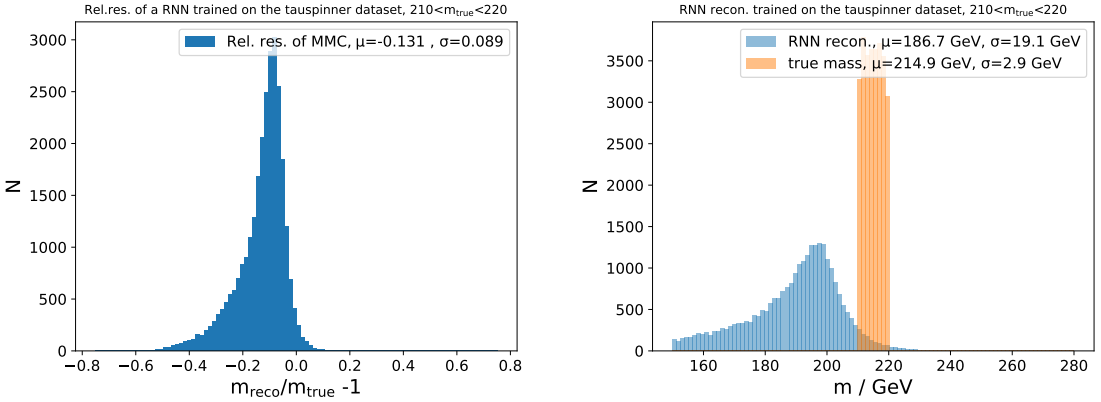


Figure 8.30: The relative mass resolution (left) and mass reconstruction (right) of a neural network trained on the γ^* sample for true masses 210 GeV $< m_{true} <$ 220 GeV

The calibration to remove the bias was also applied to this sample. The calibration plots are shown in fig. 8.31. In the second calibration step, a linear function as used for the γ^* sample is not sufficient for the fit. Instead, a logarithmic function of the form

$$m_{\tau\tau}^{corrected} = p \ln(m_{\tau\tau}^{truth}) + t \Rightarrow m_{\tau\tau}^{truth} = \exp(\frac{m_{\tau\tau}^{corrected} - t}{p}) \tag{8.59}$$

is used, and the inverse of it is applied to the data points. This modification is necessary because the tauspinner sample is a lot more asymmetric in its mass distribution, which worsens the bin-to-bin migration effects. A quadratic polynomial can also be used, however the logarithm has the benefit that it is a surjective function, so that its inverse is always defined. The logarithmic calibration results in

the successful removal of most of the bias. The result of the calibration can be seen in fig. 8.32. Most of the bias is thereby removed, although a larger bias remains than for the γ^* sample reconstruction. In addition to the negative biases at high masses, a positive bias is present in the region of 80 – 120 GeV for both the RNN and MMC reconstruction. For the lowest bin, the MMC has a bias of $b_i \approx 0.3$, which might be due to the low statistics for low masses.

The 68% and 95% quantiles are shown in fig. 8.33. For low masses $m_{truth} < 110$ GeV and for high masses $m_{truth} > 170$ GeV, the RNN has lower 68% quantile widths than the MMC reconstruction. The 95% quantile widths are larger for the RNN than for the MMC in the 60 – 70 GeV bin and in the mass region of 100 GeV to 120 GeV.

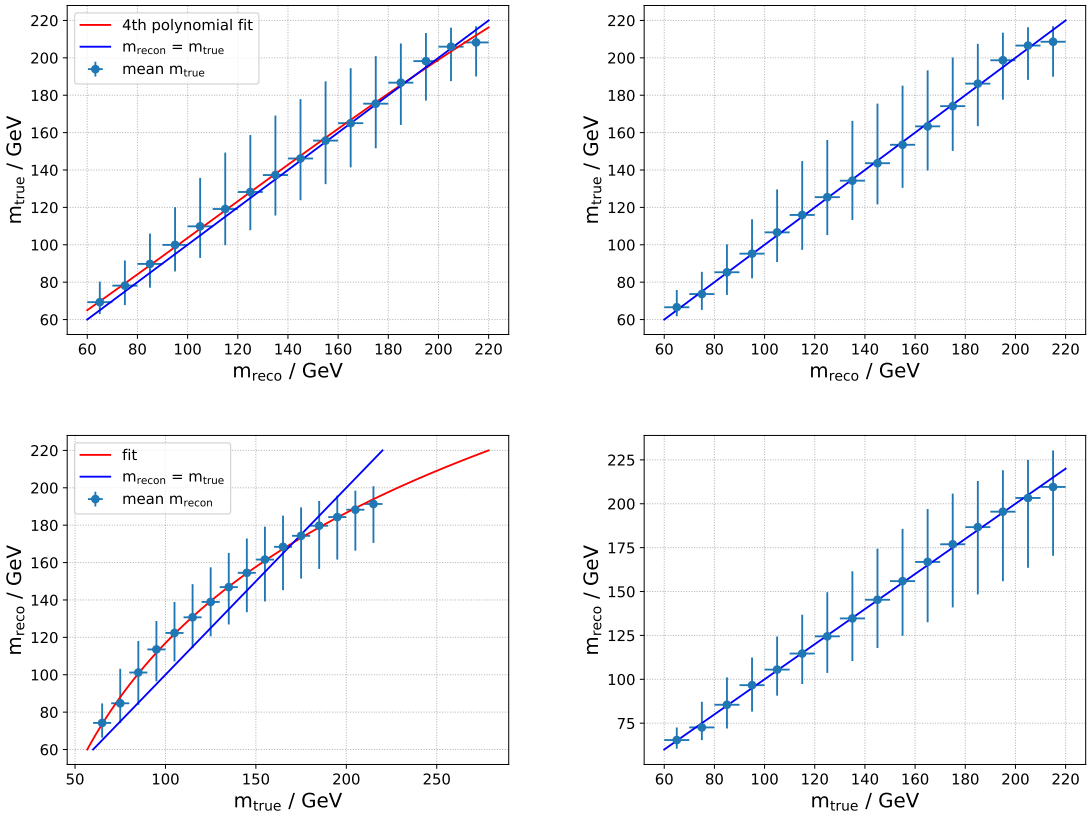


Figure 8.31: The median and 68% quantiles of the true mass in bins of the reconstructed mass for the RNN reconstruction before (top left) and after (top right) the first calibration step, as well as the median and 68% quantiles of the reconstructed mass in bins of the true mass for the RNN reconstruction before (bottom left) and after (bottom right) the second calibration step.

The ROC curves for an RNN trained on the tauspinner sample and the MMC applied to the Higgs and Z samples, are shown in fig. 8.34. The AUC of the RNN is 0.887, which is an improvement to the ROC curve of the RNN trained on the γ^* sample. Therefore, the emulation of the parity violation in the tauspinner sample is beneficial to the RNN and has a positive effect on the separating power. Nevertheless, the MMC still has an advantage over RNNs for the task of separating Higgs and Z events

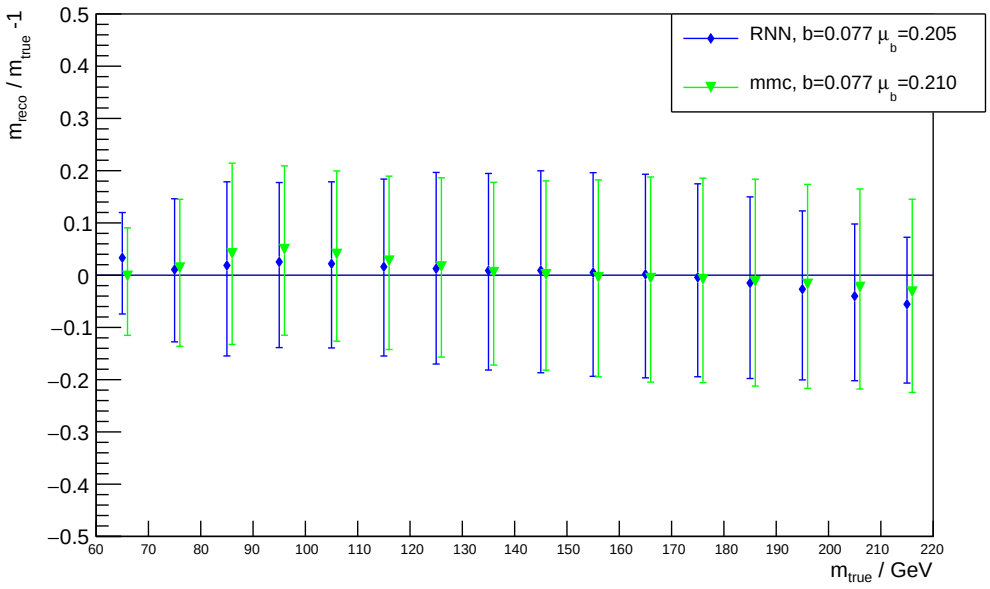


Figure 8.32: The mean relative mass resolution and the 68% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the tauspinner sample and the MMC reconstruction of the tauspinner sample after the calibration.

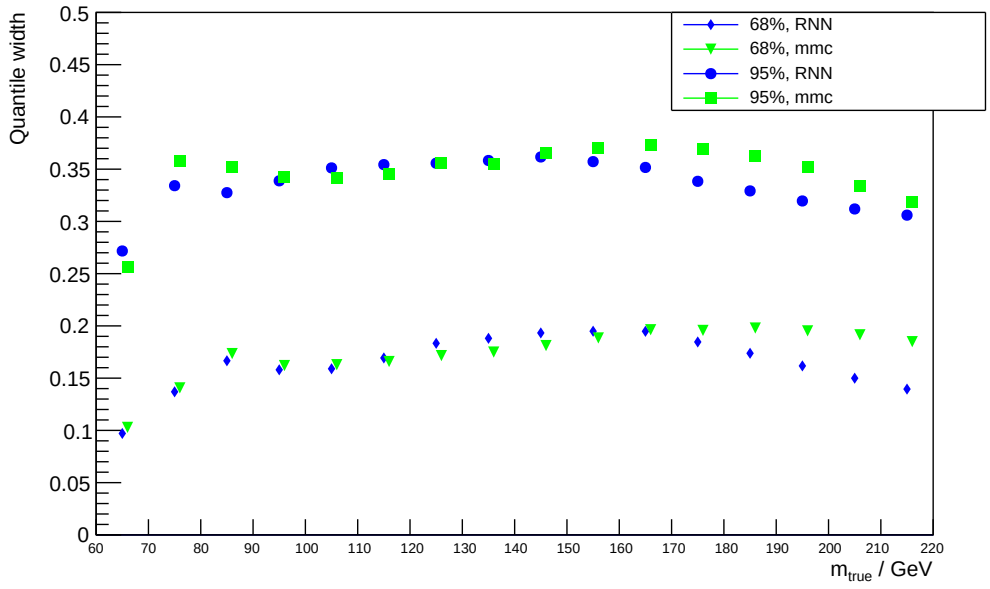


Figure 8.33: The 68% and 95% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the tauspinner sample and the MMC-reconstruction of the tauspinner sample after the calibration.

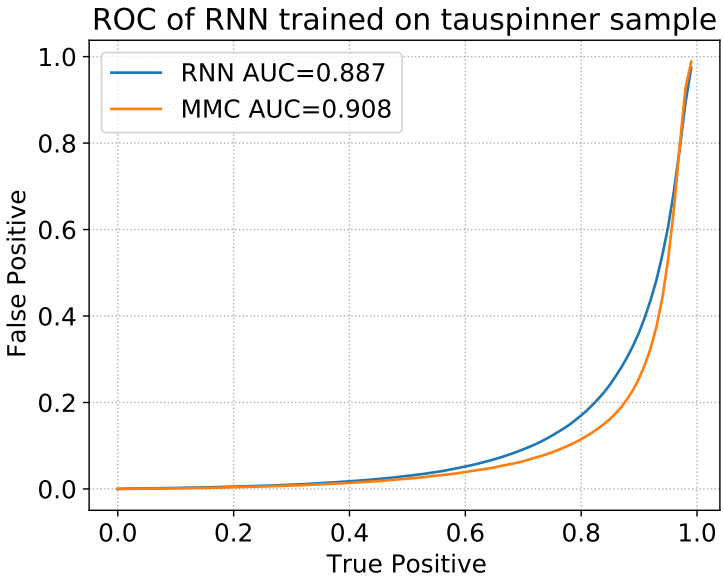


Figure 8.34: A ROCcurve of the separation power between the H sample and the Z sample for the RNN trained on the tauspinner sample and the MMC reconstruction.

8.5.3 Spin0hh Sample Reconstruction

In this section the performance of a neural network trained on the spin0hh sample is discussed. The relative mass resolution of a neural network trained on this sample is shown in fig. 8.35 (left). The distribution has a mean value of $\mu = -0.050$ and a standard deviation of $\sigma = 0.170$. The mean value is negative because the spin0hh sample is more biased to high masses than the other two training samples and the mass of high mass events is more likely to be underestimated in the reconstruction. The standard deviation is similar to the standard deviation of the tauspinner sample.

The relative mass resolution of the MMC reconstruction on the spin0hh sample is shown in fig. 8.36 (left). The distribution has a mean value of $\mu = -0.050$ with a standard deviation of $\sigma = 0.170$.

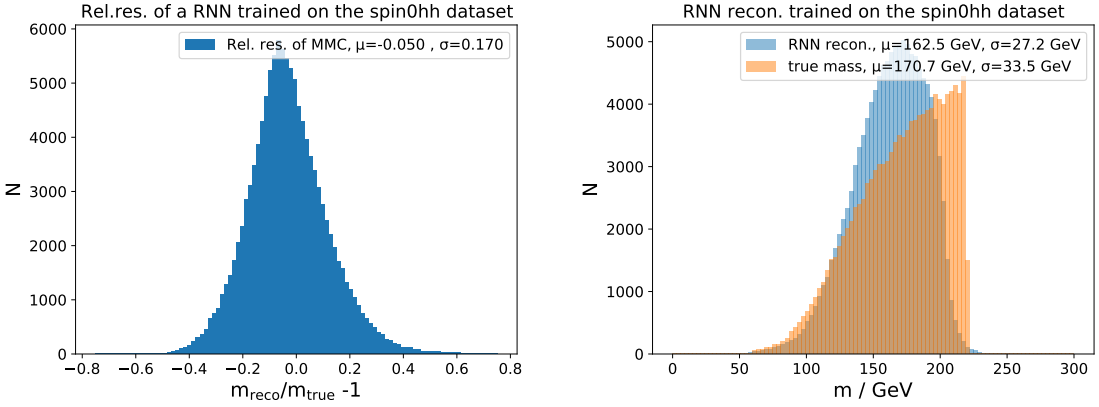


Figure 8.35: The relative mass resolution (left) and mass reconstruction (right) of a neural network trained on the spin0hh sample.

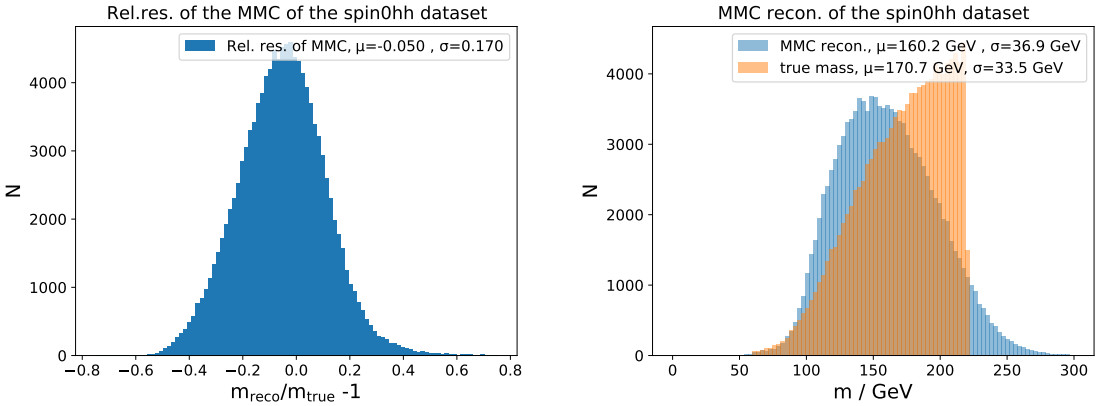


Figure 8.36: The relative mass resolution (left) and mass reconstruction (right) of the MMC reconstruction of the spin0hh sample.

The relative mass resolution in bins of true mass is shown in fig. 8.37. The plot is similar to the other two samples, but especially for low masses the 68% quantile width is very high for the RNN reconstruction. The maximum bias of the RNN reconstruction is achieved in the lower bins and has a

value of $b_{max} = 0.20$, it then continuously drops and reaches a minimum of $b_{min} = -0.15$. The MMC reconstruction has a maximum bias of $b_{max} = 0.14$ and then also continuously falls and reaches a minimum bias of $b_{min} = -0.10$.

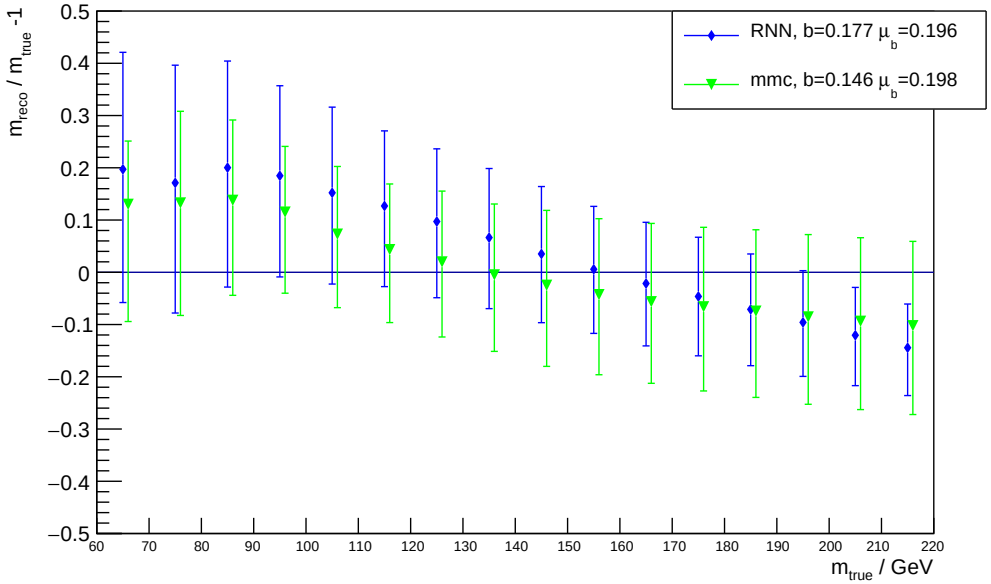


Figure 8.37: The mean relative mass resolution and the 68% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the tauspinner sample and the MMC reconstruction of the tauspinner sample.

The relative mass resolution for the lowest and highest bins are shown on the left in fig. 8.38 and 8.39. The statistics for the spin0hh sample are even lower than for the tauspinner sample so the derived value from their distributions are not accurate. For the lowest mass bin the mean relative mass resolution is computed to be $\mu = 0.099$ with a standard deviation of $\sigma = 0.204$ and for the highest mass bin it is computed to be $\mu = -0.144$ with a standard deviation of 0.091. Once again the bins on the extreme ends of the true mass spectrum show a large bias towards reconstructions that point towards masses that lie in the range of the sample ($60\,\text{GeV} < m_{recon} < 220\,\text{GeV}$).

The calibration plots for this neural network reconstruction can be seen in fig. 8.40. The first calibration step is successful. For the second calibration step, a logarithm is necessary for the fit again, since the bin-to-bin effects are similar to the bin-to-bin effects for the tauspinner sample, when changing from the binned m_{recon} plot to the binned m_{truth} plot.

The binned relative mass resolution after the calibration is shown in fig. 8.41. Most of the bias has been removed, though the remaining bias $b = 0.101$ is more significant than for the other samples for both the RNN and MMC reconstruction. The 68% and 95% quantiles are shown in fig. 8.42. While the 68% quantile width is relatively stable between 0.14 and 0.21, the 95% quantiles reach values over 0.5 for $m_{true} < 90\,\text{GeV}$. Since this is the case for both the RNN and the MMC reconstruction, this effect is not because of the low statistics for the RNN training, but instead an effect of the sample itself. The only low mass events that pass the cuts in the sample might populate a subset of the phase-space that is difficult for neural networks and for the MMC to reconstruct.

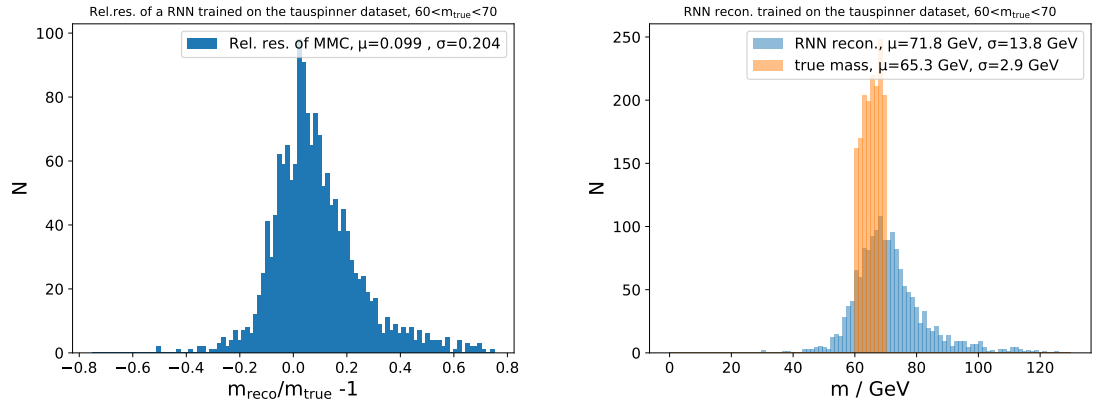


Figure 8.38: The relative mass resolution (left) and mass reconstruction (right) of a neural network trained on the spin0hh sample for true masses 60 GeV $< m_{true} <$ 70 GeV

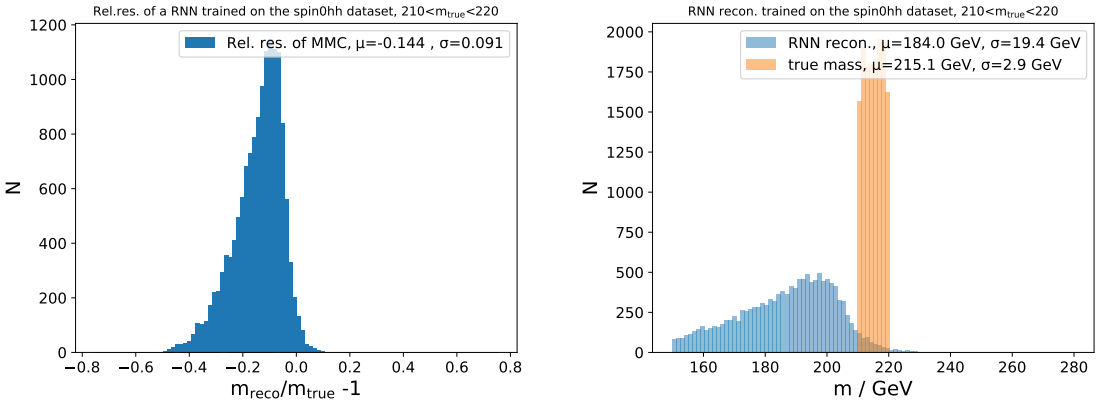


Figure 8.39: The relative mass resolution (left) and mass reconstruction (right) of a neural network trained on the spin0hh sample for true masses 210 GeV $< m_{true} <$ 220 GeV

The ROC curves for an RNN trained on the spin0hh sample and the MMC, when applied to the Higgs and Z samples, are shown in fig. 8.43. The AUC of the RNN is 0.862. This is a deterioration in separating power compared to both the γ^* and tauspinner sample. The low statistics, especially for low masses and the resulting decline in the mass resolution, might be the reason for this. RNNs trained on the spin0hh sample is not able to better separate between Higgs and Z events than the MMC.

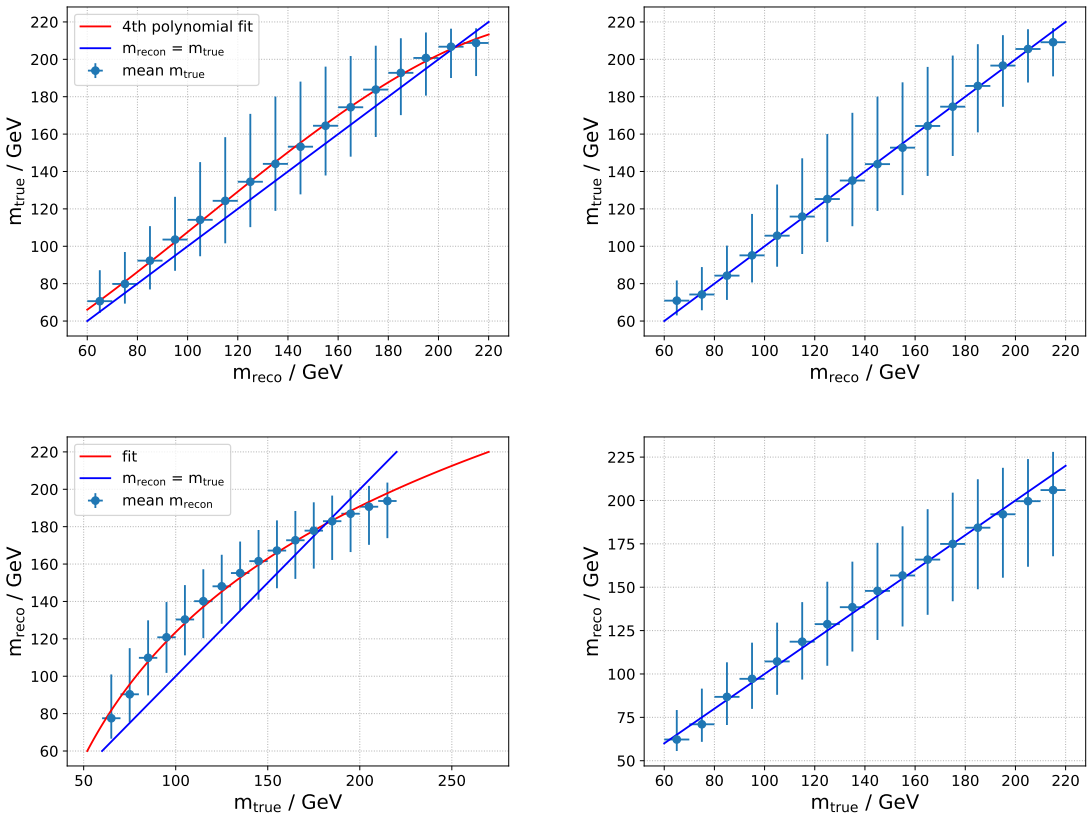


Figure 8.40: The median and 68% quantiles of the true mass in bins of the reconstructed mass for the RNN reconstruction before (top left) and after (top right) the first calibration step, as well as the median and 68% quantiles of the reconstructed mass in bins of the true mass for the RNN reconstruction before (bottom left) and after (bottom right) the second calibration step.

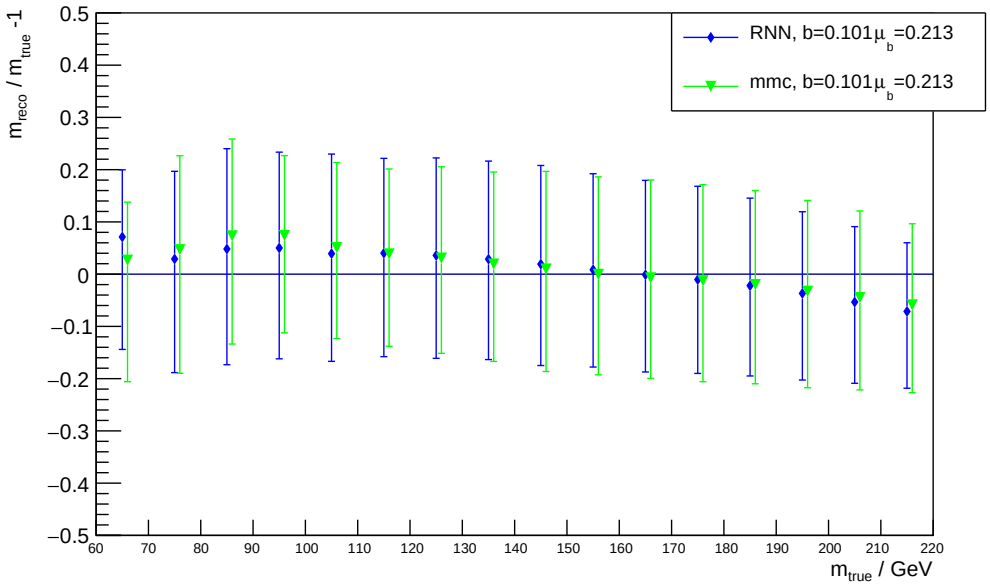


Figure 8.41: The mean relative mass resolution and the 68% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the tauspinner sample and the MMC reconstruction of the spin0hh sample after the calibration.

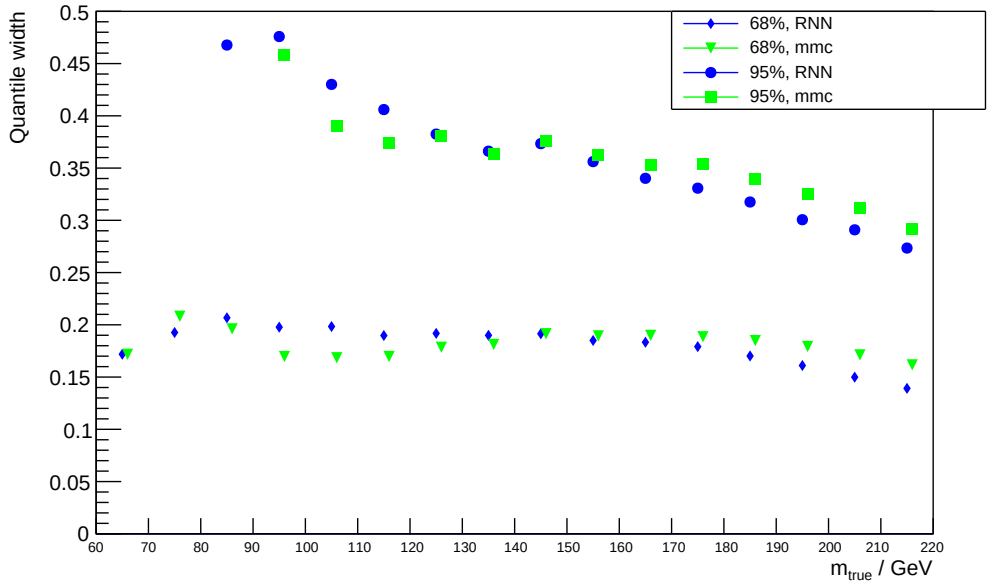


Figure 8.42: The 68% and 95% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the tauspinner sample and the MMC reconstruction of the spin0hh sample after the calibration.

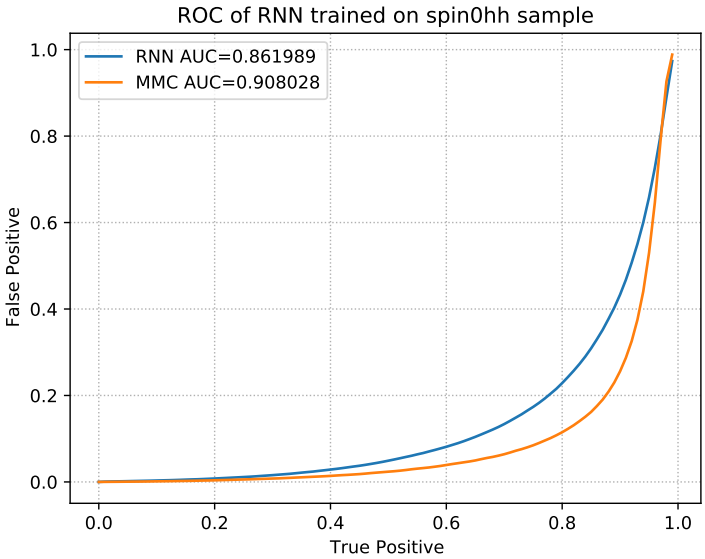


Figure 8.43: A ROC curve of the separation power between the H sample and the Z sample for the RNN trained on the spin0hh sample and the MMC reconstruction.

8.5.4 Combined sample Reconstruciton

To allow the neural network to learn the kinematic differences of both spin-1 and spin-0 events together, a combined sample is created by combining the events of the tauspinner and spin0hh samples. Since the tauspinner events greatly outnumber the spin0hh events, the events are reweighted to give both samples 50% of the total weight.

The relative resolution in bins of the true ditau mass after the calibration is seen in fig. 8.44. It follows a very similar pattern to the relative resolution of the tauspinner sample. The total bias of the RNN is $b_{RNN} = 0.077$ and it is $b_{MMC} = 0.082$ for the MMC. The total and the mean 68% quantile of the RNN is $\mu_{RNN} = 0.204$ of the MMC is $\mu_{MMC} = 0.211$ of the MMC. The 68% quantiles and 95% quantiles in bins of m_{true} are shown in fig. 8.45. With the exception of the $80\,\text{GeV} < m_{true} < 100\,\text{GeV}$ range, the RNN reconstruction has lower 68% quantiles than the MMC reconstruction. The 95% quantiles of the RNN are lower than the MMC across the entire mass range. For both the 68% and 95% quantiles, the difference in the width between the MMC and the RNN grows for higher masses. This is a trend that occurs for all samples studied.

The ROC curves for an RNN trained on the combined sample and the MMC, when applied to the Higgs and Z samples, are shown in fig. 8.46. The AUC of the RNN is 0.897. This is the best performance of all the samples, reaching values very close to the AUC of the MMC. While the low statistics of spin0hh sample, when used to train an RNN on its own, leads to a low RNN performance, it can successfully be used as an addition to the tauspinner sample to mimic spin-1 events and therefore improve the separating power of the RNN. However, it is still less able to separate between Higgs and Z events than the MMC.

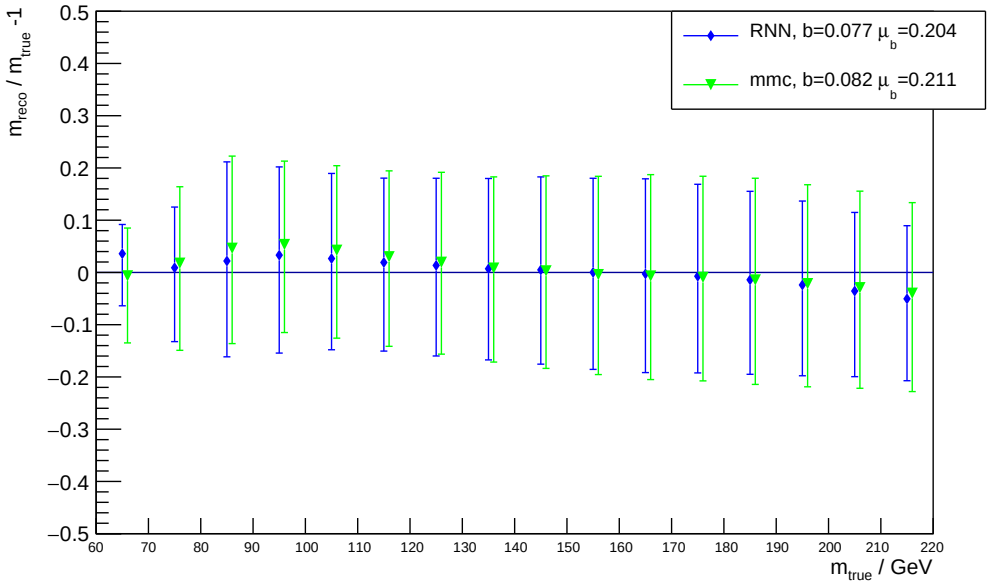


Figure 8.44: The mean relative mass resolution and the 68% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the combined sample and the MMC reconstruction of the spin0hh sample after the calibration.

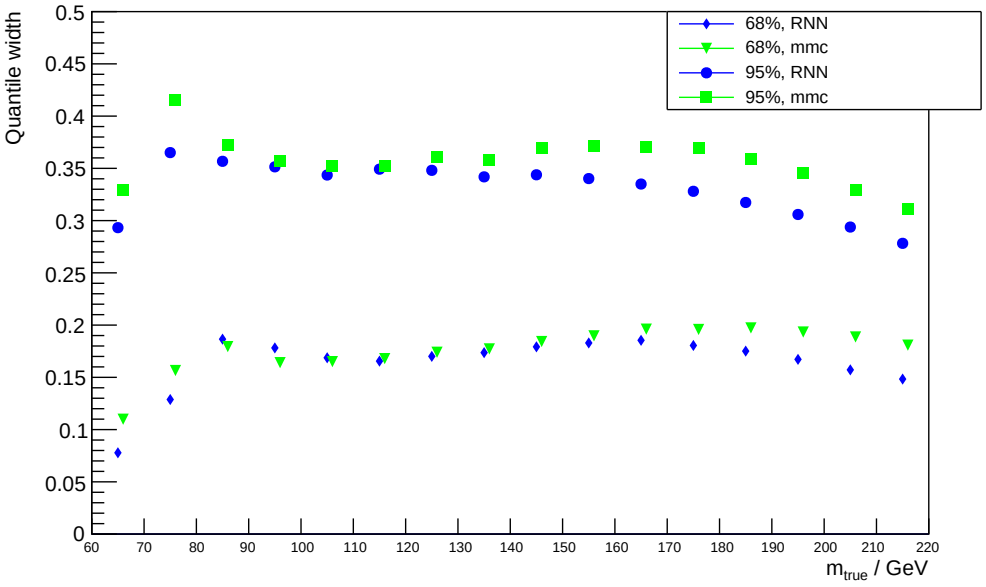


Figure 8.45: The 68% and 95% quantile width of the relative mass resolution in bins of the true ditau mass of a RNN trained on the combined sample and the MMC reconstruction of the spin0hh sample after the calibration.

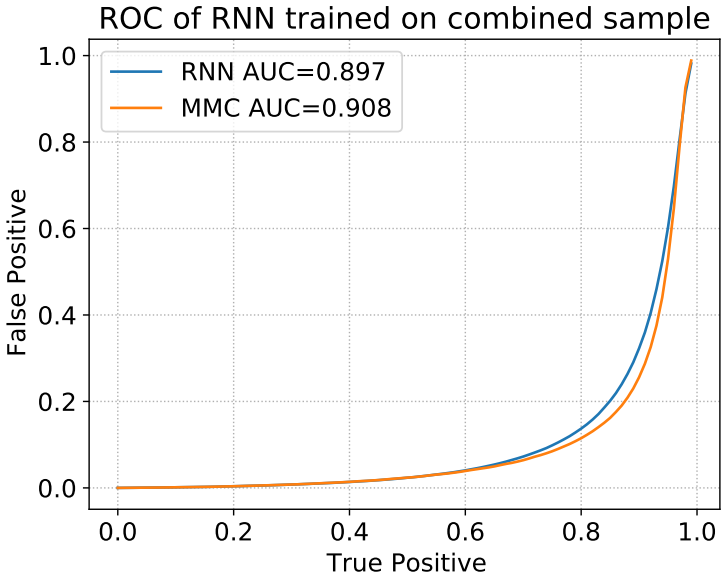


Figure 8.46: ROC curves of the separation power between the H sample and the Z sample for the RNN trained on the combined sample and the MMC reconstruction.

8.6 Target of the Neural Network

In the previous sections, the true mass $t_1 = m_{true}$ of the event is the target variable of the RNN which it learns to reconstruct. In this section some other targets for the RNN training are also considered. These are as follows:

$$t_2 = m_{true}^{inv} = m_{true} - m_{true}^{vis} \tag{8.60}$$

$$t_3 = m_{recon}^{inv} = m_{true} - m_{reco}^{vis} \tag{8.61}$$

$$t_4 = \frac{m_{true}}{m_{true}^{vis}} \tag{8.62}$$

$$t_5 = \frac{m_{true}}{m_{reco}^{vis}} \tag{8.63}$$

where m_{true}^{inv} is the matched true invisible neutrino mass, m_{true}^{vis} is the matched true visible mass of the tau decays, m_{recon}^{inv} is the invisible neutrino mass as determined through the reconstructed MET, and m_{recon}^{vis} is the reconstructed visible mass of the tau decays in the detector. t_1, t_3 and t_5 can be used in analysis with real data. t_5 could diverge for very small m_{recon}^{inv}, though there are cuts in the samples on this variable. t_2 and t_4 can not be used in analysis with real data, since they require the true visible mass m_{true}^{vis} to be perfectly reconstructed for use.

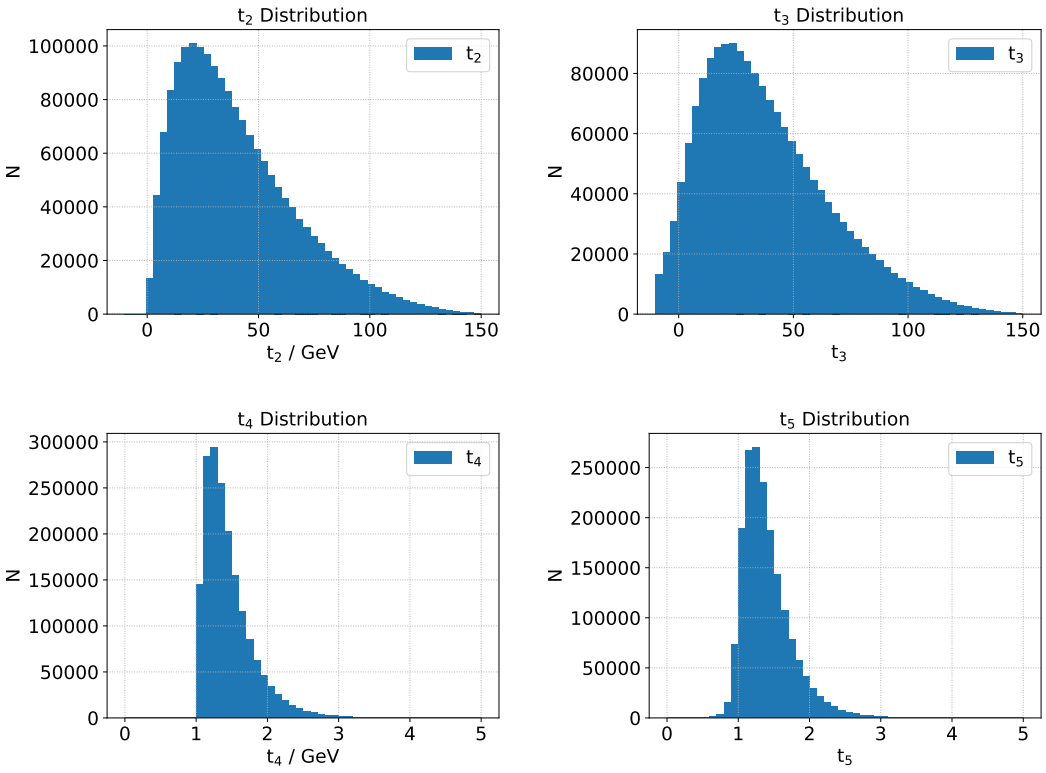


Figure 8.47: Targets t_2-t_5 of the combined dataset.

The targets t_2-t_5 might be helpful for the RNN training since it moves the focus of the reconstruction onto the invisible neutrino momentum, which is the main problem of the ditau mass reconstruction.

The distributions of the targets t_2-t_5 are shown in fig. 8.47 for the γ^* sample. As expected, there are some negative values for target t_3 since it is possible that the reconstructed energy is larger than the actual ditau mass. For t_5, the same effect is observed when some events have a value of under 1.

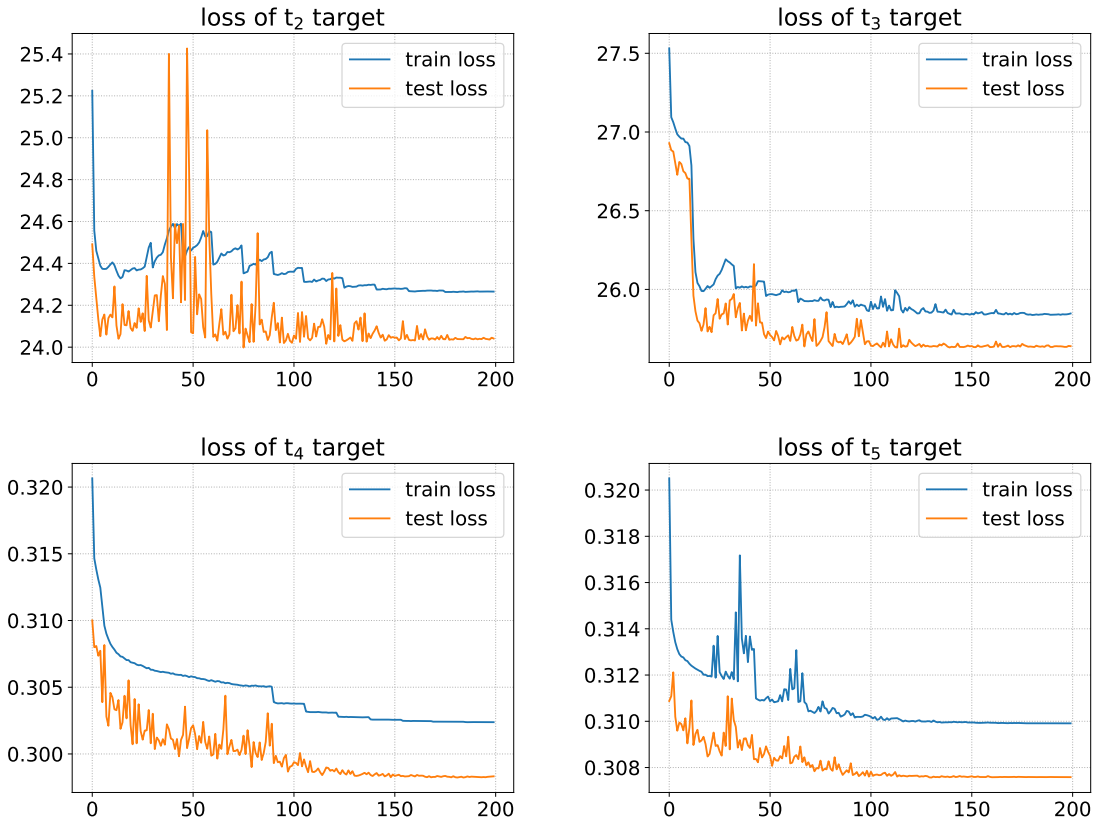


Figure 8.48: Loss graphs of RNNs using the targets t_2-t_4 trained on the combined spin0hh and tauspinner sample.

The loss graphs of neural networks trained on each of the four new targets of the combined dataset are shown in fig. 8.48. Noteworthy is that the targets consisting of the ratio of the total mass and the visible mass follow a smoother trend than the loss graphs of the other targets. Especially target t_4 is very smooth in its training loss and has no spiking whatsoever.

The resulting prediction is reconverted back into a reconstruction of the true mass by applying the inverse of the equations 8.60 to 8.63. The relative mass resolution of the reconstructions can be seen in fig. 8.49. The neural networks trained on targets that include the true visible mass have a residual bias in the relative mass resolution that is slightly higher ($\mu_b = -0.013$ for t_4 and $\mu_b = -0.014$ for t_2, compared to $\mu_b = -0.005$ for t_5 and $\mu_b = -0.007$ for t_3), and the standard deviations of the relative mass resolution are higher for the targets that include the reconstructed visible mass ($\sigma = 0.190$ for t_3 and $\sigma = 0.202$ for t_5, compared to $\sigma = 0.173$ for t_2 and $\sigma = 0.178$ for t_4).

The reconstruction of neural networks trained on all five targets are compared in a binned relative

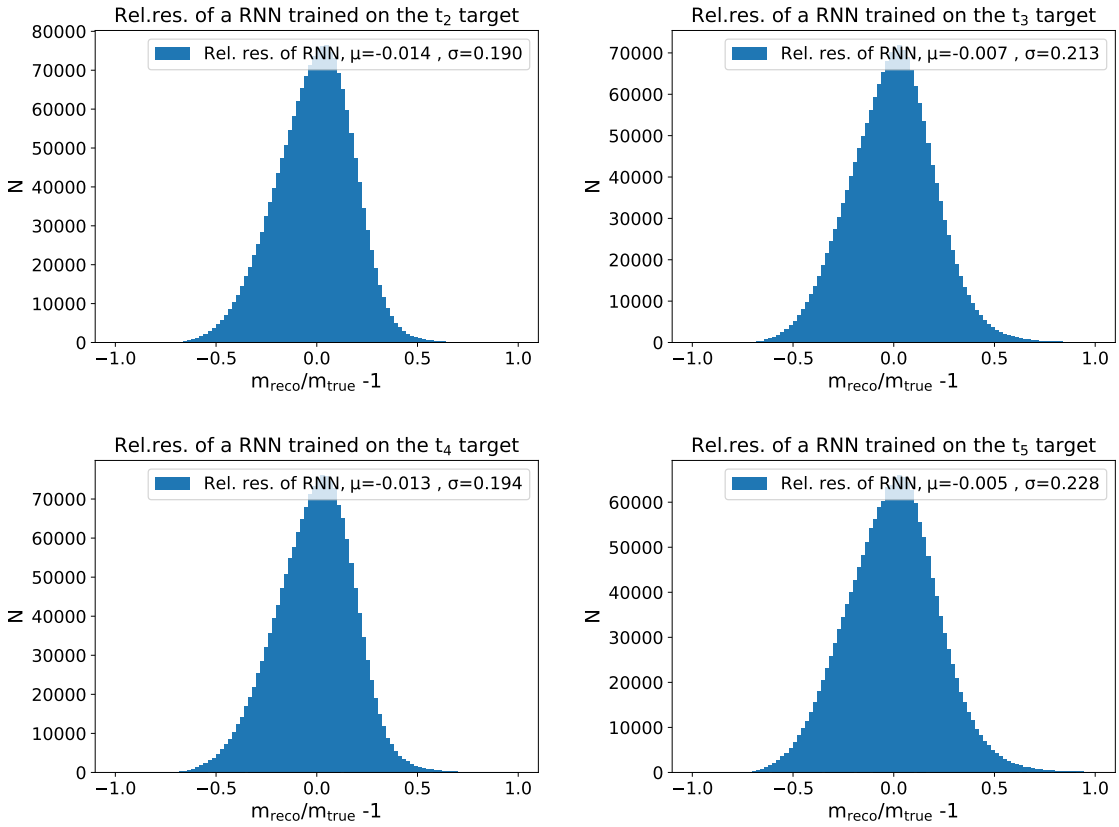


Figure 8.49: The relative mass resolution of the four additional targets after the calibration, using the combined spin0hh and tauspinner sample.

mass resolution plot after the calibration in fig. 8.50. The calibration manages to remove most of the bias for all targets. The 68% and 95% quantiles after the calibration are shown in fig. 8.51. Both t_2 and t_4 lead to better reconstructions and have lower quantile widths for all mass bins where $m_{true} < 180\,\text{GeV}$. For higher mass bins the standard t_1 target has smaller quantile widths for both the 68% and 95% quantiles. t_3 and t_5 lead to higher quantile widths for almost all bins.

ROC curves of the separation power between the H sample and the Z sample for the RNN trained on the targets t_1-t_4, using the combined spin0hh and tauspinner sample, are shown in fig. 8.52. While the AUC of the neural network trained on t_1 is 0.897, it is 0.904 and 0.903 for t_2 and t_4 respectively. For t_3 and t_5, the AUC value are 0.876 and 0.850 respectively. Neural networks trained on the targets t_2 and t_4 have an improved separation power between Higgs and Z events. However, since the the reconstruction for these targets requires the true visible mass to be known, they can not be used in analysis with real data. Therefore, the optimal target out of the five that are analyzed in this section is $t_1 = m_{true}$

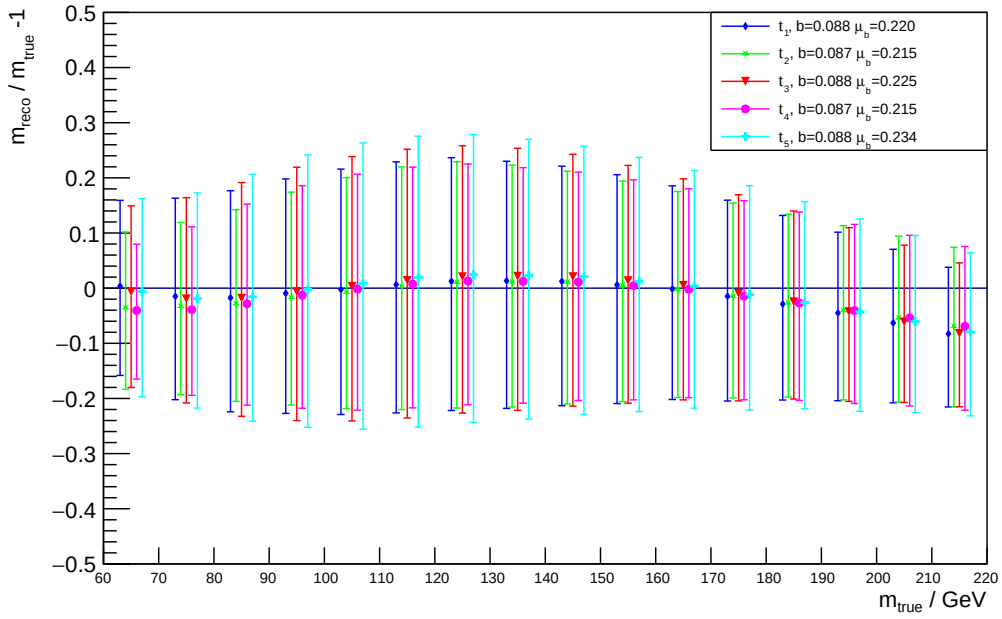


Figure 8.50: The mean relative mass resolution and the 68% quantile width of the relative mass resolution in bins of the true ditau mass of RNNs trained on the targets $t_2 - t_4$, using the combined spin0hh and tauspinner sample after the calibration.

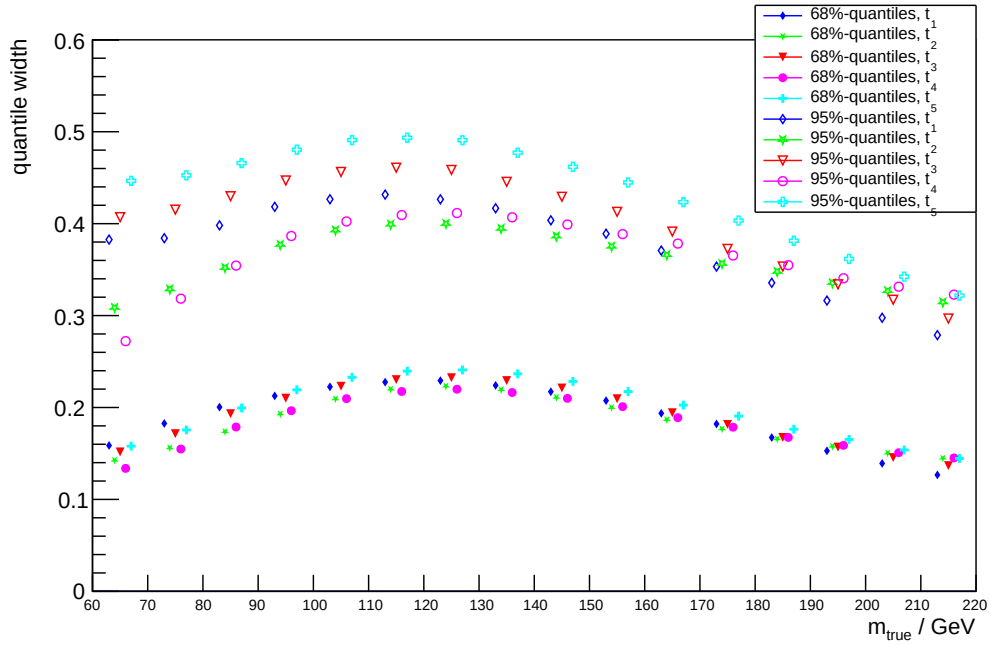


Figure 8.51: The 68% and 95% quantile width of the relative mass resolution in bins of the true ditau mass of RNNs trained on the targets t_2-t_4, using the combined spin0hh and tauspinner sample after the calibration.

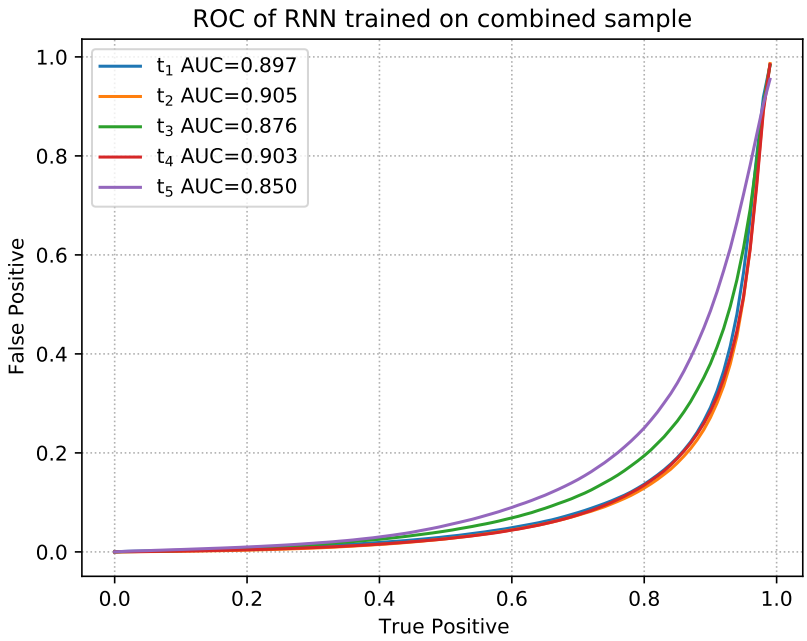


Figure 8.52: ROC curves of the separation power between the H sample and the Z sample for RNNs trained on the five targets $t_1 - t_5$.

8.7 Expanded Input Variables

In addition to the standard input set, three further sets of input variables are presented: the Auxiliary set, the AuxiliaryExtra set, and the Jets set. In the Auxiliary variable set, the number of charged tracks reconstructed is replaced with the reconstructed decaymode of the taus. This variable is included to allow the neural network to learn the kinematics of the different decaymodes, especially the difference between the 1p0n and 1p1n decay modes, which have the same number of charged tracks. It also includes the MET significance MET_{sig}, the pile-up of the event and the number of reconstructed jets that are assigned to the event. MET_{sig} is a measure of how likely the measured MET is a consequence of random fluctuations. This can be used by the neural network to learn how much relevance to put on the MET. Pile-up describes the number of proton-antiproton collisions that occur during the same bunch crossing in the ATLAS interaction point, and generally leads to a deterioration in the quality of reconstructing jets. In total, the Auxiliary variable set has 13 variables.

The AuxiliaryExtra variable set builds on the Auxiliary set by adding the number of muons n_μ as an additional input and splitting up the Lorentz vectors of the taus into their charged and neutral components. The transverse momentum, the pseudorapidity and the azimuthal angle are then considered separately for the neutral as well as the charged component for a total of 23 variables. As discussed in chapter 2.5, the charged and the neutral parts of a tau decay both contain information on the helicity of the tau, and, therefore, also contain information on the spin of the source resonance.

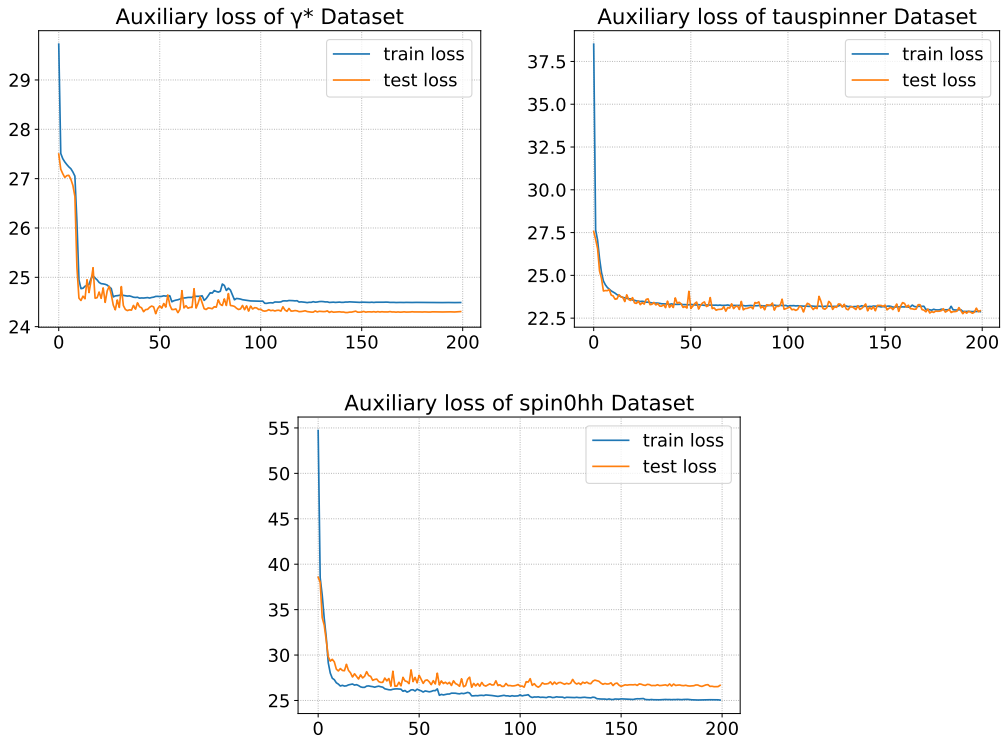


Figure 8.53: Loss graphs of RNNs using the Auxiliary variable set trained on the γ^* sample (top left), tauspinner sample (top right) and spin0hh sample (bottom).

The final input variable set is the Jets set. It has the variables of the standard set but also includes the 4-vector information (p_T, ϕ, η and m) of the three most energetic jets that are not allocated to the tau leptons of the event. These can be used to reconstruct the boost of the resonance. The Jets set has 22 variables. All added variables of the three variable sets are shown in the appendix A for the five samples.

The loss plots of the Auxiliary variable set for the γ^*, tauspinner- and spin0hh samples are shown in fig. 8.53. Both the training loss and the testing loss are shown. The losses converge to similar values, as they do for a neural network trained on the standard variable input set, and the training is generally stable. The same plots for the Jets set are shown in fig. 8.54. No significant difference in the training process or final loss values can be seen compared to the other variable input sets. The loss plots for the AuxiliaryExtra variable set are shown in fig. 8.55. The training process set seems a lot more inconsistent compared to the other sets. This might be because the neural network has to learn how to accurately add Lorentz vectors to each other, which is a highly nonlinear process. While neural networks are supposed to be able to handle nonlinear problems, the entire ditau reconstruction might be too complex for the given architecture. However, the loss does eventually stabilize and converges as it nears the end of the training.

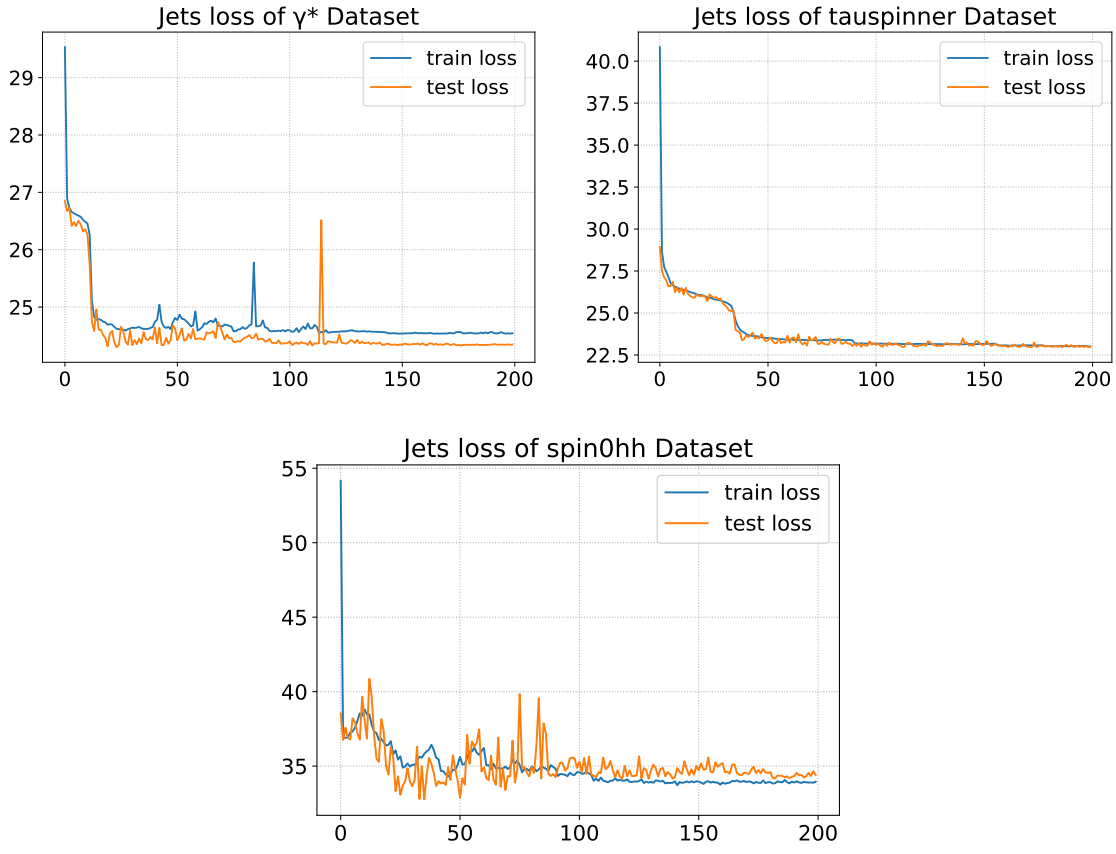


Figure 8.54: Loss graphs of RNNs using the Jets variable set trained on the γ^* sample (top left), tauspinner sample (top right) and spin0hh sample (bottom).

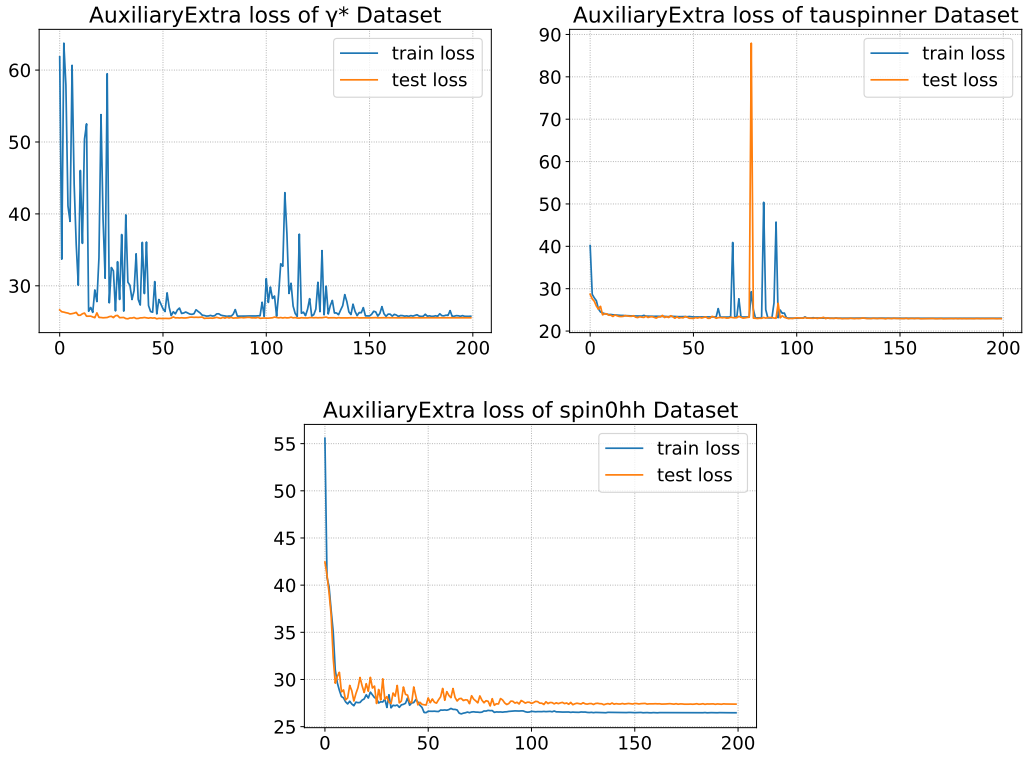


Figure 8.55: Loss graphs of RNNs using the AuxiliaryExtra variable set trained on the γ^* sample (top left), tauspinner sample (top right) and spin0hh sample (bottom).

In general, the addition of further variables into the training process does not seem to alter the final loss values significantly. Binned relative mass resolution plots are used in order to further study any effect that the additional inputs might have on the reconstruction performance of the neural networks. Fig. 8.56 shows the binned relative mass resolution of the four sets of input variables trained on the γ^* sample. Neural networks trained on the four variable input sets show very similar patterns. Some notable differences, however, are that the Jets set of variables seems to have a significantly greater bias at lower masses and also a slightly greater bias at higher high masses. In general, the total bias is smallest for the standard set, and the 68% quantile width is the most favorable for the Auxiliary set.

The same fit after the calibration is shown in fig. 8.57. The calibration is able to remove most of the bias for all four sets of input variables, though some bias remains for the higher bins. In addition, a bias of $b = 0.07$ is present in the lowest mass bin for the Jets set of input variables. The 68% quantile and 95% quantile widths are shown in fig. 8.58. The standard and Auxiliary sets of input variables have the smallest 68% and 95% quantile widths. As these two are the simplest of the four sets of variables, this indicates that the additional variables do not help in the reconstruction and complicate the training instead, leading to less satisfying ditau mass reconstruction performances. ROC curves of the Higgs sample and Z sample reconstruction of neural networks trained with these variable sets on the γ^* dataset are shown in fig. 8.59. The ROC curves of the RNNs trained on the Auxliary and AuxiliaryExtra variable sets show a slight decrease in separation power between the Higgs sample and Z sample. The Jets set has a significantly worse separation power with an AUC = 0.773.

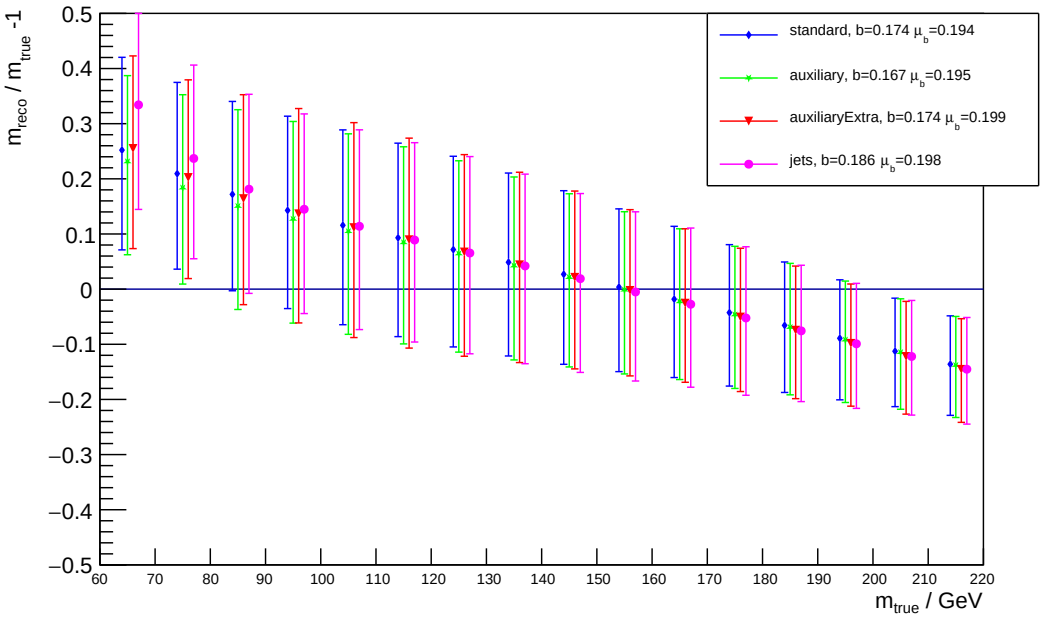


Figure 8.56: Relative resolution of the four sets of input variables in bins of the true mass trained on the γ^* sample before the calibration.

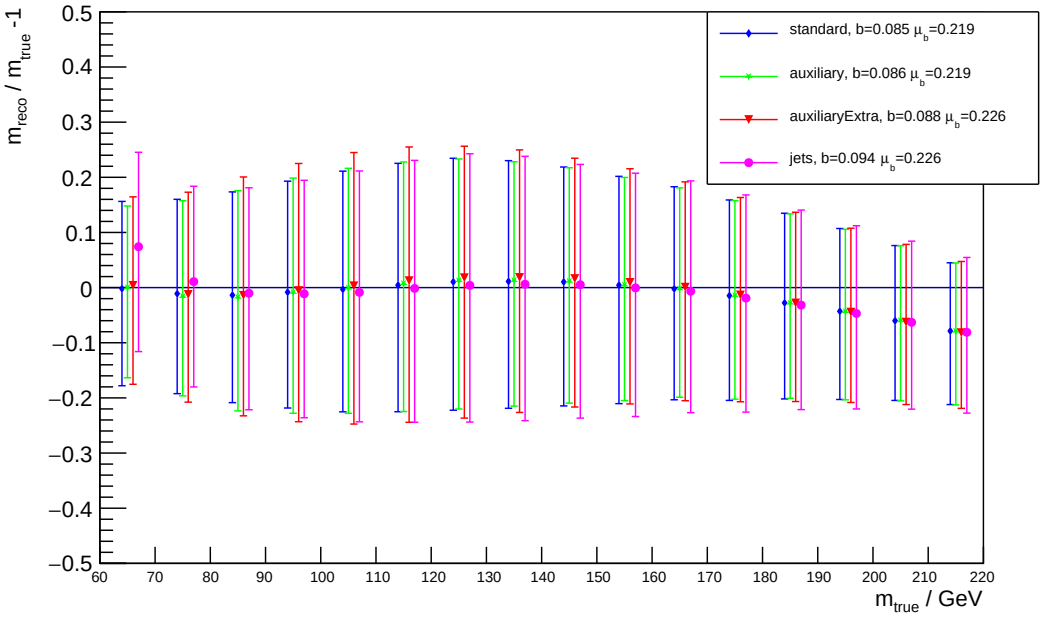


Figure 8.57: Relative resolutions in bins of the true mass after the calibration of RNNs trained on the γ^* sample using the four sets of input variables

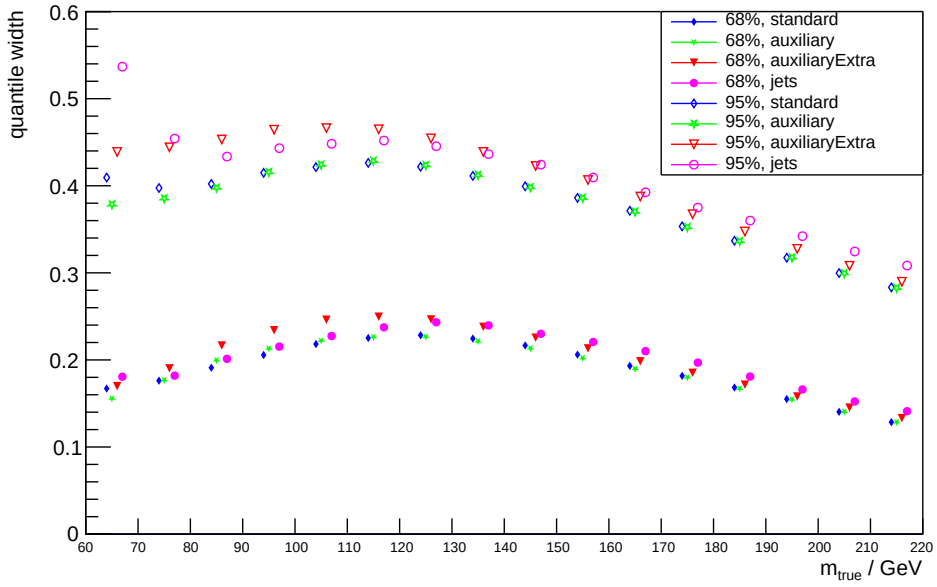


Figure 8.58: 68% and 95% quantiles of the relative mass resolution of the four sets of input variables in bins of the true mass after the calibration.

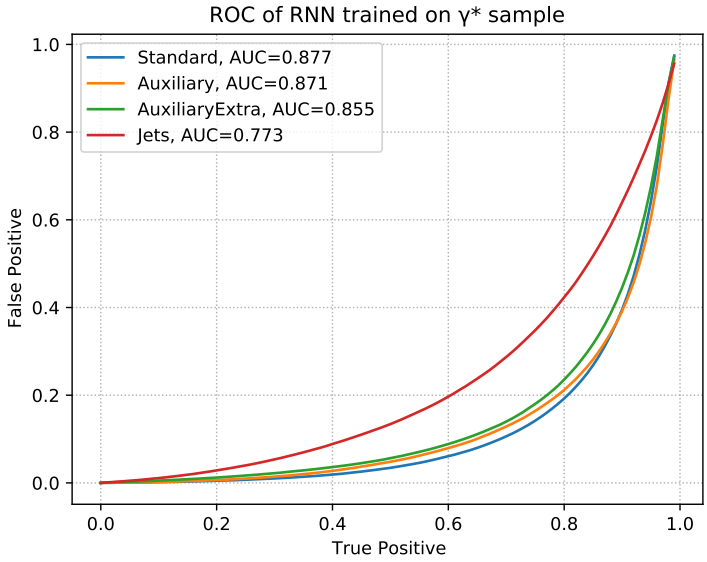


Figure 8.59: ROC curves of the Higgs sample and Z sample reconstruction of neural networks trained with with the four variable sets on the γ^* dataset

The binned relative mass resolution after the calibration of the four sets of input variables for the tauspinner sample is shown in fig. 8.60 and the same plot for the spin0hh sample is shown in fig. 8.61. For the tauspinner sample, all variable input sets lead to a bias of $b = 0.073$. The biases of the individual bins are also very similar. For the spin0hh sample, The bias of the standard, Auxiliary and AuxiliaryExtra sets are also very similar, between 0.102 and 0.104. However, the Jets set follows a very different pattern and has a significantly higher bias. Especially in the low mass range, there is a very large positive bias, reaching $b_i = 0.36$ for the $70\,\text{GeV} < m_{true} < 80\,\text{GeV}$ bin.

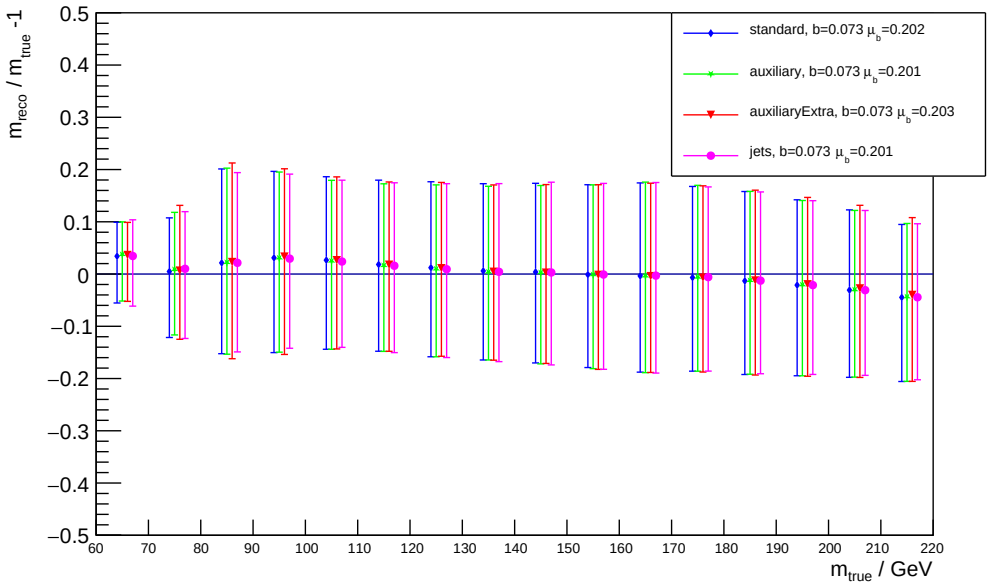


Figure 8.60: Relative resolutions in bins of the true mass after the calibration of RNNs trained on the tauspinner sample using the four sets of input variables.

The 68% quantile and 95% quantile widths for the tauspinner sample after the calibration can be seen in fig. 8.62. The only significant difference between the four sets are the slightly higher 68% and 95% quantile widths for the AuxiliaryExtra set in most bins. The quantiles of RNN reconstructions of the spin0hh sample can be seen in fig. 8.63. The standard, Auxiliary and AuxiliaryExtra sets show very similar performances for both the 68% and 95% quantiles, though in the 95% quantile, there are very large values of over 0.5 for the bins where $m_{true} < 90\,\text{GeV}$. As already discussed in chapter 8.5.3, this is not an effect exclusive to the RNN, since it also appears in the MMC reconstruction.

ROC curves of the Higgs sample and Z sample reconstruction of neural networks trained with these variable sets on the tauspinner sample and spin0hh sample are shown in fig. 8.64. The ROC curves of the RNNs trained on the Auxliary variable set show a slight decrease in separation power between the Higgs sample and Z sample. The ROC curve of the neural networks trained with the AuxiliaryExtra variable set have an AUC = 0.621. For the spin0hh sample, the neural network trained with the jets variable set also shows a significant decrease in separation power with an AUC = 0.730.

Generally it can be said that the extended input sets do not lead to an increase in the reconstruction performance of the neural network, and instead leads to a performance deterioration. The added

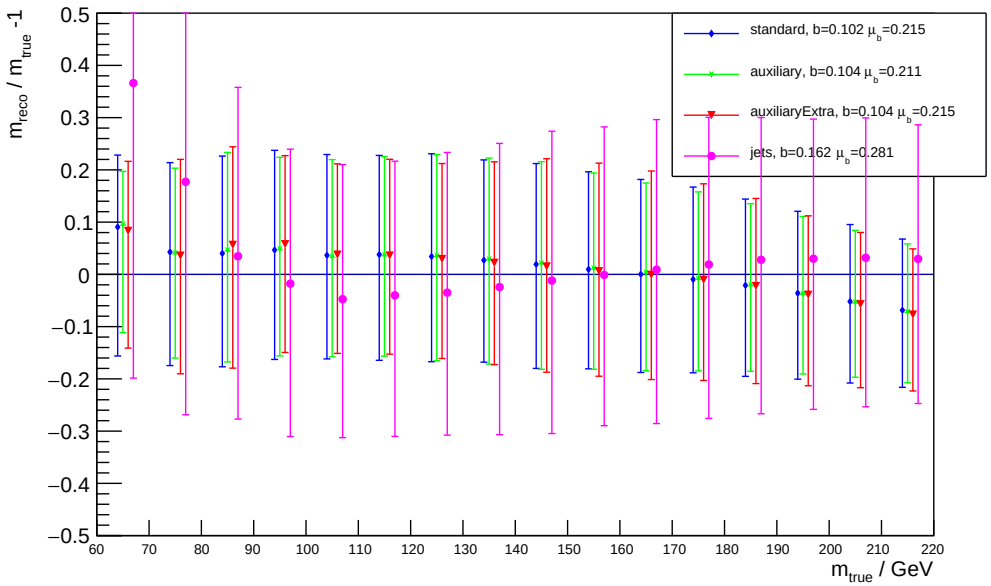


Figure 8.61: Relative resolutions in bins of the true mass after the calibration of RNNs trained on the spin0hh sample using the four sets of input variables.

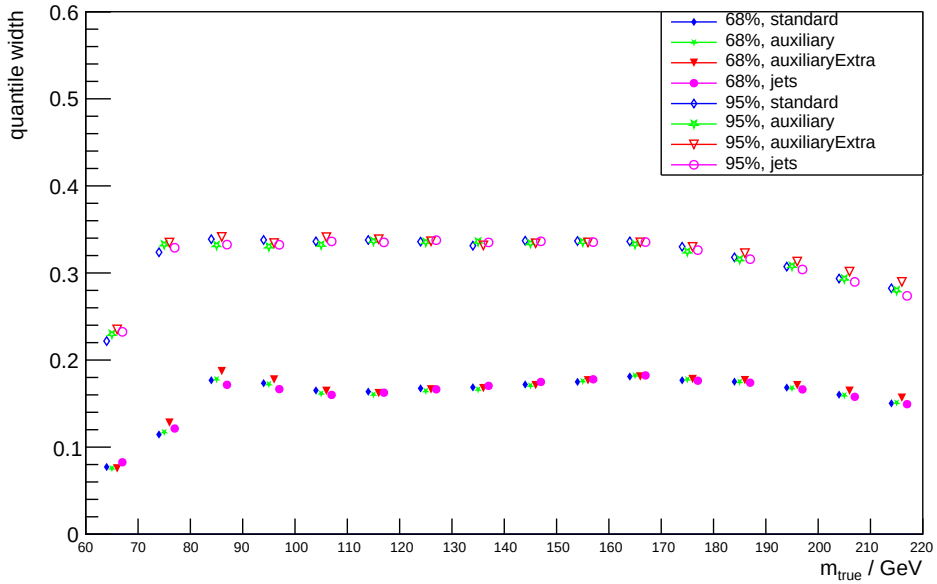


Figure 8.62: 68% and 95% quantiles of the relative mass resolutions of RNNs trained on the tauspinner sample using the four sets of input variables after the calibration.

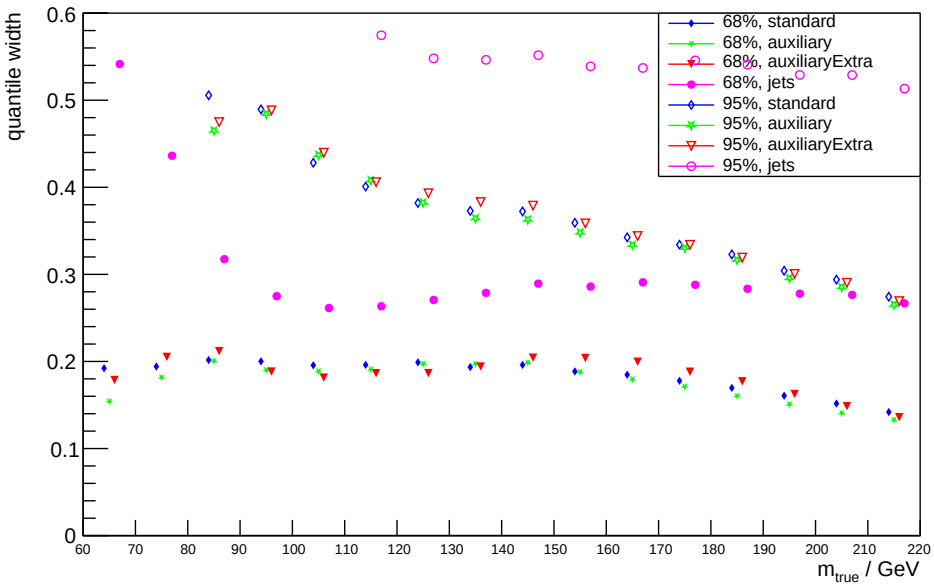


Figure 8.63: 68% and 95% quantiles of the relative mass resolutions of RNNs trained on the spin0hh sample using the four sets of input variables after the calibration.

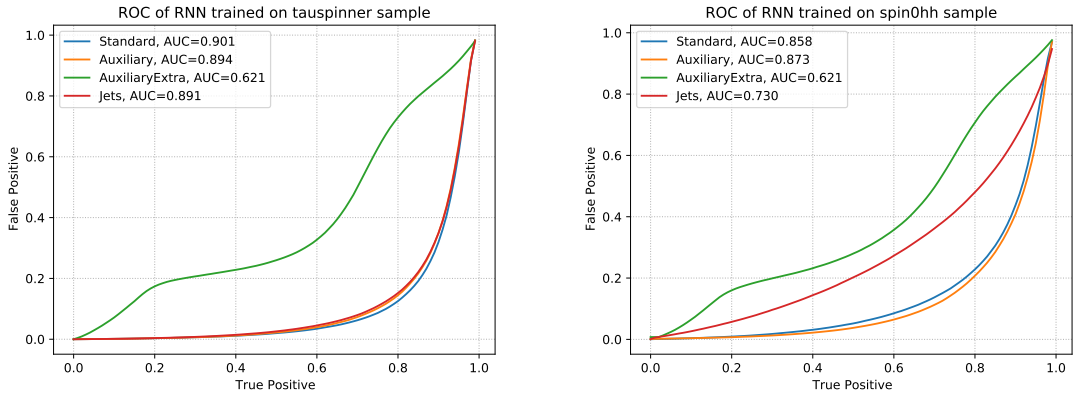


Figure 8.64: ROC curves of the Higgs sample and Z sample reconstruction of neural networks trained with the four variable sets on the tauspinner sample (left) and spin0hh sample (right)

variables might make the problem too complex for the neural network to solve. Especially the AuxiliaryExtra set seems to demand too much of the neural networks since it requires the neural network to add up the lorentz vectors of the charged part and neutral part of the tau decay. Having too many variables can also make neural networks prone to overtraining. It is therefore recommended to either use the standard variable set or the Auxiliary set.

8.8 Other Loss Functions

In the previous sections the squared error loss $L = (m_{true} - m_{reco})^2$ is always used for the neural network training. In this section two alternative loss functions are considered. The first one is an alternative to the calibration to correct for the bias. The second loss function includes a penalty term that adds loss if there is a correlation between the helicity and mass in the reconstruction.

8.8.1 Edge Corrected Loss Function

The inclusion of an edge correction in the loss function is an approach studied in [93]. Although it manages to reduce the bias in the reconstruction, there is an additional variable which must be optimized and it leads to a deterioration in neural network performance. The edge corrected loss function is

$$L(y;\theta) = \frac{1}{m}\sum_{0}^{m}\left(\left(\frac{\hat{y}-y}{2\sigma_i^2}\right)^2 + \ln\left(\mathrm{erf}\left(\frac{\hat{y}-y_{min}}{\sqrt{2\sigma_i^2}}\right) - \mathrm{erf}\left(\frac{\hat{y}-y_{max}}{\sqrt{2\sigma_i^2}}\right)\right) + \ln\left(\sqrt{2\pi\sigma_i^2}\right)\right) \tag{8.64}$$

where $y_{min} = 60\,\mathrm{GeV}$ and $y_{max} = 220\,\mathrm{GeV}$ are the minimum and maximum true mass values of the sample. The error function terms in the loss function simulate a normal distribution at the boundaries of the sample, which rewards the neural network if an event is reconstructed outside of the edges of the sample. The standard deviation σ_i of the normal distribution must be estimated. Since events with low masses and high masses are expected to have different standard deviations, σ_i should also depend on the individual mass y_i of each event. In this thesis a linear relation $\sigma_i = \sigma\hat{y}$ is used, whereby σ must be optimized.

The relative mass resolution in bins of the true mass is given in fig. 8.65 for $\sigma = 0.00, 0.05, 0.10, 0.18$. The bias decreases with growing σ, but significantly more bias remains as in the calibration approach. The 68% and 95% quantile widths are shown in fig. 8.66. The quantile widths increase with growing σ. The largest increase is for high masses, while the widths stay nearly constant for the mass range $90\,\mathrm{GeV} < m_{true} < 130\,\mathrm{GeV}$.

The ROC curves of the four neural networks presented in fig. 8.65 applied to the Higgs and Z sample are shown in fig. 8.67. For $\sigma = 0.00$, the AUC is 0.896 and for $\sigma = 0.18$, the AUC is 0.870. Therefore, while the edge corrected loss function does partially remove the bias, it also leads to a deterioration in the separation power between the Higgs and Z sample.

The edge corrected loss function is able to lower the bias of the binned relative mass resolution, but some bias remains. The calibration is a more capable method of bias removal. Contrary to the calibration, the edge corrected loss function does lead to a deterioration of the separation power between Higgs and Z events, and is therefore an inferior way to remove the bias.

8.8.2 Helicity Correlation Loss Function

As discussed in section chapter 2.5, the kinematics of the ditau decay depend on the helicity of the tau leptons. This can lead to a bias in the reconstruction that depends on the spin of the resonance. A modification to the loss function that measures this bias and punishes the neural network for it can be used to resolve this issue.

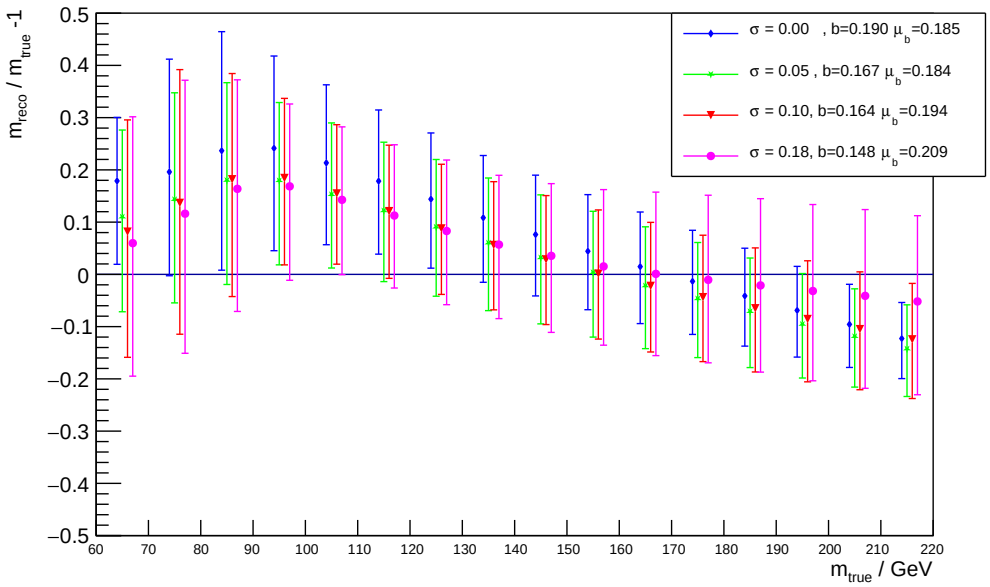


Figure 8.65: Relative resolutions in bins of the true mass of RNNs trained on the spin0hh sample using the edge corrected loss function on the combined sample.

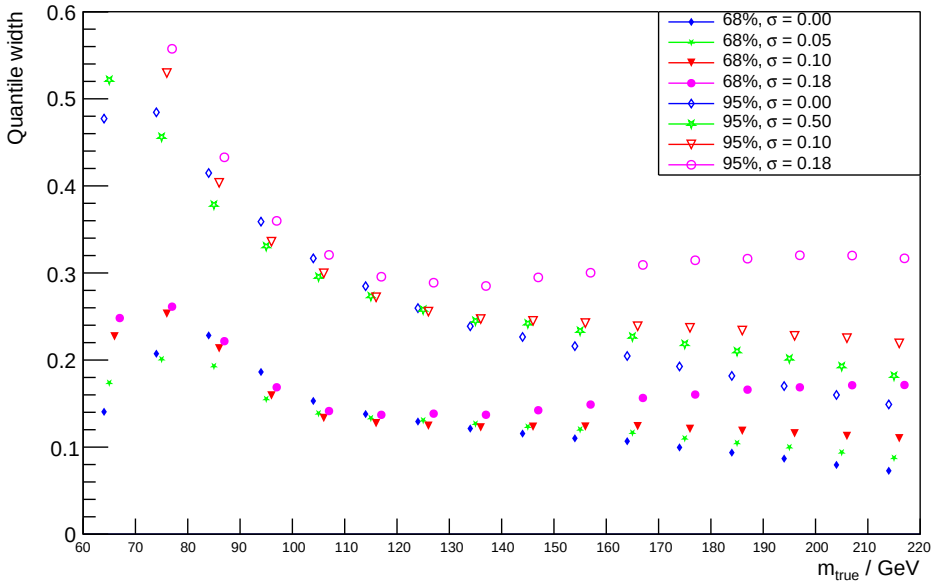


Figure 8.66: 68% and 95% quantiles of the relative mass resolutions of RNNs trained on the spin0hh sample using the edge corrected loss function on the combined sample.

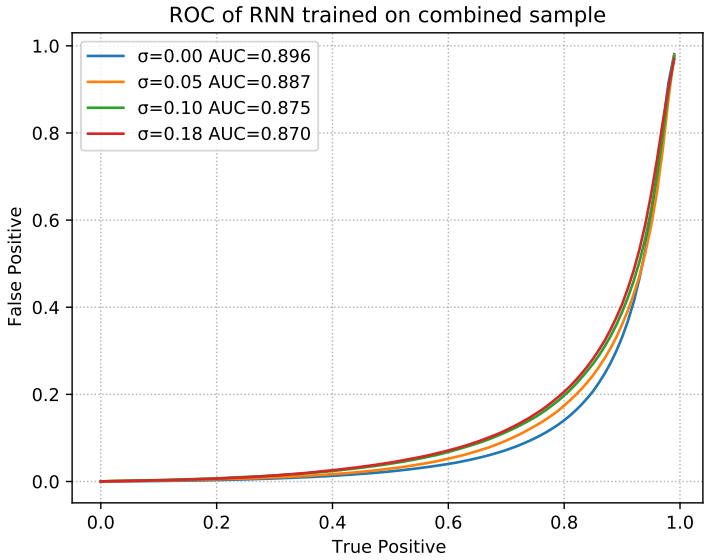


Figure 8.67: ROC curves of the separation power between the H sample and the Z sample for RNNs trained using the edge corrected loss function on the combined sample.

For this, a helicity correlation loss function is used:

$$L(y;\theta) = \frac{1}{m}\sum_{0}^{m}\left((\hat{y}-y)^2 + \lambda \cdot \left|\mathrm{Corr}(y_\mathrm{H},\hat{y}_\mathrm{H}) - \mathrm{Corr}(y_\mathrm{Z},\hat{y}_\mathrm{Z})\right|\right) \tag{8.65}$$

where the correlation between two variables is defined as

$$\mathrm{Corr}(x,y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2+\epsilon}}. \tag{8.66}$$

y_H and $\hat{y}_\mathrm{H}$ refer to only those true and reconstructed masses that are from the H-sample and y_Z and $\hat{y}_\mathrm{Z}$ refer to only those true and reconstructed masses that are from the Z-sample. $\epsilon = 0.00001$ ensures that the denominator is never zero. The variable λ must be optimized.

The relative mass resolutions after the calibration of RNNs that are trained with the correlation loss function in bins of the true mass with $\lambda = 0, 500, 1000, 1800$ are shown in fig. 8.68. The 68% and 95% quantile widths for the four RNN reconstructions are shown in fig. 8.69. There are no significant differences between the four RNNs, although the performance does degrade slightly for larger λ values. For $\lambda = 0$, the mean 68% quantile width is $\mu_b = 0.204$ and for $\lambda = 1800$, the mean 68% quantile width is $\mu_b = 0.206$. The deterioration in the relative resolution mostly appears in the mass range $80\,\mathrm{GeV} < m_{true} < 110\,\mathrm{GeV}$ and above 180 GeV.

Fig. 8.70 shows the ROC curves of the four neural networks applied to the Higgs and Z sample. For small values of $\lambda \leq 1000$, a slight difference in the AUC is observed with growing λ values. For $\lambda = 1800$ a more significant deterioration in the separating power becomes apparent. The AUC is 0.897 for $\lambda = 0$ and falls to 0.890 for $\lambda = 1800$.

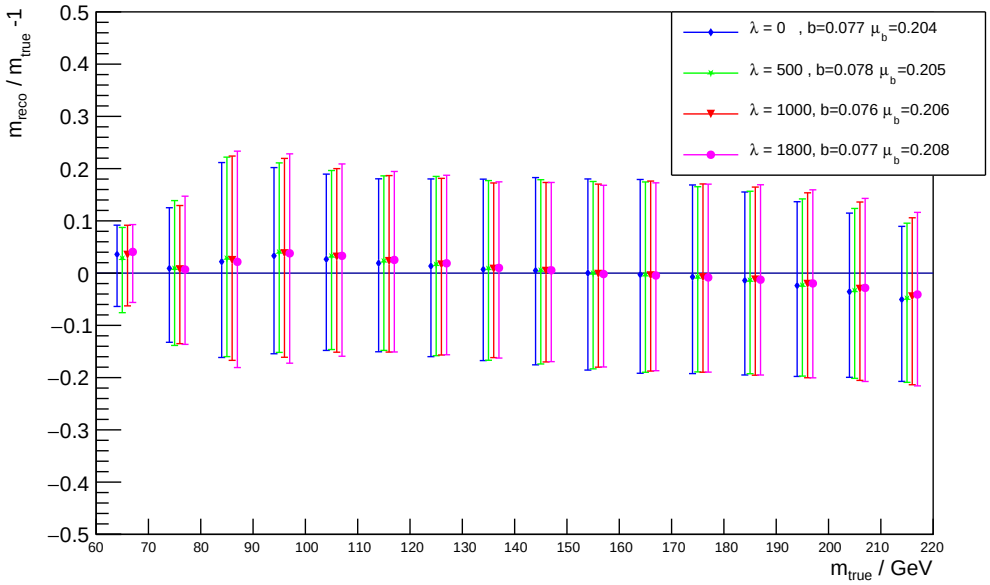


Figure 8.68: Relative resolutions in bins of the true mass of RNNs trained on the combined sample using the helicty correlation loss function.

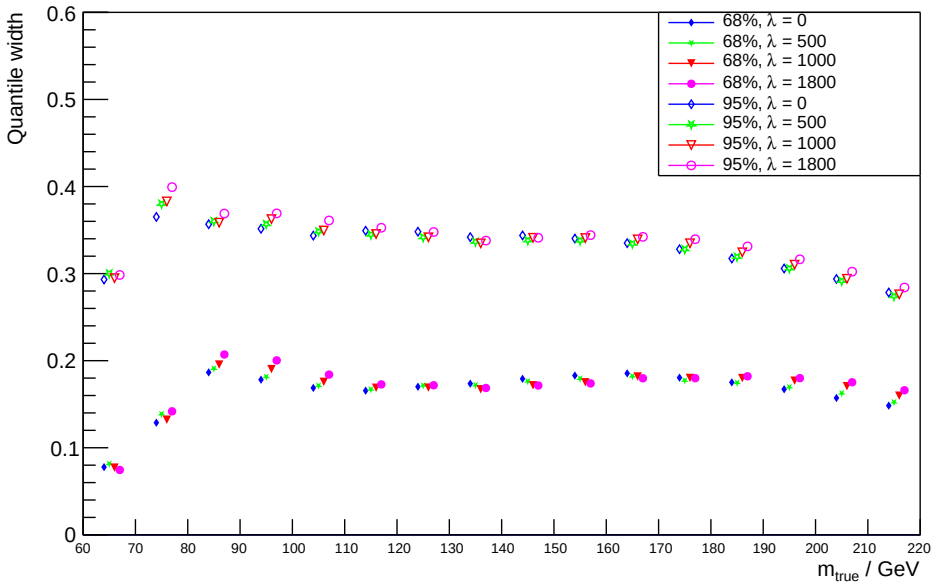


Figure 8.69: 68% and 95% quantiles of the relative mass resolutions of RNNs trained on the combined sample using the helicty correlation loss function.

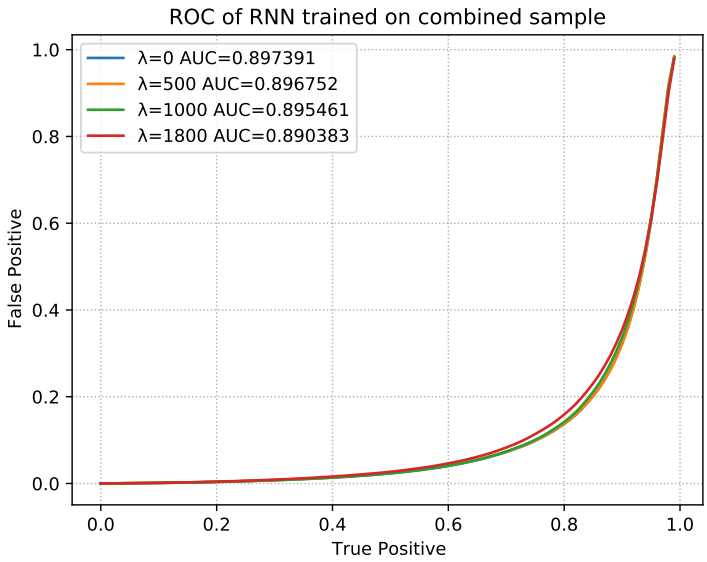


Figure 8.70: ROC curves of the separation power between the H sample and the Z sample for RNNs trained using the helicty correlation loss function on the combined sample.

The correlation loss function does not seem to lead to a superior separation power between Higgs and Z events. This might be a consequence of low statistics for the tauspinner Higgs sample in the low mass region, or it might imply that the reconstruction has no significant bias in the reconstruction between spin-1 and spin-0 events.

8.9 Summary

Tab. 8.2 shows a summary of the reconstruction performances of the various neural networks that were trained in this thesis. The MMC reconstruction performances are also shown.

Sample	Loss	Input Variables	Target	b	μ_b	AUC
γ^*	Squared Error	Standard	m_{true}	0.086	0.219	0.875
Tauspinner	Squared Error	Standard	m_{true}	0.077	0.205	0.887
Spin0hh	Squarred Error	Standard	m_{true}	0.101	0.213	0.862
Combined	Squarred Error	Standard	m_{true}	0.077	0.204	0.897
Combined	Squarred Error	Standard	$m_{true} - m_{true}^{vis}$	0.087	0.215	0.905
Combined	Squarred Error	Standard	$m_{true} - m_{reco}^{vis}$	0.088	0.225	0.876
Combined	Squarred Error	Standard	m_{true}/m_{true}^{vis}	0.087	0.215	0.903
Combined	Squarred Error	Standard	m_{true}/m_{reco}^{vis}	0.088	0.234	0.850
γ^*	Squared Error	Aux	m_{true}	0.086	0.219	0.871
Tauspinner	Squared Error	Aux	m_{true}	0.073	0.201	0.894
Spin0hh	Squarred Error	Aux	m_{true}	0.104	0.211	0.872
γ^*	Squared Error	AuxExtra	m_{true}	0.088	0.226	0.855
Tauspinner	Squared Error	AuxExtra	m_{true}	0.073	0.203	0.621
Spin0hh	Squarred Error	AuxExtra	m_{true}	0.104	0.215	0.621
γ^*	Squared Error	Jets	m_{true}	0.094	0.226	0.772
Tauspinner	Squared Error	Jets	m_{true}	0.073	0.201	0.891
Spin0hh	Squarred Error	Jets	m_{true}	0.162	0.281	0.730
Combined	Edge Correc. $\sigma = 0.00$	Standard	m_{true}	0.190	0.185	0.896
Combined	Edge Correc. $\sigma = 0.05$	Standard	m_{true}	0.167	0.184	0.887
Combined	Edge Correc. $\sigma = 0.10$	Standard	m_{true}	0.164	0.194	0.875
Combined	Edge Correc. $\sigma = 0.18$	Standard	m_{true}	0.148	0.209	0.870
Combined	Helicity Corr. $\lambda = 0$	Standard	m_{true}	0.077	0.204	0.897
Combined	Helicity Corr. $\lambda = 500$	Standard	m_{true}	0.076	0.205	0.897
Combined	Helicity Corr. $\lambda = 1000$	Standard	m_{true}	0.076	0.206	0.895
Combined	Helicity Corr. $\lambda = 1800$	Standard	m_{true}	0.077	0.208	0.890
γ^*	MMC			0.085	0.225	0.908
Tauspinner	MMC			0.077	0.210	0.908
Spin0hh	MMC			0.101	0.213	0.908
Combined	MMC			0.082	0.211	0.908

Table 8.2: A summary of all the neural networks that are presented in this thesis. With the exception of the neural network trained with the edge corrected error loss, the neural network performance after the calibration is given.

CHAPTER 9

Conclusion

The focus of this thesis is the ditau mass reconstruction using regression neural networks trained on Monte-Carlo-generated ditau events that simulate the ATLAS detector. This is a difficult task due to the neutrinos that get created in every tau decay. In this thesis it is shown that RNNs can be used for the ditau mass reconstruction of such events. First, the stability of the training process is studied. A setup that creates stable training most of the time is found. However, the training does sometimes finish suboptimally. Because of this, three neural networks with different initial seeds are trained at the same time.

The reconstruction performance of neural networks is measured using figures of merit: the mean bias of the relative mass resolution in bins of the true mass b and the mean 68%-quantile width μ_b. At first there is a large bias in the relative mass resolution for the low and high regions of the mass range. This is because the neural networks learn the shape of the training samples and therefore create a bias in the reconstruction. A two step calibration can be used to remove most of the bias. When measuring the performance of neural networks after the calibration, it is shown that neural networks can match and, in some mass regions, exceed the performance of the MMC reconstruction. The relative mass resolution as a function of the true mass varies when using different samples for the neural network training. The training sample choice that leads to the smallest quantile widths is a combined sample of the tauspinner and spin0hh sample. However, the 68%-quantile widths of a neural network trained on this combined sample is larger for the $80 < m_{true} < 110\,\text{GeV}$ mass range. For lower and higher masses as well as for the entire mass range of the 95%-quantiles, neural networks are able to consistently create smaller quantile widths.

When applying these neural networks to a Standard Model Higgs sample and a Standard Model Z sample, where stricter cuts are applied, the separation power between the two can be measured using an ROC curve. For the MMC, the AUC of the ROC curve is 0.908. Using the combined tauspinner and spin0hh samples, an AUC of the ROC curve of 0.897 is achieved. This means that the neural network is not better able to differentiate between Higgs and Z events.

Furthermore, the effects of using different targets for the training are shown. Training on the difference between the true mass and the visible mass can be beneficial to the training, with a mean 68%-quantile width of $\mu_b = 0.215$ compared to $\mu_b = 0.220$ when using the true mass as a target. However, this is only the case if the true visible mass is considered, meaning when the detector resolution effects are ignored. Using the reconstructed visible mass leads to a deterioration of the relative mass resolution ($\mu_b = 0.225$). A similar effect is seen when training on the ratio of the true

mass to the visible mass. When the true visible mass is used in the ratio, a similar improvement in the relative mass resolution of the reconstructed ditau mass can be observed ($\mu_b = 0.215$), but when the reconstructed visible mass is used in the ratio instead, the mean quantile width of the relative mass resolution is larger ($\mu_b = 0.234$). Since the true visible mass is not a quantity that can be used for the ditau mass reconstruction of real data, it is concluded that the true mass is the optimal target of the five target choices.

The effect of different input variables for the neural network training is also demonstrated. Four different sets of variables are considered. The standard set, consisting of the transverse momentum p_T, the azimuthal angle ϕ, the pseudorapidity η, and the number of charged tracks of both reconstructed tau leptons, as well as the MET and azimuthal angle of the MET, is concluded to be the optimal choice of the four sets. Adding further variables, such as MET_{sig}, the pile-up, and the number of reconstructed hadronic jets, does not lead to an improvement in the reconstruction performance.

Apart from the squared error loss, two further loss functions are studied. The edge corrected loss function can be used as an alternative to the calibration, by rewarding the neural network for reconstructing events outside of the mass range of the training samples. It is able to reduce the bias, but a large part of the bias remains and the separation power between Higgs and Z events deteriorates. Therefore, using the squared error loss in combination with the calibration is a better approach to handling the bias. To remove a possible spin bias in the reconstruction, a helicity correlation loss function is also studied. It does not lead to an improved reconstruction and it negatively affects the AUC of the ROC curve. Therefore, it is concluded that training with the squared error loss leads to a better reconstruction performance.

Generally it can be said that the neural networks studied in this thesis are not able to create a better $\text{H} \rightarrow \tau\tau$ and $\text{Z} \rightarrow \tau\tau$ mass reconstruction than the MMC. It does, however, have better reconstruction capabilities than the MMC for masses larger than the Higgs mass. It is also a lot simpler to create a new training for a neural network than to crate a new parametrization for the MMC and the reconstruction speed of NNs is significantly faster than for the MMC. Going forward it is advised to apply a neural network in an analysis and measure the systematic effects of the reconstruction in comparison to the MMC.

Bibliography

[1] M. Tanabashi et al. (Particle Data Group), *Leptons*, 2019, URL: https://pdg.lbl.gov/2019/tables/rpp2019-sum-leptons.pdf.

[2] Morad Aaboud et al., *Measurement of $\tau\tau$ polarisation in $Z\gamma* \rightarrow \tau\tau$ decays in proton–proton collisions at $\sqrt{s}$= 8 TeV with the ATLAS detector*, The European Physical Journal C **78** (2018) 1.

[3] Georges Aad et al., *The ATLAS experiment at the CERN large hadron collider*, Journal of instrumentation **3** (2008).

[4] Georges et. al. Aad, *Operation and performance of the ATLAS semiconductor tracker. Operation and performance of the ATLAS semiconductor tracker*, JINST **9** (2014) P08009, arXiv: 1404.7473, URL: https://cds.cern.ch/record/1698966.

[5] E et. al. Abat, *The ATLAS TRT end-cap detectors*, Journal of Instrumentation **3** (2008) P10003.

[6] M Aharrouche et al., *Response uniformity of the ATLAS liquid argon electromagnetic calorimeter*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **582** (2007) 429.

[7] A Airapetian et al., *ATLAS detector and physics performance: Technical Design Report, 1*, CERN-LHCC-99-014, ATLAS-TDR-014, 1999.

[8] Ugur Akgun, *CMS HF calorimeter PMTS and Ξ+ C lifetime measurement*, PhD thesis: University of Iowa, 2003.

[9] Christian Bierlich et. al., *PYTHIA 8.3*, 2022, URL: https://pythia.org/.

[10] James Gillies et. al., *The coupling constants and unification of interactions*, URL: https://www.physicsmasterclasses.org/exercises/keyhole/en/projects/running_alphas.html.

[11] Simone Alioli et. al., *The POWHEG BOX*, 2022, URL: https://powhegbox.mib.infn.it.

[12] Tanabashi et. al., *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001, URL: https://pdg.lbl.gov/2019/tables/rpp2019-sum-gauge-higgs-bosons.pdf.

[13] Claude Amsler et al., *Review of particle physics*, Physics Letters B **667** (2008) 1.

[14] ATLAS Collaboration, *Observation of the gamma to tau tau process in Pb+Pb collisions and constraints on the tau-lepton anomalous magnetic moment with the ATLAS detector*, 2022, URL: https://arxiv.org/abs/2204.13478.

[15] *ATLAS tile calorimeter: Technical Design Report*, Technical design report. ATLAS, CERN, 1996, URL: http://cds.cern.ch/record/331062.

[16] Frank T Avignone III, Steven R Elliott and Jonathan Engel, *Double beta decay, Majorana neutrinos, and neutrino mass*, Reviews of Modern Physics **80** (2008) 481.

[17] Elisabetta Barberio, Bob van Eijk and Zbigniew Was, *PHOTOS—a universal Monte Carlo for QED radiative corrections in decays*, Computer Physics Communications **66** (1991) 115.

[18] Gerald W Bennett et al., *Final report of the E821 muon anomalous magnetic moment measurement at BNL*, Physical Review D **73** (2006) 072003.

[19] Christopher M Bishop et al., *Neural networks for pattern recognition*, Oxford university press, 1995.

[20] Enrico Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Physics **7** (2019) 034.

[21] Ian C. Brock, *Lecture notes for Physik V: Kerne und Teilchen*, 2019.

[22] Matteo Cacciari, Gavin P Salam and Gregory Soyez, *The anti-kt jet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063.

[23] ATLAS Electromagnetic Liquid Argon Endcap Calorimeter et al., *Construction, assembly and tests of the ATLAS electromagnetic end-cap calorimeters*, Journal of Instrumentation **3** (2008) P06002.

[24] F. Capozzi et al., *Neutrino masses and mixings: Status of known and unknown 3 ν parameters*, Nuclear Physics B **908** (2016) 218, URL: https://doi.org/10.1016%2Fj.nuclphysb.2016.02.016.

[25] Steven Carlip, *Quantum gravity: a progress report*, Reports on progress in physics **64** (2001) 885.

[26] Pedro Castro, *Dipole current, bending angle and beam energy in bunch compressors at TTF/VUV-FEL*, tech. rep., CM-P00050885, 2004.

[27] S Catani, F Fiorani and G Marchesini, *QCD coherence in initial state radiation*, Physics Letters B **234** (1990) 339.

[28] CERN, *CERN's accelerator complex*, 2022, URL: https://home.cern/science/accelerators/accelerator-complex.

[29] CERN, *Gluon fusion (ggF)*, 2022, URL: https://atlas.cern/glossary/gluon-fusion.

[30] CERN, *Higgs cross sections for HL-LHC and HE-LHC*, 2016, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HiggsEuropeanStrategy.

[31] CERN, *LHC The guide faq*, Brochure, 2021, URL: https://cds.cern.ch/record/2809109/files/CERN-Brochure-2021-004-Eng.pdf.

[32] CERN, *The Large Hadron Collider*, 2022, URL: https://www.home.cern/science/accelerators/large-hadron-collider.

[33] CERN, *The Large Hadron Collider*, 2022, URL: https://www.home.cern/science/accelerators.

[34] CERN, *The matter-antimatter asymmetry problem*, Accessed: 2022–11-14, URL: https://en.wikipedia.org/wiki/File:Weinberg_angle_(relation_between_coupling_constants).svg.

[35] CERN, *Vector boson fusion (VBF)*, 2022, URL: https://atlas.cern/glossary/vector-boson-fusion.

[36] Serguei Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30.

[37] ATLAS Collaboration, Physical Review Letters **125** (2020), URL: https://doi.org/10.1103%2Fphysrevlett.125.051801.

[38] ATLAS Collaboration et al., *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton-proton collisions*, arXiv preprint arXiv:2009.04986 (2020).

[39] ATLAS collaboration, *ATLAS: Calorimeter*, https://atlas.cern/Discover/Detector/Calorimeter, Accessed: 2022–11-14.

[40] ATLAS collaboration, *ATLAS: Magnet System*, https://atlas.cern/Discover/Detector/Magnet-System, Accessed: 2022–11-14.

[41] ATLAS collaboration, *ATLAS: Muon Spectrometer*, https://atlas.cern/Discover/Detector/Muon-Spectrometer, Accessed: 2022–11-14.

[42] ATLAS collaboration, *ATLAS: The Inner Detector*, https://atlas.cern/Discover/Detector/Inner-Detector, Accessed: 2022–11-14.

[43] ATLAS collaboration, *ATLAS: Trigger and Data Acquisition*, https://atlas.cern/Discover/Detector/Trigger-DAQ, Accessed: 2022–11-14.

[44] ATLAS collaboration et al., *Technical design report for the ATLAS inner tracker strip detector*, tech. rep., LHC/ATLAS Experiment, 2017.

[45] Ingrid Deigaard, *Measurement of the Tau Polarization in $Z\rightarrow \tau\tau$ Decays with the ATLAS Detector*, PhD thesis: Bohr Inst., 2012.

[46] Alexandre Deur, Stanley J. Brodsky and Guy F. de Téramond, *The QCD running coupling*, Progress in Particle and Nuclear Physics **90** (2016) 1, URL: https://doi.org/10.1016%2Fj.ppnp.2016.04.003.

[47] Alexandre Deur et al., *Experimental determination of the effective strong coupling constant*, Physics Letters B **650** (2007) 244.

[48] Luigi Di Lella and Carlo Rubbia, "The Discovery of the W and Z Particles", *60 Years of CERN Experiments and Discoveries*, World Scientific, 2015 137.

[49] PRESIDIDA POR EL EXCMO SR and D AMADOR SCHÜLLER PÉREZ, *EL CERN Y LA MEGACIENCIA THE CERN AND THE MEGASCIENCE*, REAL ACADEMIA NACIONAL DE MEDICINA ().

[50] A. Elagin et al., *A new mass reconstruction technique for resonances decaying to*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2011) 481, URL: https://doi.org/10.1016%2Fj.nima.2011.07.009.

[51] John et al Ellis, *A Historical Profile of the Higgs Boson*, https://cds.cern.ch/record/2012465/plots, Accessed: 2022–11-09, 2015.

[52] Despoina Evangelakou, "diTau Mass Reconstruction Methods focus on ATLAS", II. Physikalisches Institut, Georg-August-Universit at Göttingen, 2011.

[53] Kjell Magne Fauske, *TikZ Example: Neural Network*, https://texample.net/tikz/examples/neural-network/, Accessed: 2022–11-12, 2006.

[54] Paolo Francavilla, ATLAS Collaboration et al., "The ATLAS tile hadronic calorimeter performance at the LHC", *Journal of Physics: Conference Series*, vol. 404, 1, IOP Publishing, 2012 012007.

[55] Stefano Frixione, Paolo Nason and Carlo Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics **2007** (2007) 070, URL: https://doi.org/10.1088%2F1126-6708%2F2007%2F11%2F070.

[56] Camilla Galloni, *Hadronic Tau reconstruction and identification performance in ATLAS and CMS*, tech. rep., 2018.

[57] Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning*, http://www.deeplearningbook.org, MIT Press, 2016.

[58] Particle Data Group, *Particle Physics Booklet*, 2019, URL: https://pdg.lbl.gov.

[59] Chris Hays, *Higgs boson measurements at ATLAS*, tech. rep., ATL-COM-PHYS-2016-1765, 2016.

[60] Universität Heidelberg, *Electron-positron annihilation*, https://uebungen.physik.uni-heidelberg.de/c/image/f/vorlesung/20172/818/Chapter16_slides.pdf, Accessed: 2022–11-11, 2020.

[61] Lena Herrmann, *Optimization of a Regression-Based Deep Neural Network for Di-Tau Mass Reconstruction in ATLAS*, Master's Thesis: Physikalisches Institut Universität Bonn, 2020, URL: https://web.physik.uni-bonn.de/group/view.php?&group=1&lang=de&c=t&id=121.

[62] Peter W Higgs, *Broken symmetries and the masses of gauge bosons*, Physical Review Letters **13** (1964) 508.

[63] Michael Hübner, *Effects of tau decay product reconstruction in a Higgs CP analysis with the ATLAS experiment*, 2016, URL: https://web.physik.uni-bonn.de/group/view.php?&group=1&lang=de&c=t&id=81.

[64] Robert L Jaffe, *Quark confinement*, Nature **268** (1977) 201.

[65] Jeremy Jordan, *Normalizing your data*, https://www.jeremyjordan.me/batch-normalization/, Accessed: 2022–11-14, 2018.

[66] Borut Paul Kersevan and Elzbieta Richter-Wacs, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, Computer Physics Communications **184** (2013) 919, URL: https://doi.org/10.1016%2Fj.cpc.2012.10.032.

[67] Diederik P Kingma and Jimmy Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980 (2014).

[68] George Kirakosyan, *Deduction of coupling constant (a= 1/137) as a wave peculiarity: Possible laboratory confirmation*, Physics essays **28** (2015) 283.

[69] Hagen Kleinert, *Particles and quantum fields*, world scientific, 2016 1457.

[70] Jeannette Lawrence, *Introduction to neural networks*, California Scientific Software, 1993.

[71] Michelangelo L Mangano, *QCD and the physics of Hadronic Collisions*, tech. rep., CERN, 2009.

[72] Maschen, *Weinberg angle (relation between coupling constants)*, URL: https://en.wikipedia.org/wiki/File:Weinberg_angle_(relation_between_coupling_constants).svg.

[73] C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*, ed. by Misner, C. W., Thorne, K. S., & Wheeler, J. A., 1973.

[74] Cush MissMJ, *Standard Model of Elementary Particles*, URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[75] C.R. Nave, *Coupling Constants for the Fundamental Forces*, 2022, URL: http://hyperphysics.phy-astr.gsu.edu/hbase/Forces/couple.html.

[76] C. Oleari, *The POWHEG BOX*, Nuclear Physics B - Proceedings Supplements **205-206** (2010) 36, URL: https://doi.org/10.1016%2Fj.nuclphysbps.2010.08.016.

[77] Rhys Edward Owen, *The ATLAS Trigger System*, (2018), URL: https://cds.cern.ch/record/2302730.

[78] Joao Pequenao, *Event Cross Section in a computer generated image of the ATLAS detector.* https://cds.cern.ch/record/1096081, Accessed: 2022–11-11, 2008.

[79] Michael Edward Peskin and Daniel V. Schroeder, *An Introduction to Quantum Field Theory*, Reading, USA: Addison-Wesley (1995) 842 p, Westview Press, 1995.

[80] James L. Pinfold, *The MoEDAL - MAPP Detector*, https://indico.cern.ch/event/868940/contributions/3815870/attachments/2081551/3496382/MoEDAL-Talk-ICHEP-2020.pdf, Accessed: 2022–11-14, 2020.

[81] Lokal Profil, *Quark confinement*, URL: https://commons.wikimedia.org/wiki/File:Quark_confinement.svg.

[82] T. Przedzinski, E. Richter-Was and Z. Was, *Documentation of TauSpinner algorithms: program for simulating spin effects in tau tau-lepton production at LHC*, The European Physical Journal C **79** (2019), URL: https://doi.org/10.1140%2Fepjc%2Fs10052-018-6527-0.

[83] *Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data*, tech. rep., CERN, 2022, URL: https://cds.cern.ch/record/2827111.

[84] Giulia Ricciardi and Marcello Rotondo, *Determination of the Cabibbo-Kobayashi-Maskawa matrix element V_cb*, Journal of Physics G: Nuclear and Particle Physics (2020), URL: https://doi.org/10.1088%2F1361-6471%2Fab9f01.

[85] Michaela Roggendorf, *Studies of Tau-Lepton Polarisation in Decays of Higgs and Z Bosons with the ATLAS Experiment*, MA thesis: Bonn U., 2014.

[86] David E Rumelhart et al., *Backpropagation: The basic theory*, Backpropagation: Theory, architectures and applications (1995) 289.

[87] Steven Schramm, "Atlas jet reconstruction, calibration, and tagging of lorentzboosted objects", *EPJ Web of Conferences*, vol. 182, EDP Sciences, 2018 02113.

[88] Sagar Sharma, Simone Sharma and Anidhya Athaiya, *Activation functions in neural networks*, towards data science **6** (2017) 310.

[89] *Sherpa Generator*, https://gitlab.cern.ch/atlas/athena/-/tree/master/Generators/Sherpa_i, 2022.

[90] Masaharu Tanabashi et al., *Review of particle physics*, Physical Review D **98** (2018) 030001.

[91] HHl Ten Kate, *Superconducting magnet system for the ATLAS detector at CERN*, IEEE transactions on applied superconductivity **9** (1999) 841.

[92] Mark Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 978-1-107-03426-6.

[93] Martin Werres, *Estimating the Mass of Di-Tau Systems in the ATLAS Experiment Using Neural Network Regression*, 2019, URL: https://web.physik.uni-bonn.de/group/view.php?&group=1&lang=de&c=t&id=102.

[94] Akira Yamamoto et al., *Design and development of the ATLAS central solenoid magnet*, IEEE Transactions on applied superconductivity **9** (1999) 852.

APPENDIX A

Variable Distributions

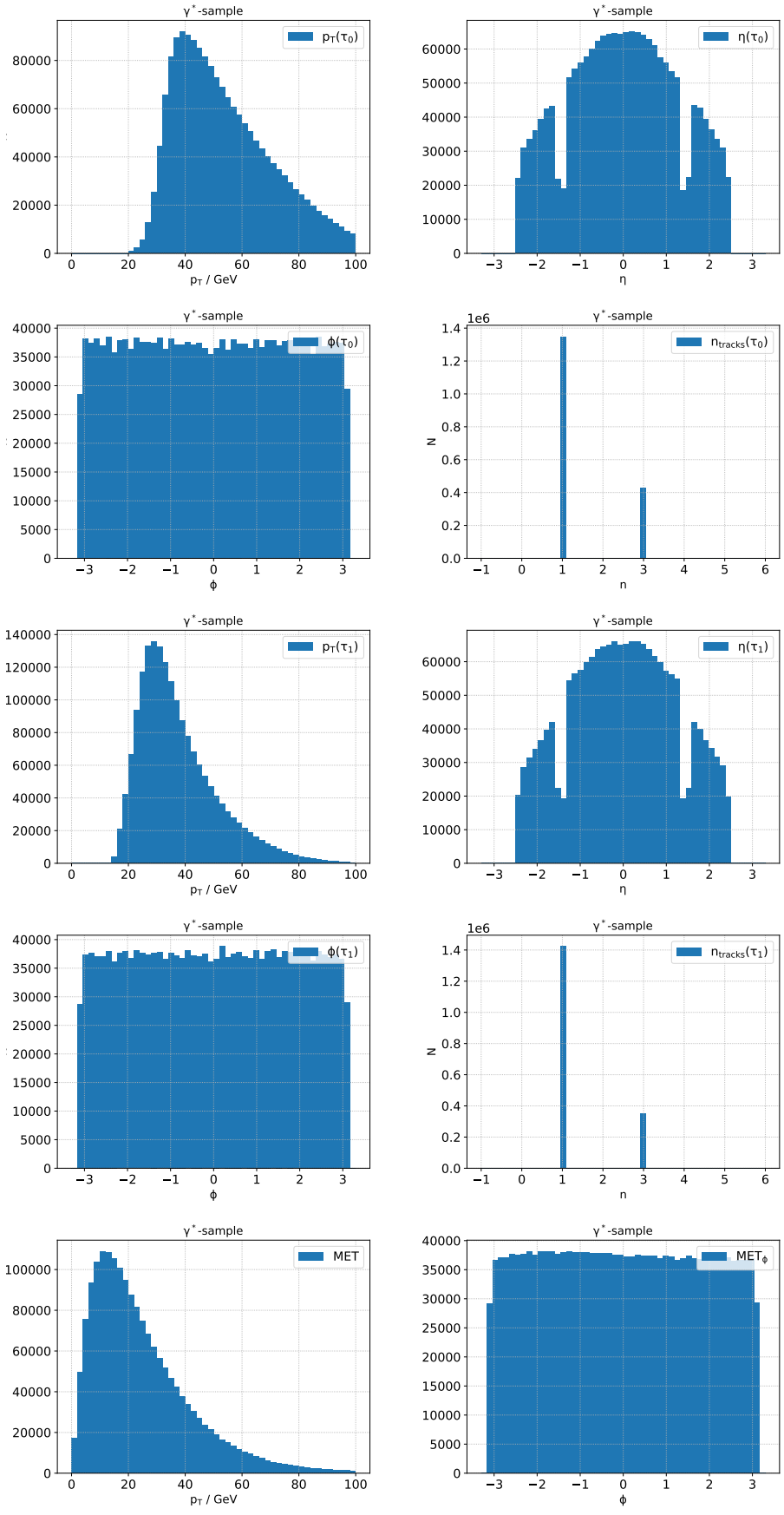


Figure A.1: Distributions of the variables of the γ^*-sample used in the standard set.

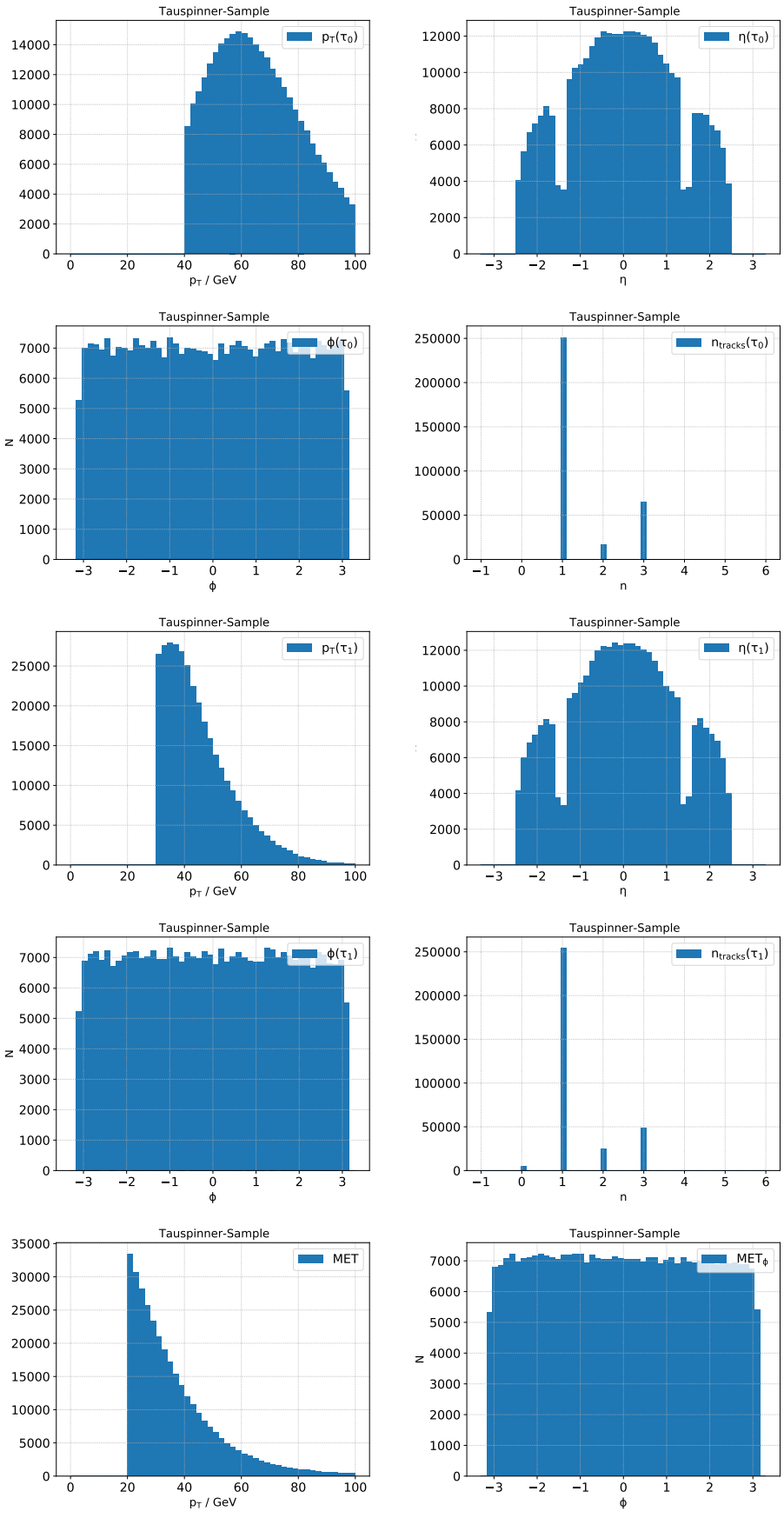


Figure A.2: Distributions of the variables of the tauspinner sample used in the standard set.

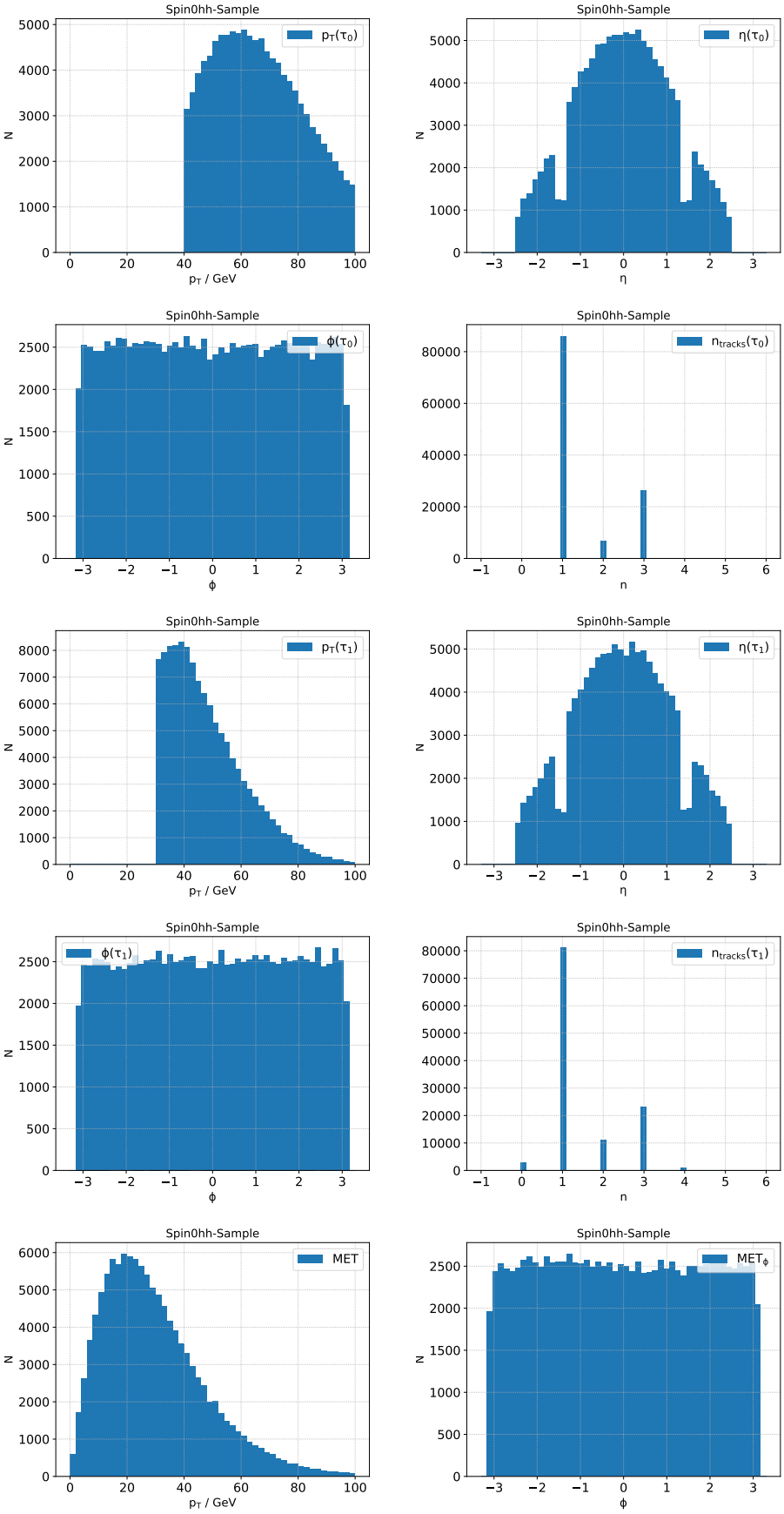


Figure A.3: Distributions of the variables of the spin0hh sample used in the standard set.

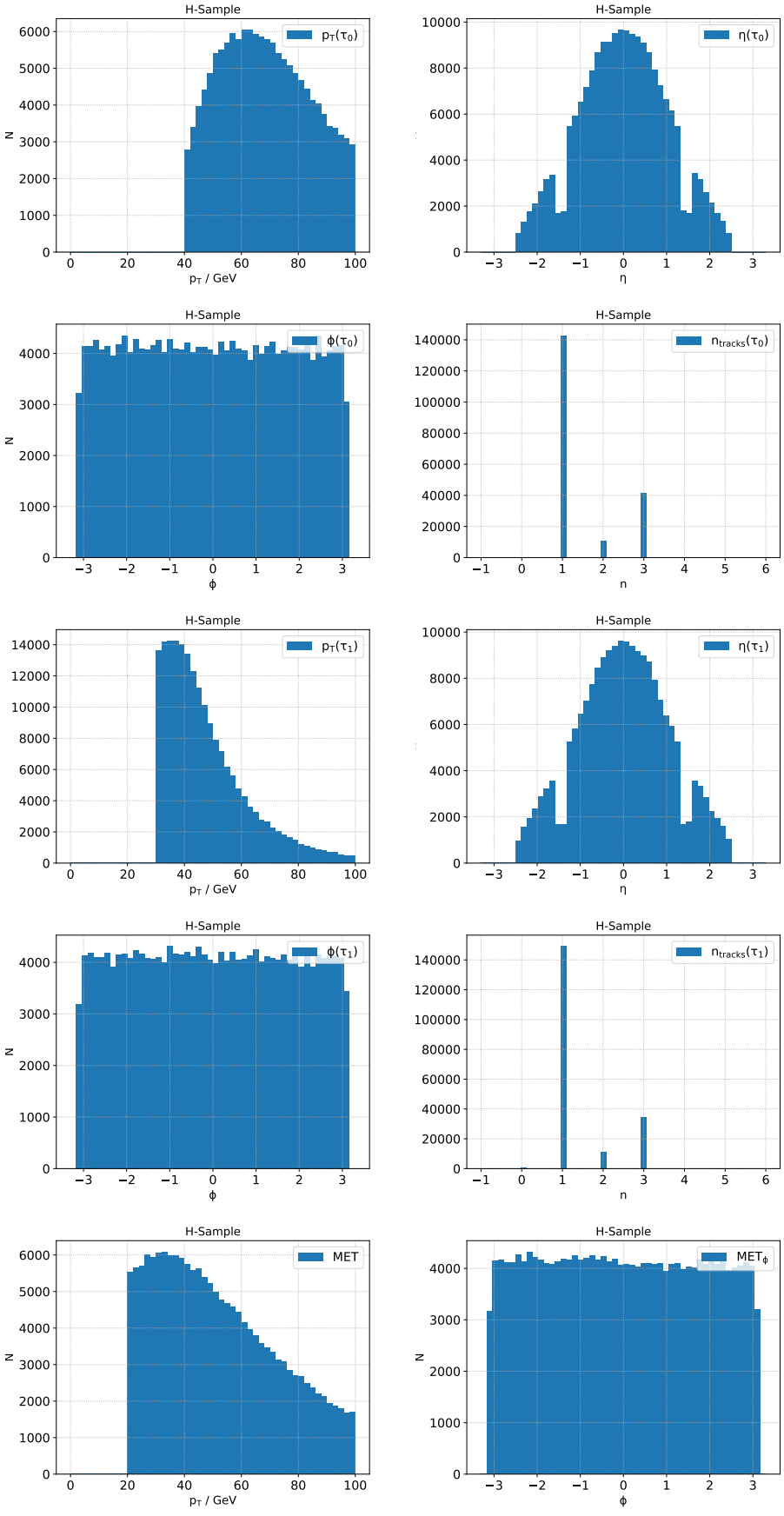


Figure A.4: Distributions of the variables of the H-sample used in the standard set.

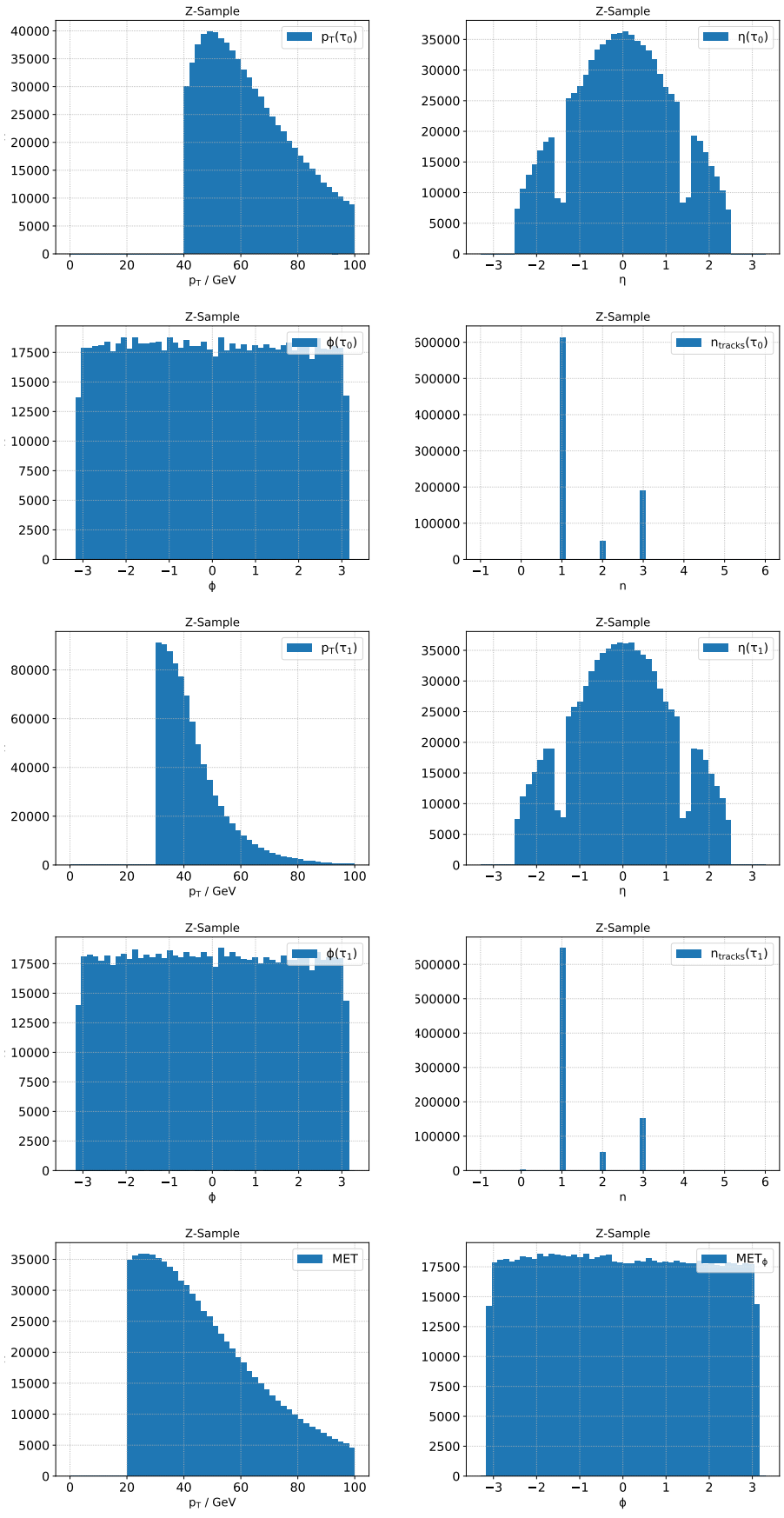


Figure A.5: Distributions of the variables of the Z-sample used in the standard set.

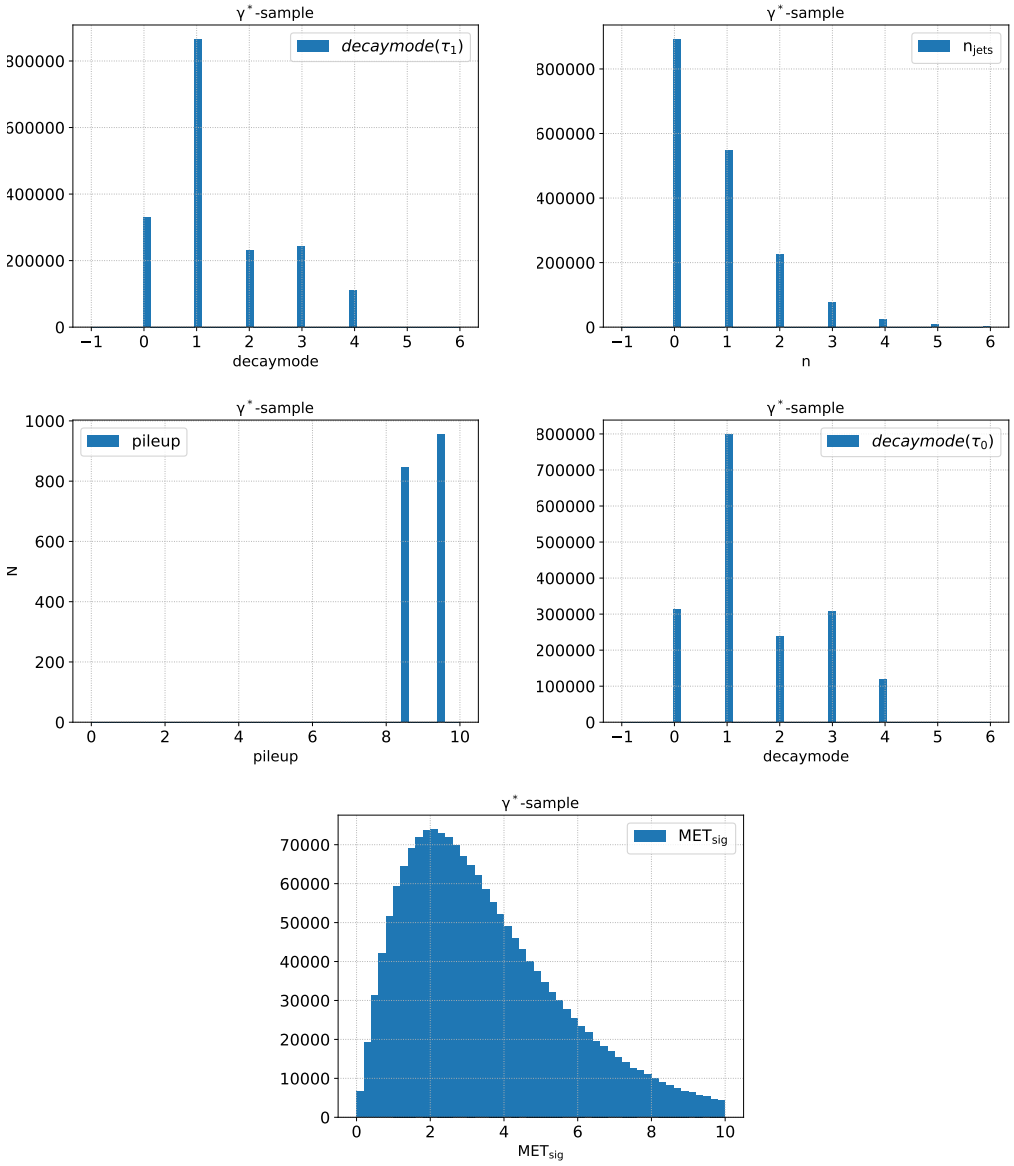


Figure A.6: Distributions of the variables of the γ^*-sample used in the Auxiliary set.

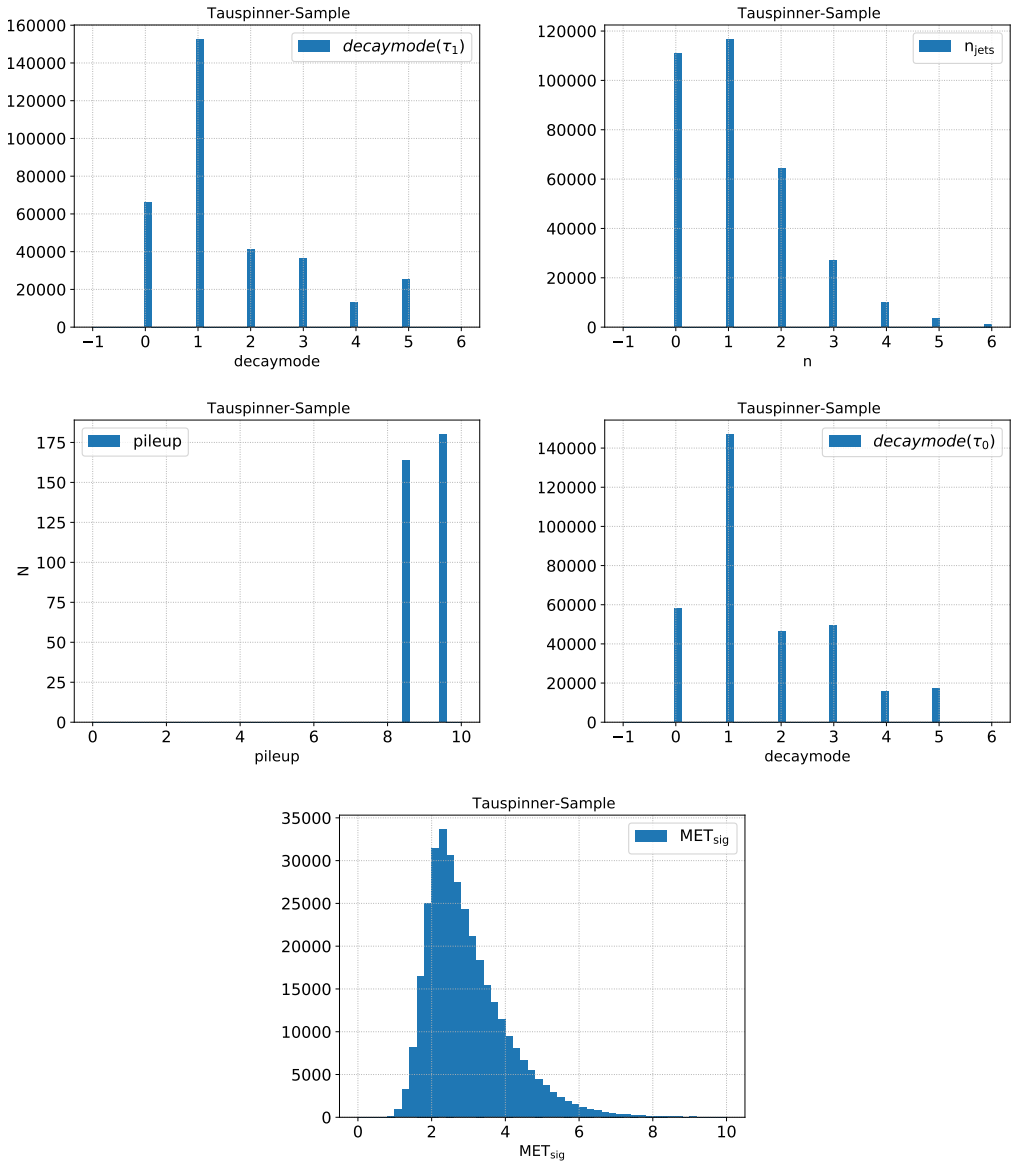


Figure A.7: Distributions of the variables of the tauspinner sample used in the Auxiliary set.

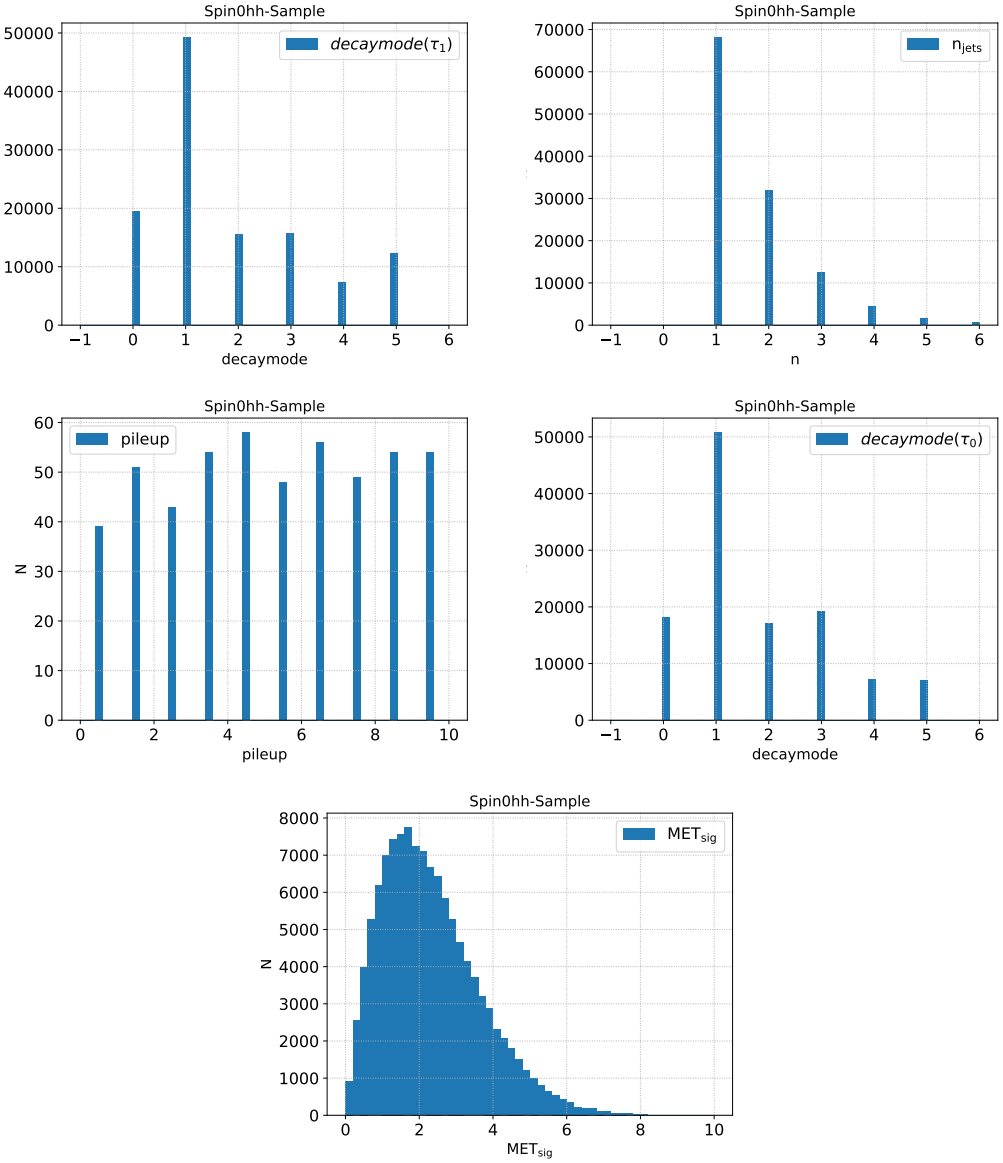


Figure A.8: Distributions of the variables of the spin0hh sample used in the Auxiliary set.

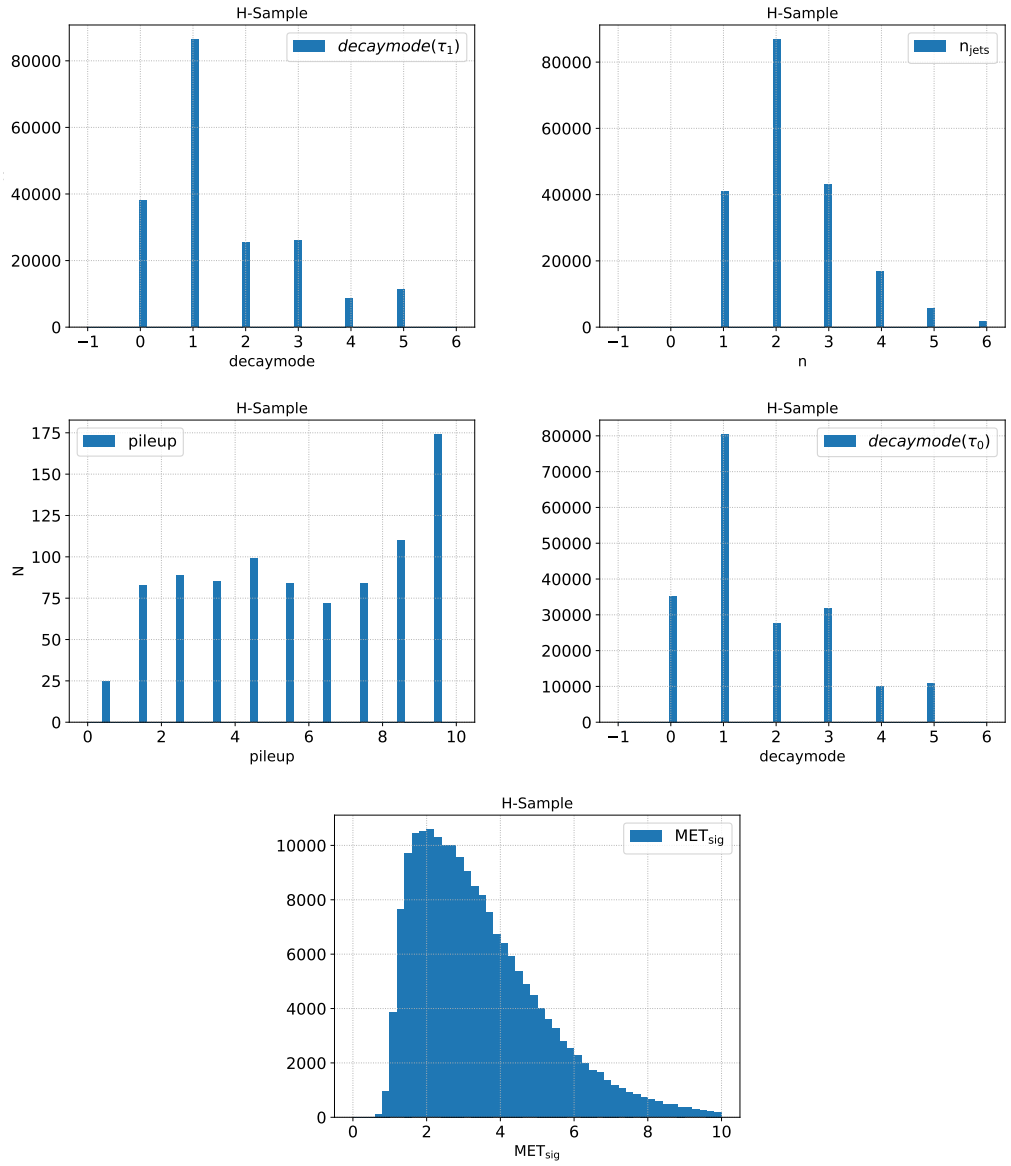


Figure A.9: Distributions of the variables of the H-sample used in the Auxiliary set.

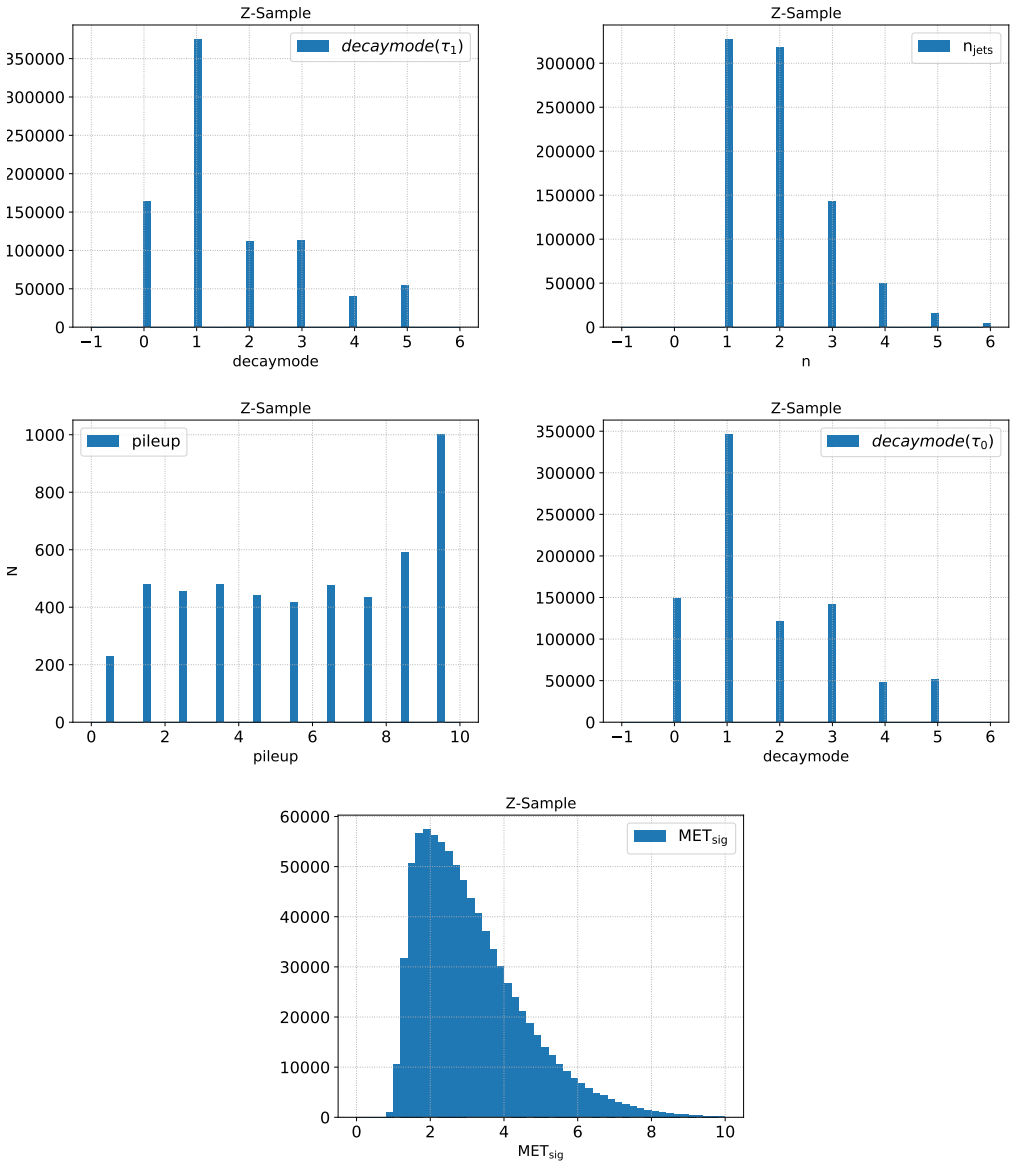


Figure A.10: Distributions of the variables of the Z-sample used in the Auxiliary set.

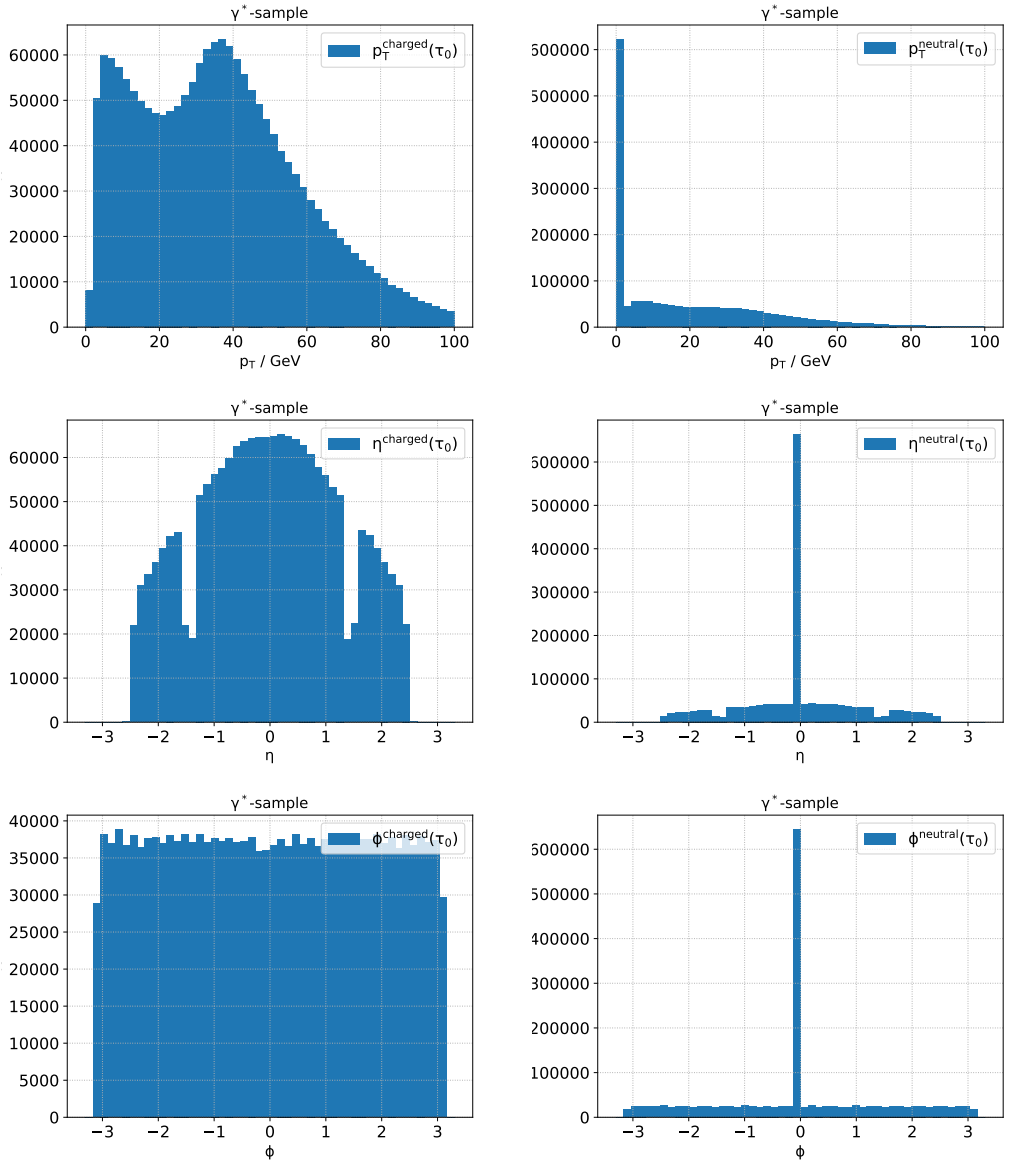


Figure A.11: Distributions of the variables of the γ^*-sample used in the AuxiliaryExtra set (1/2).

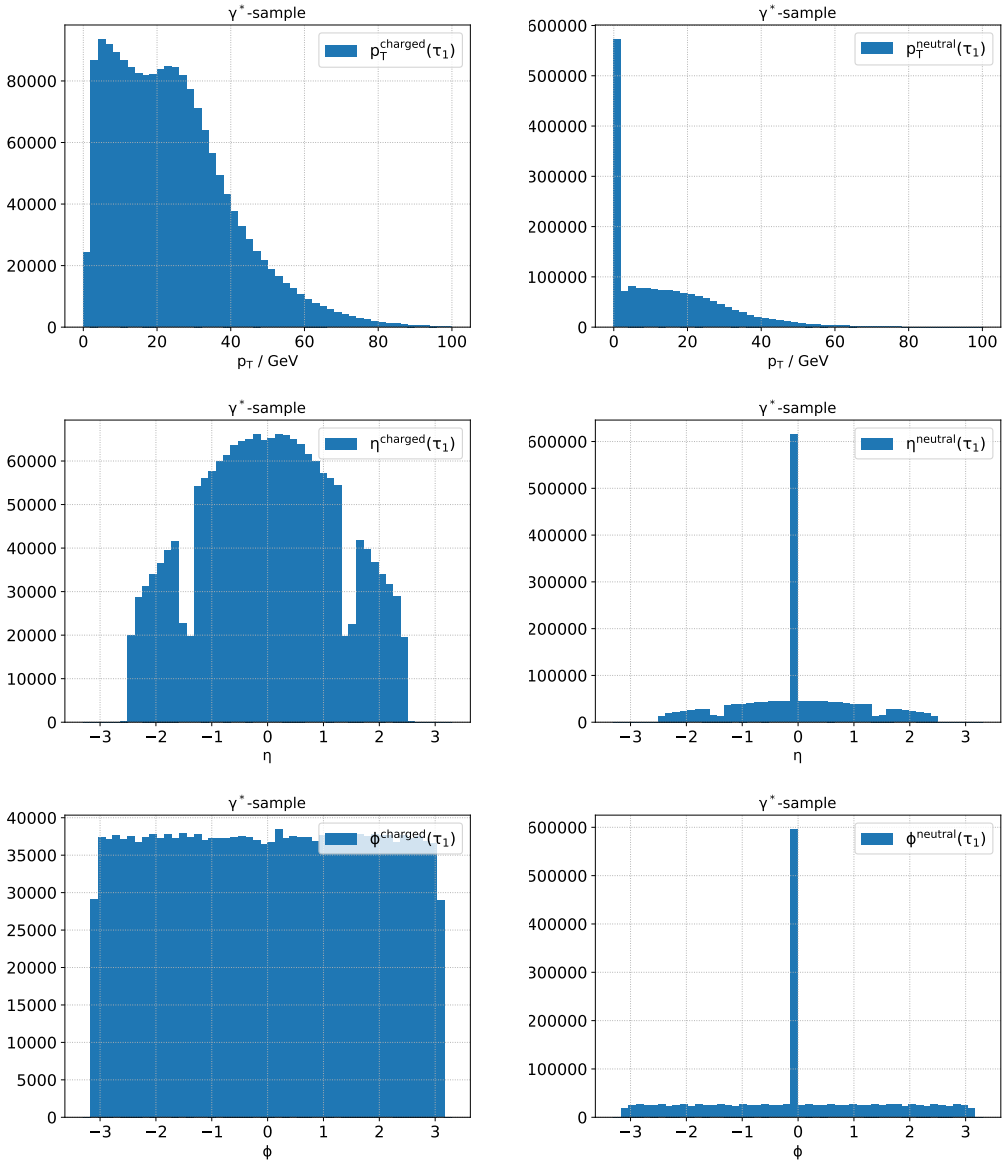


Figure A.12: Distributions of the variables of the γ^*-sample used in the AuxiliaryExtra set (2/2).

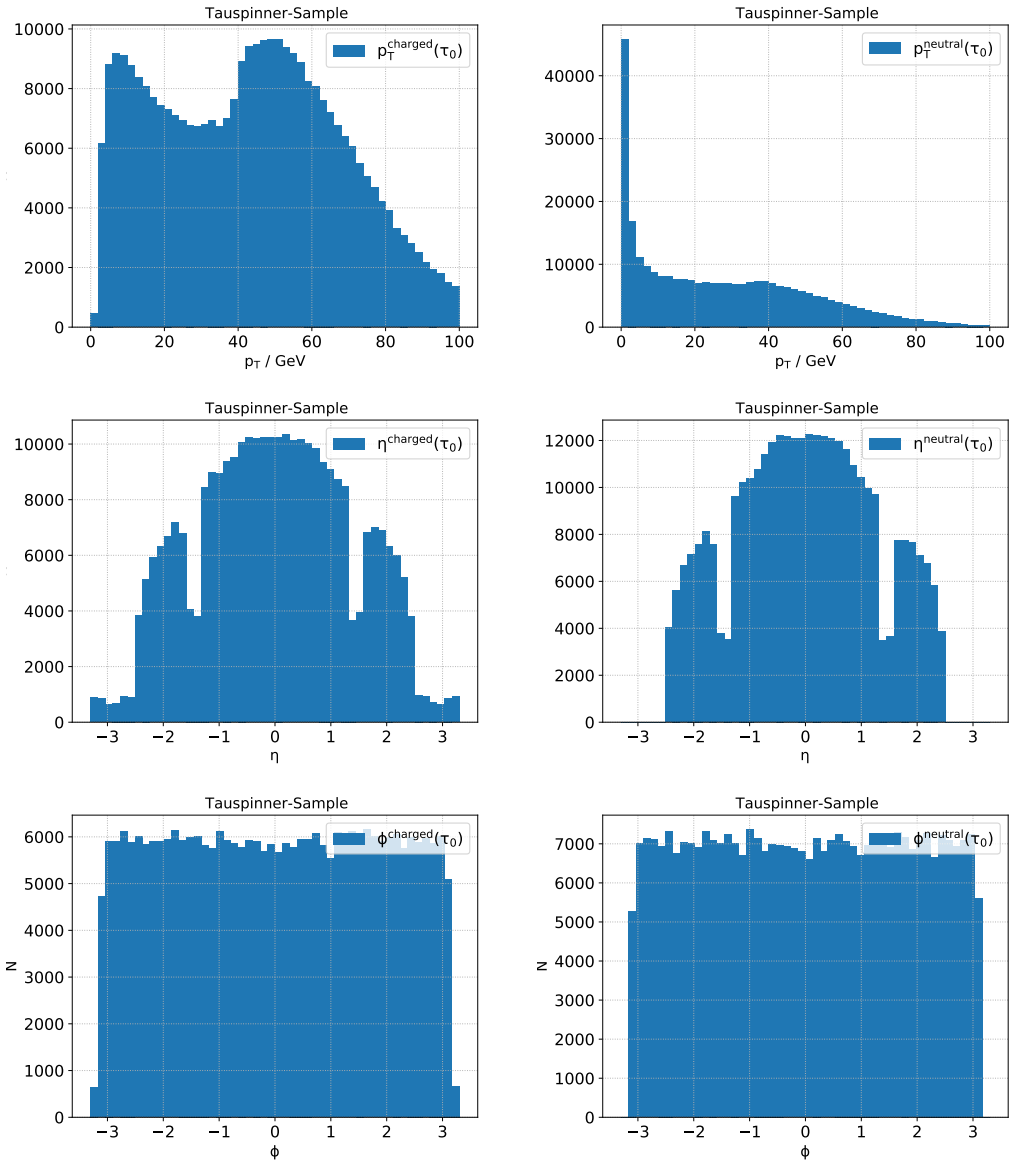


Figure A.13: Distributions of the variables of the tauspinner sample used in the AuxiliaryExtra set (1/2).

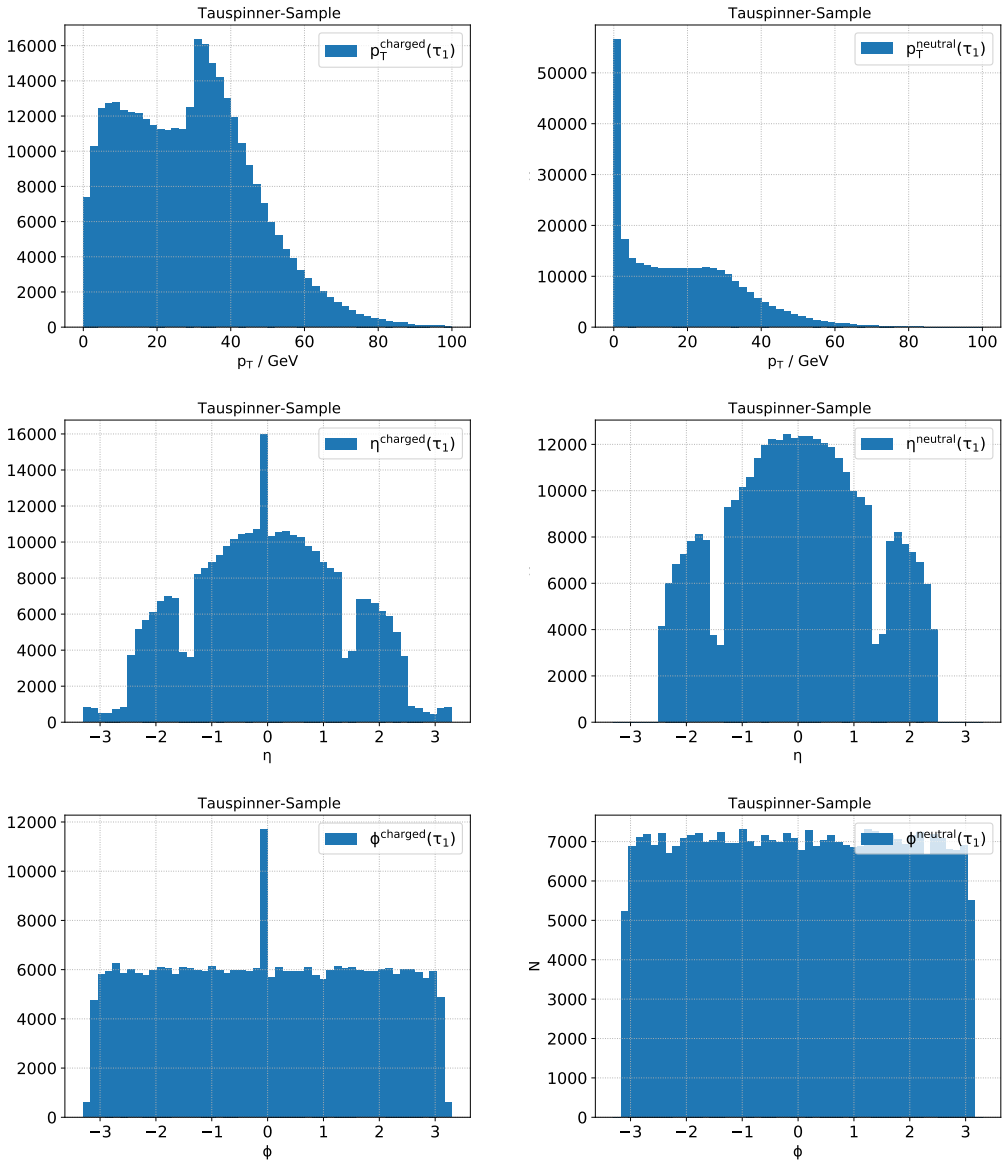


Figure A.14: Distributions of the variables of the tauspinner sample used in the AuxiliaryExtra set (2/2).

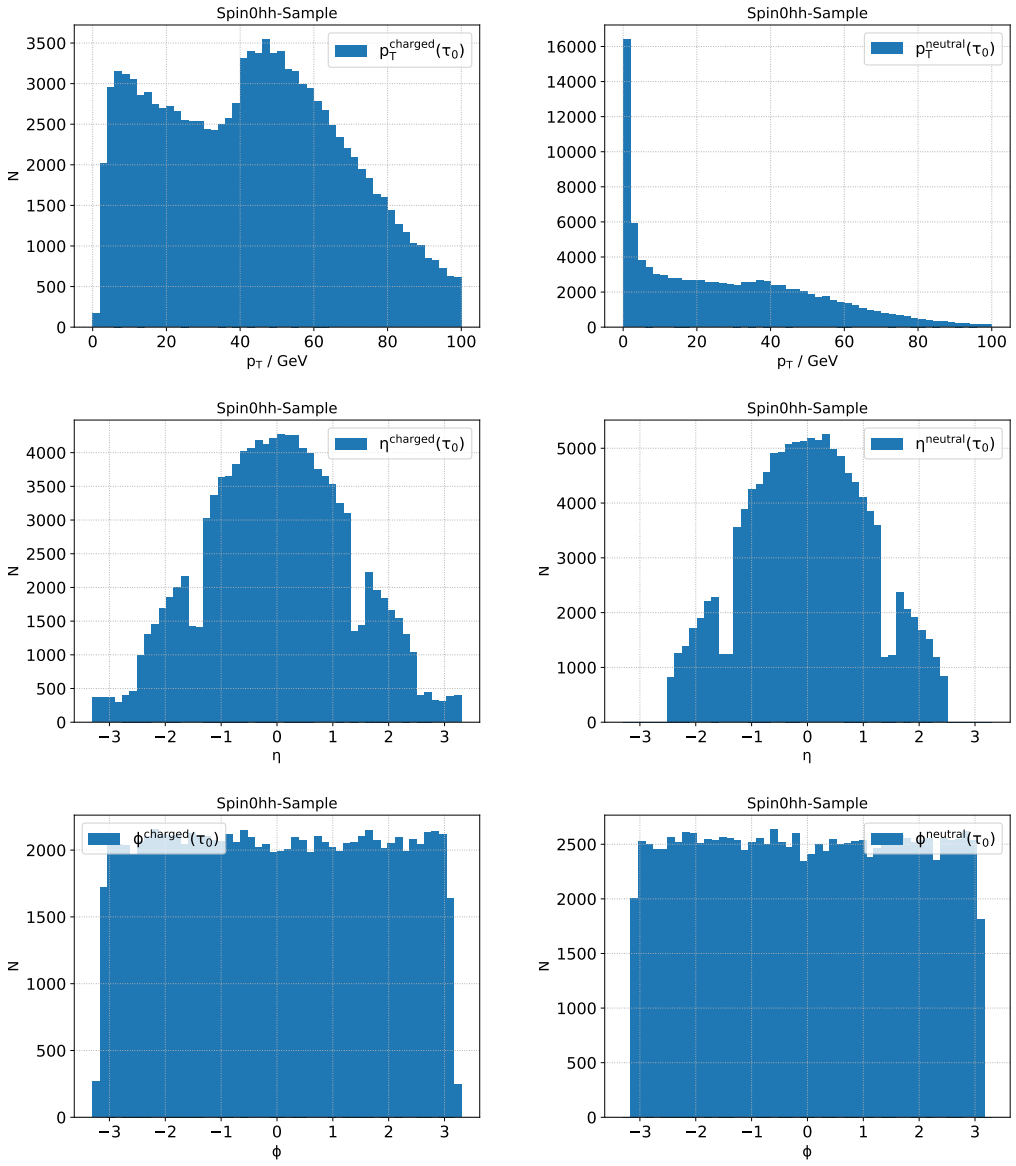


Figure A.15: Distributions of the variables of the spin0hh sample used in the AuxiliaryExtra set (1/2).

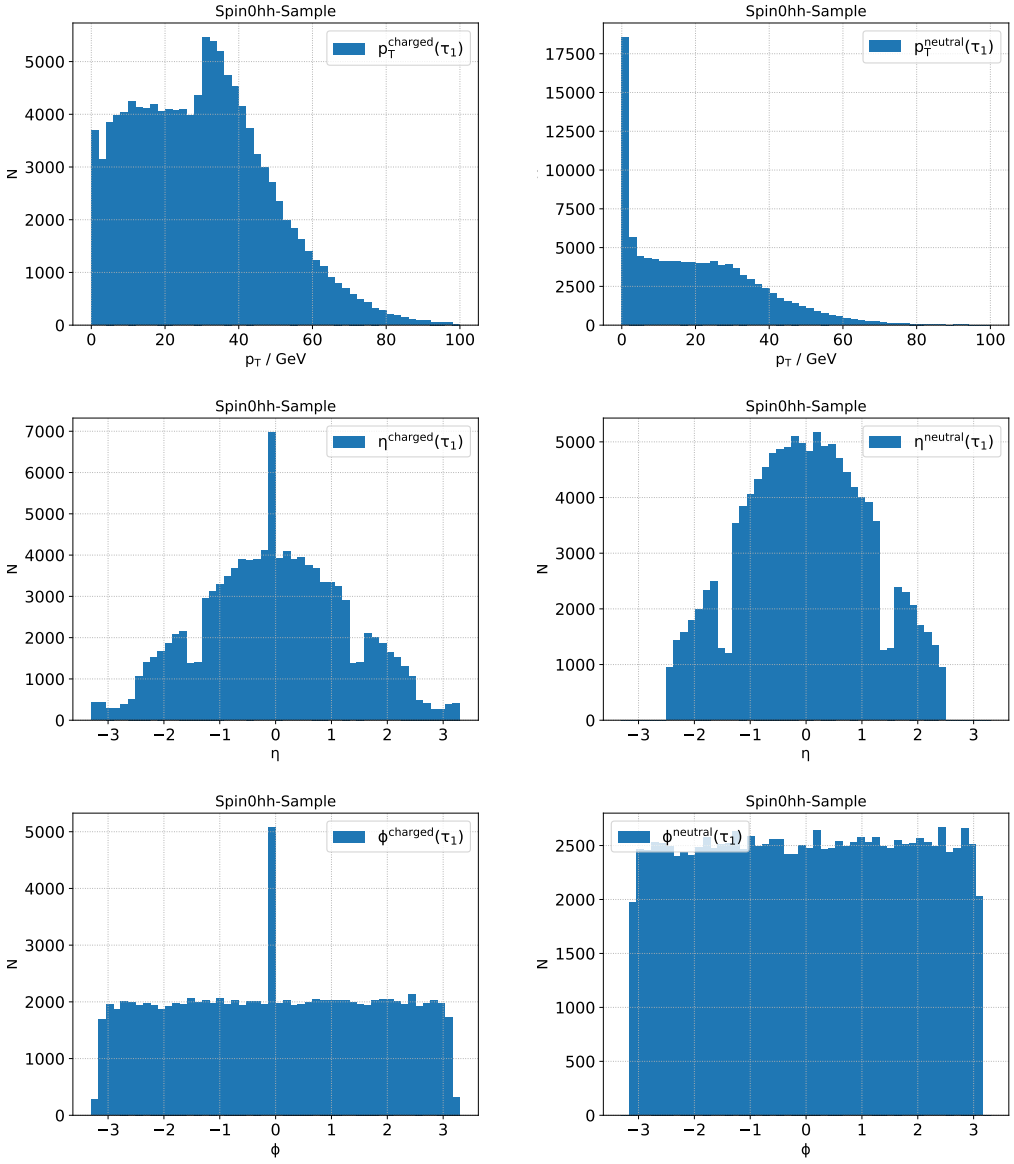


Figure A.16: Distributions of the variables of the spin0hh sample used in the AuxiliaryExtra set (2/2).

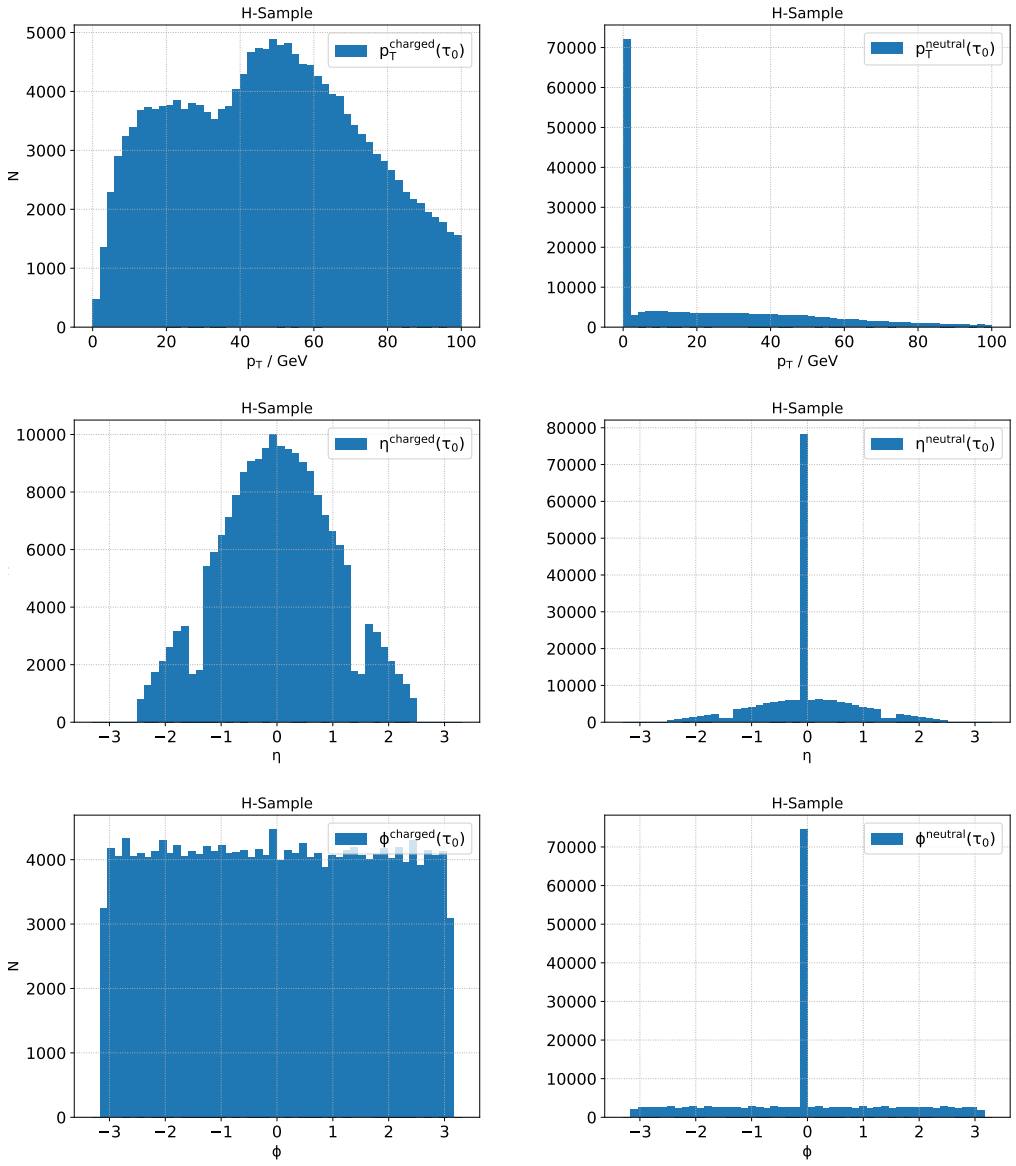


Figure A.17: Distributions of the variables of the H-sample used in the AuxiliaryExtra set (1/2).

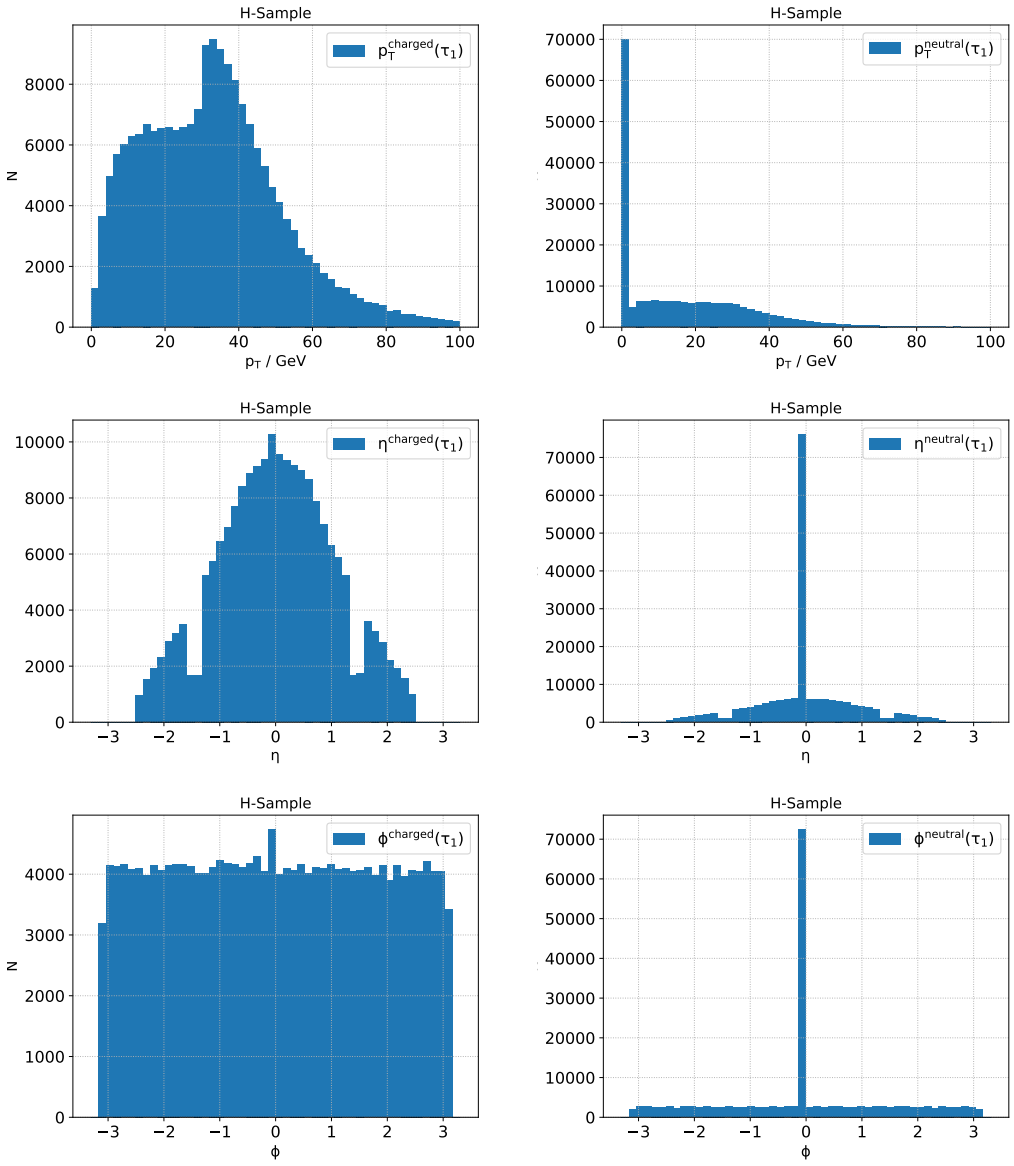


Figure A.18: Distributions of the variables of the H-sample used in the AuxiliaryExtra set (2/2).

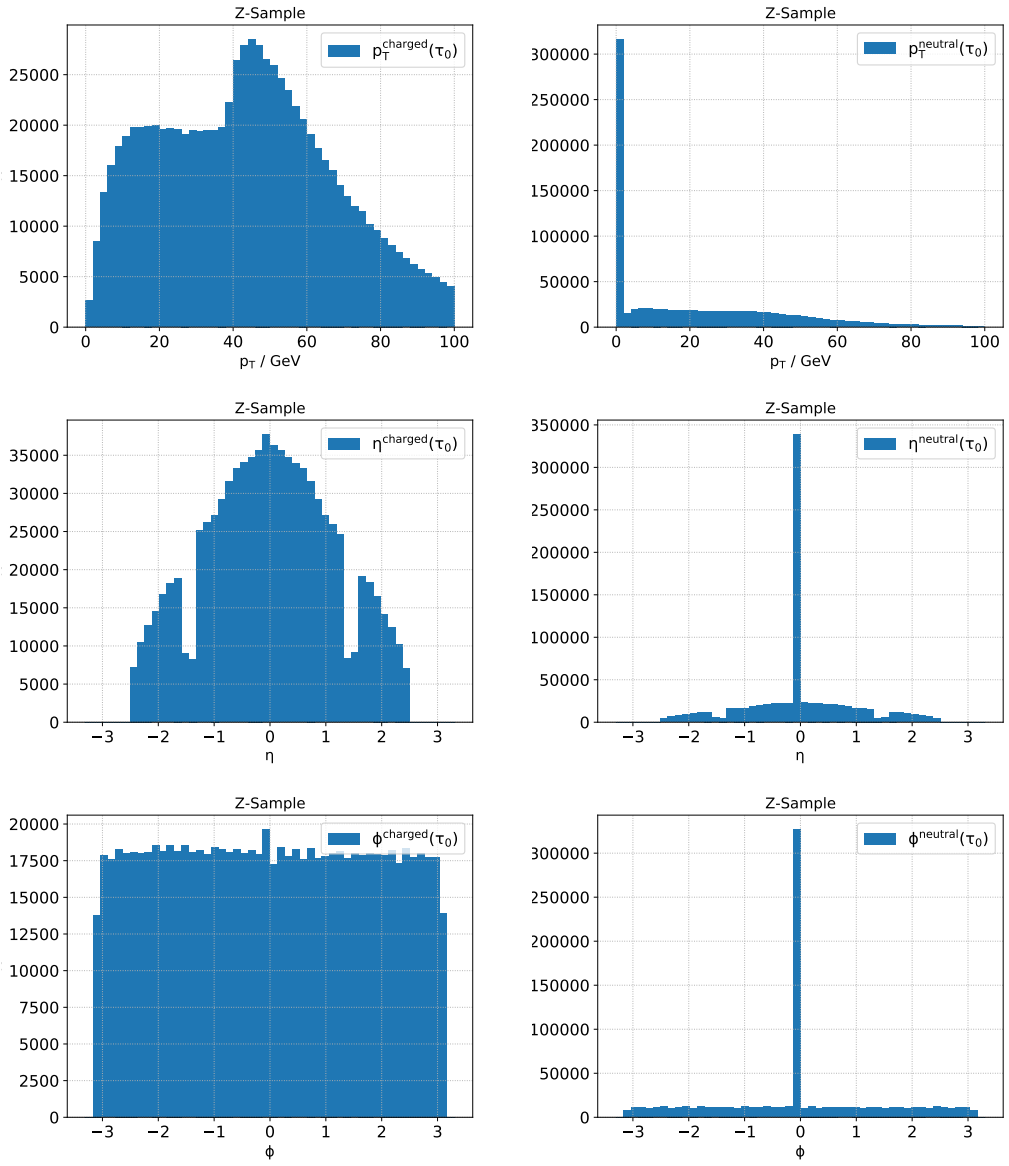


Figure A.19: Distributions of the variables of the Z-sample used in the AuxiliaryExtra set (1/2).

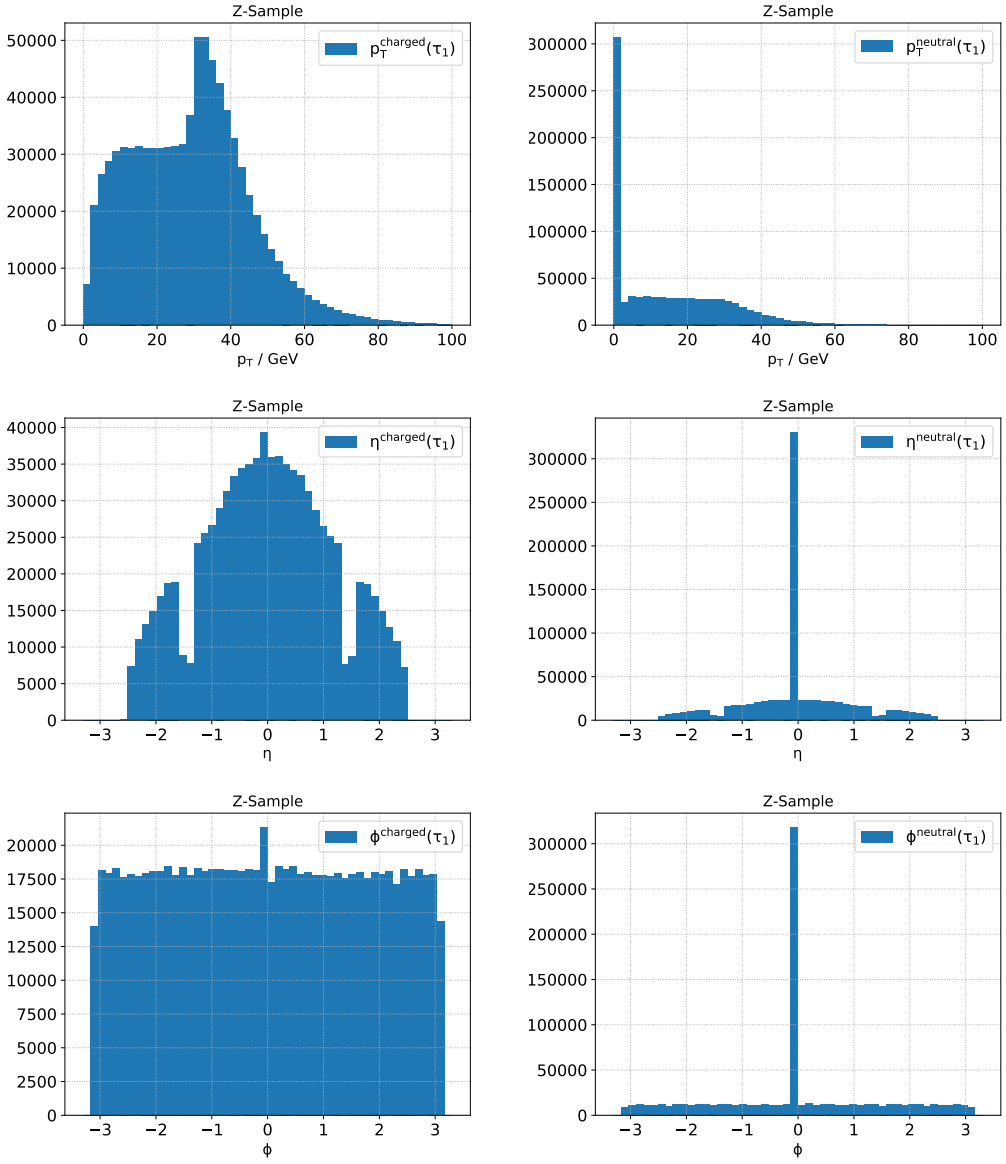


Figure A.20: Distributions of the variables of the Z-sample used in the AuxiliaryExtra set (2/2).

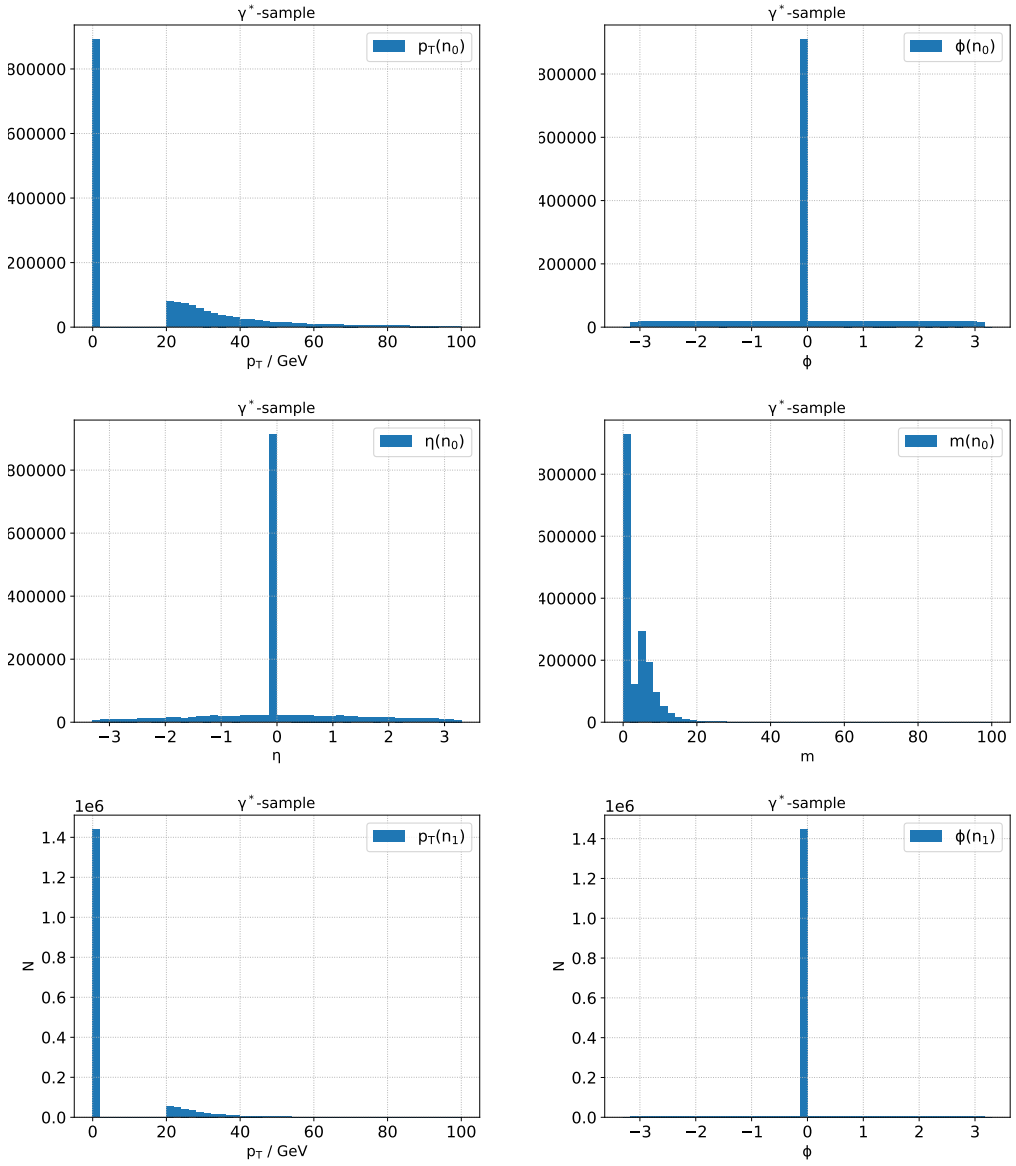


Figure A.21: Distributions of the variables of the γ^*-sample used in the Jets set (1/2).

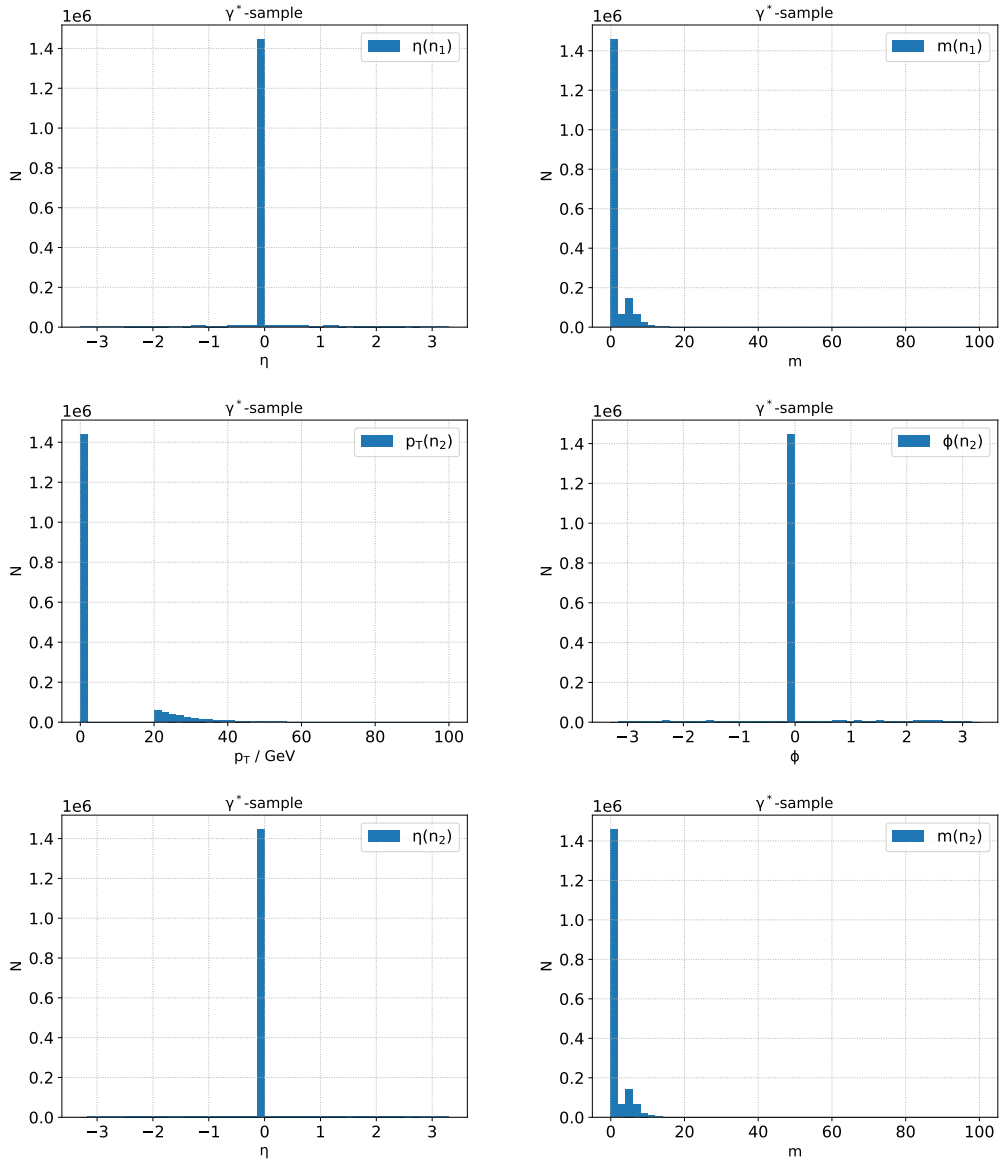


Figure A.22: Distributions of the variables of the γ^*-sample used in the Jets set (2/2).

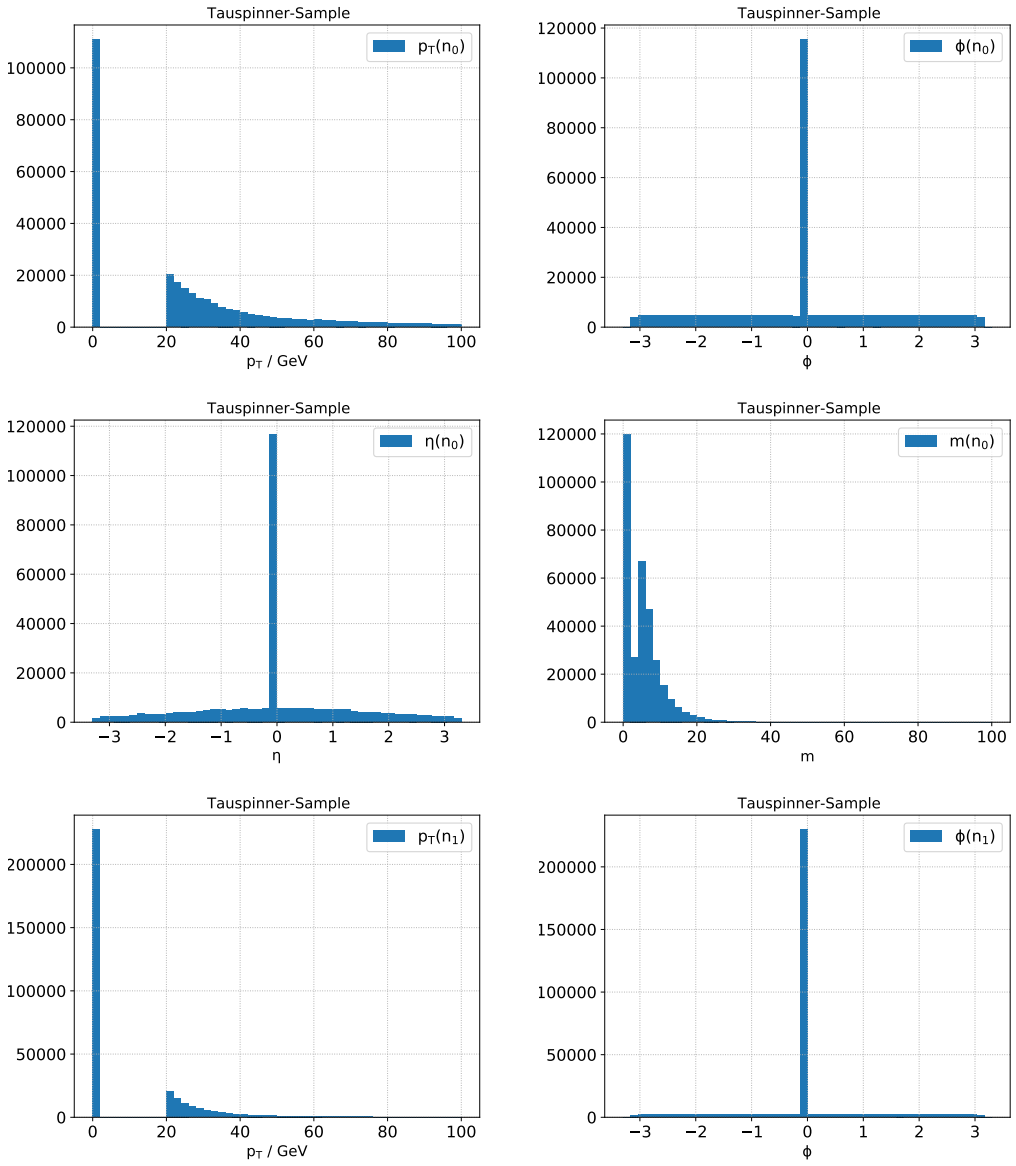


Figure A.23: Distributions of the variables of the tauspinner sample used in the Jets set (1/2).

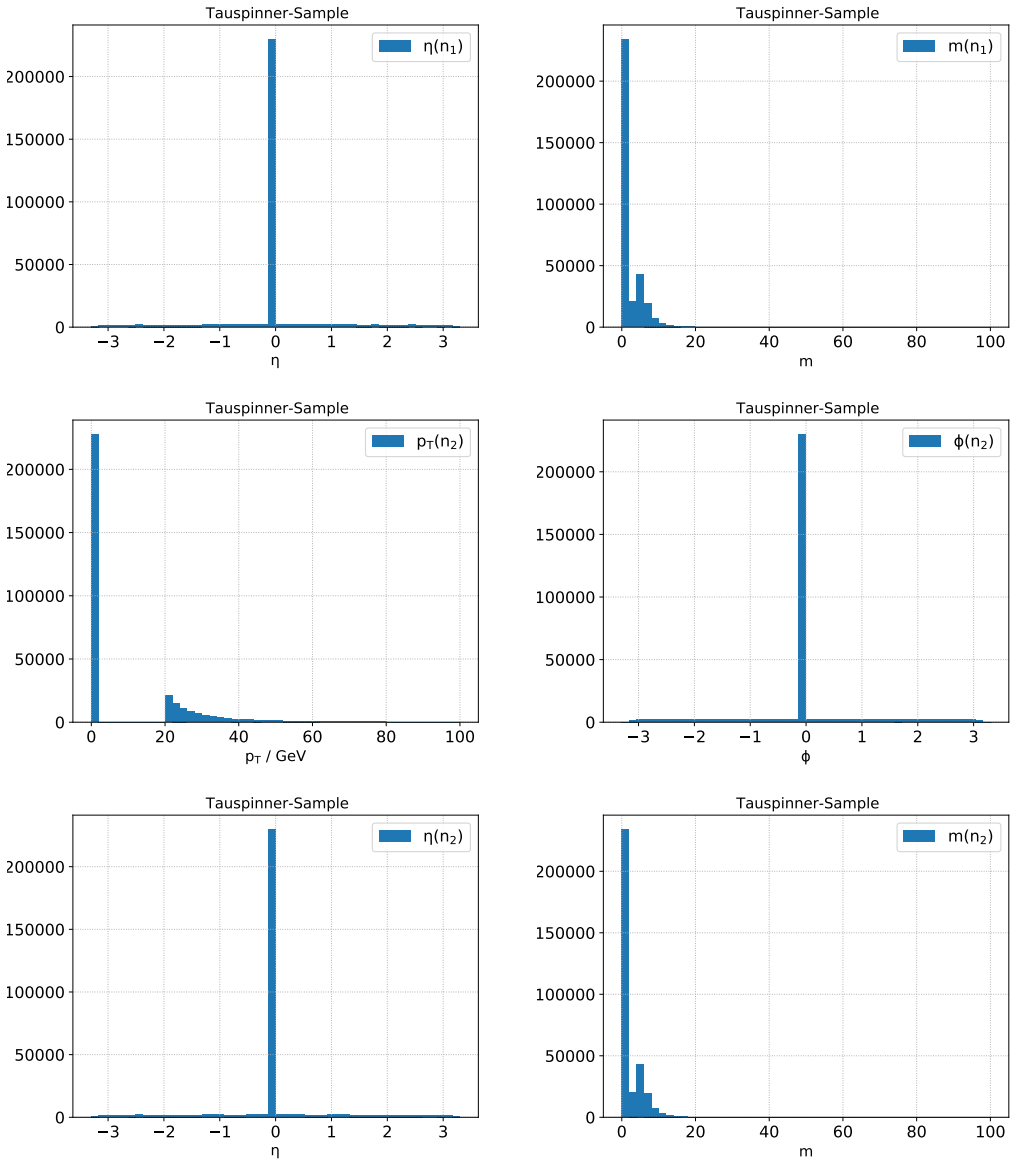


Figure A.24: Distributions of the variables of the tauspinner sample used in the Jets set (2/2).

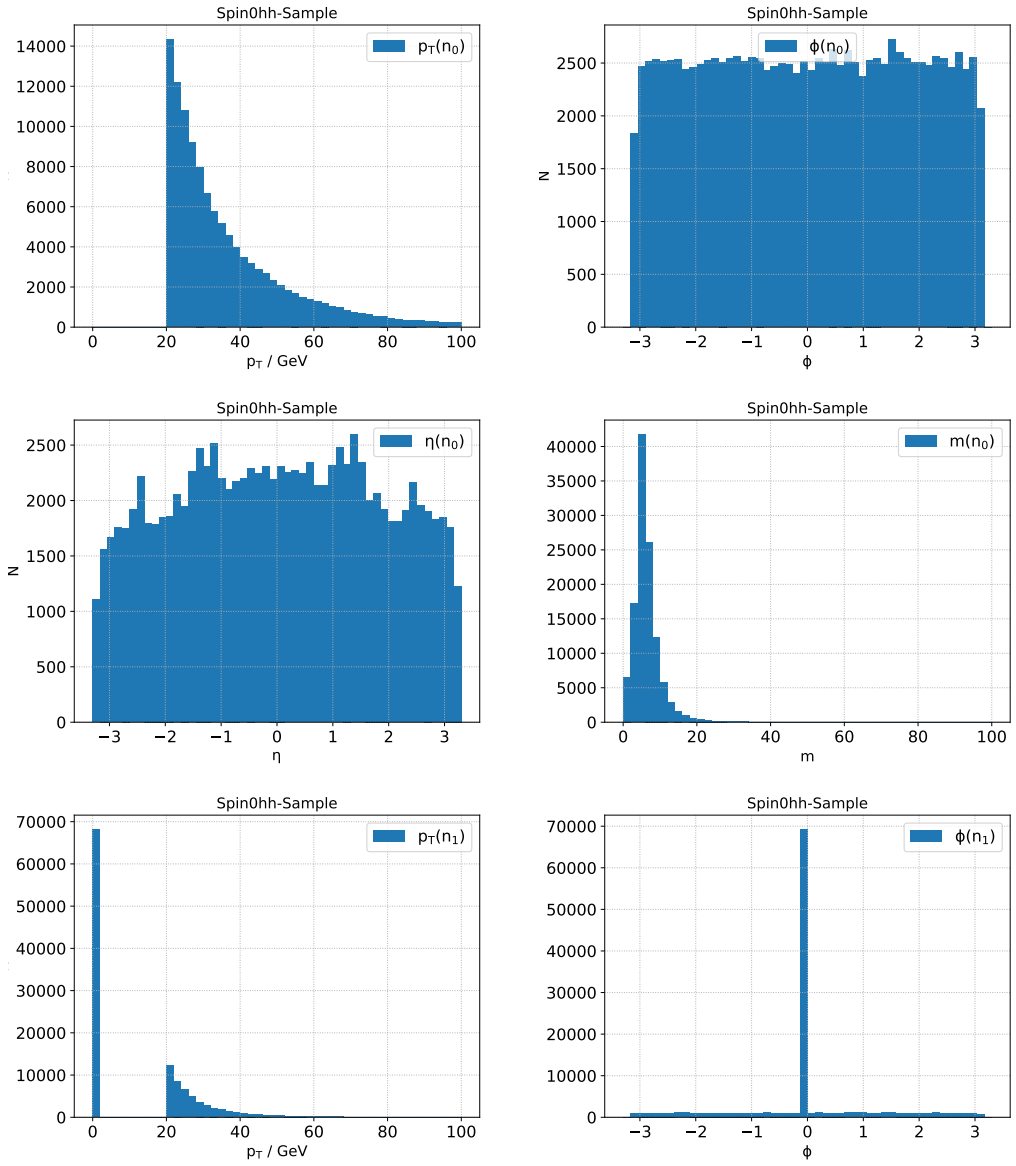


Figure A.25: Distributions of the variables of the spin0hh sample used in the Jets set (1/2).

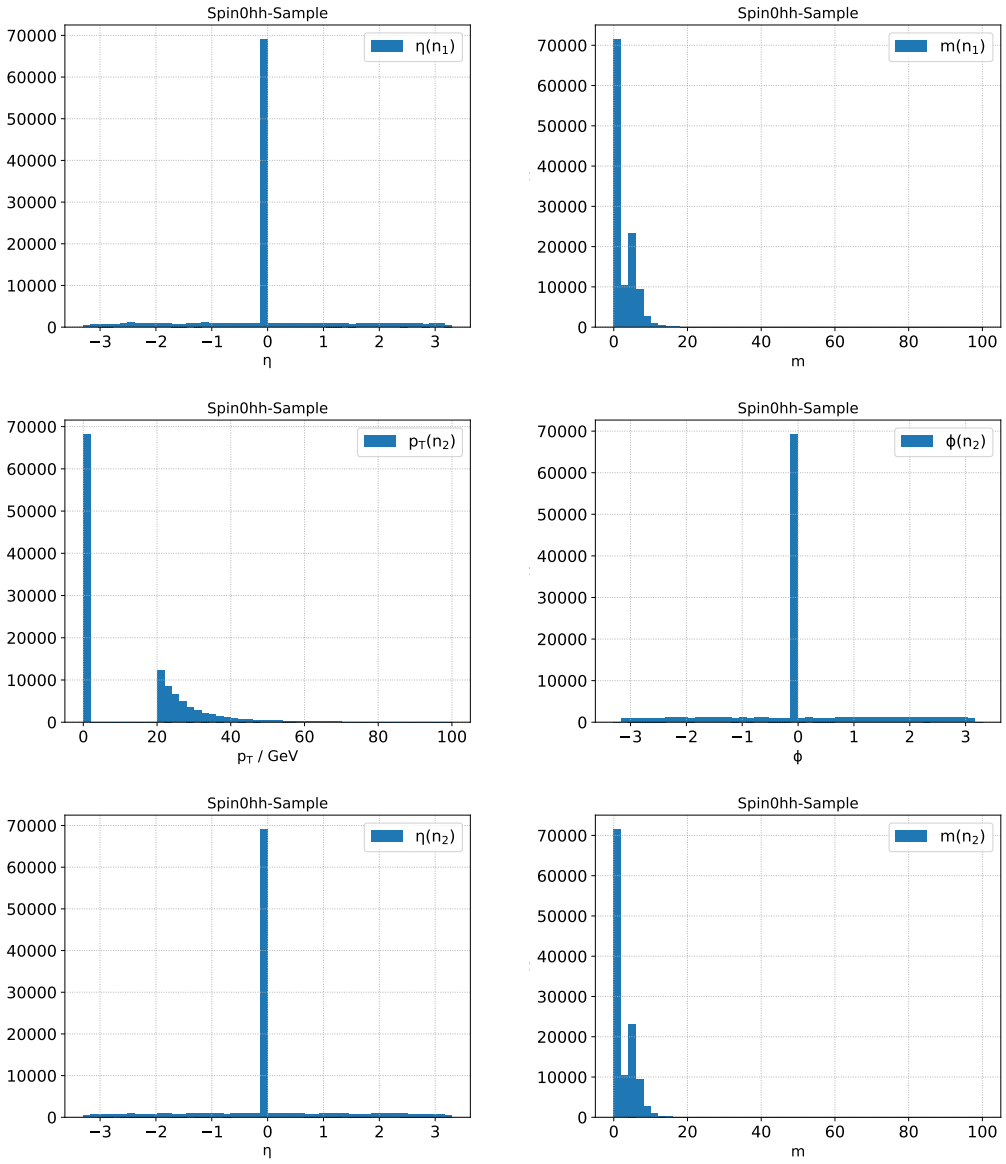


Figure A.26: Distributions of the variables of the spin0hh sample used in the Jets set (2/2).

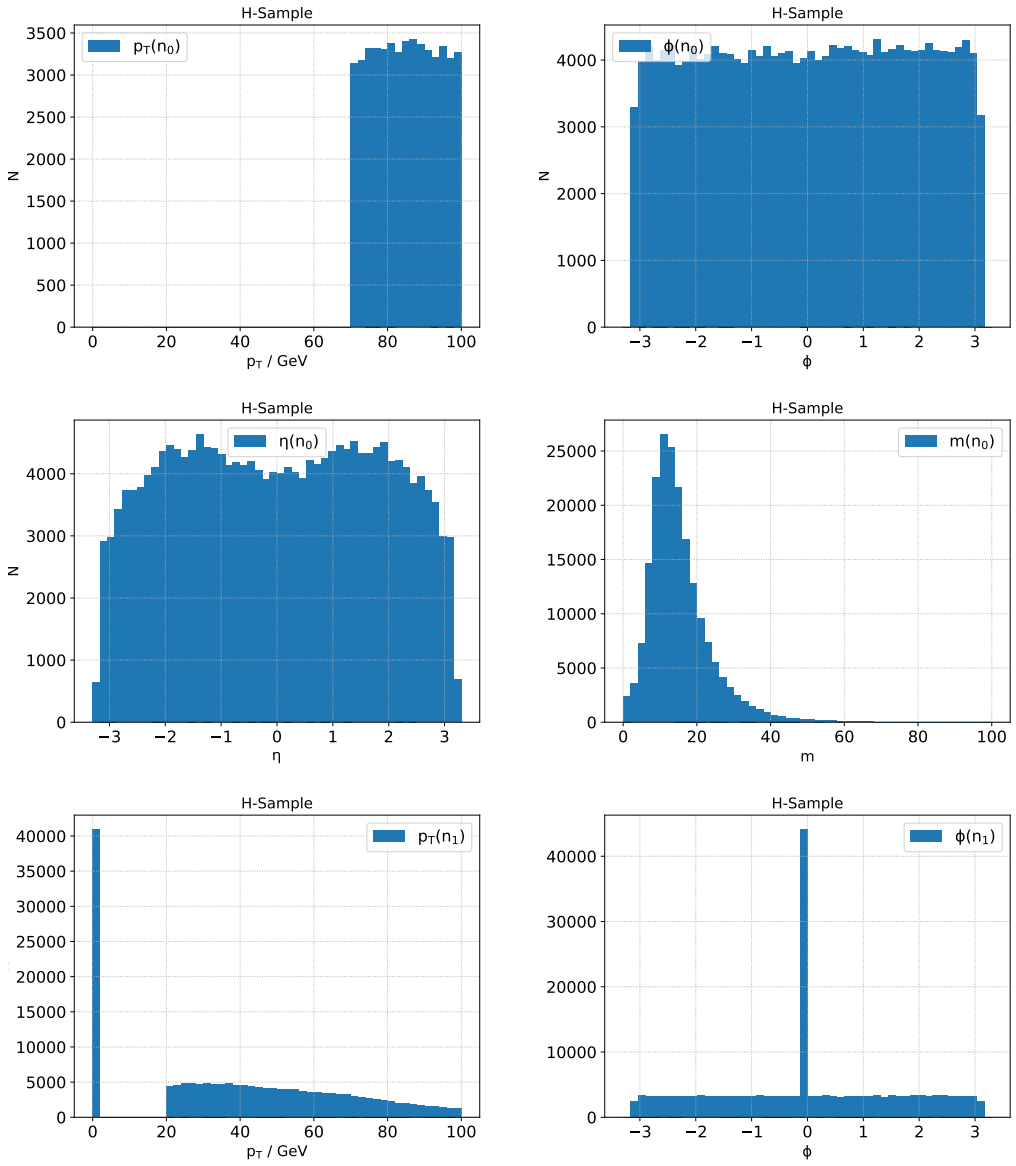


Figure A.27: Distributions of the variables of the H-sample used in the Jets set (1/2).

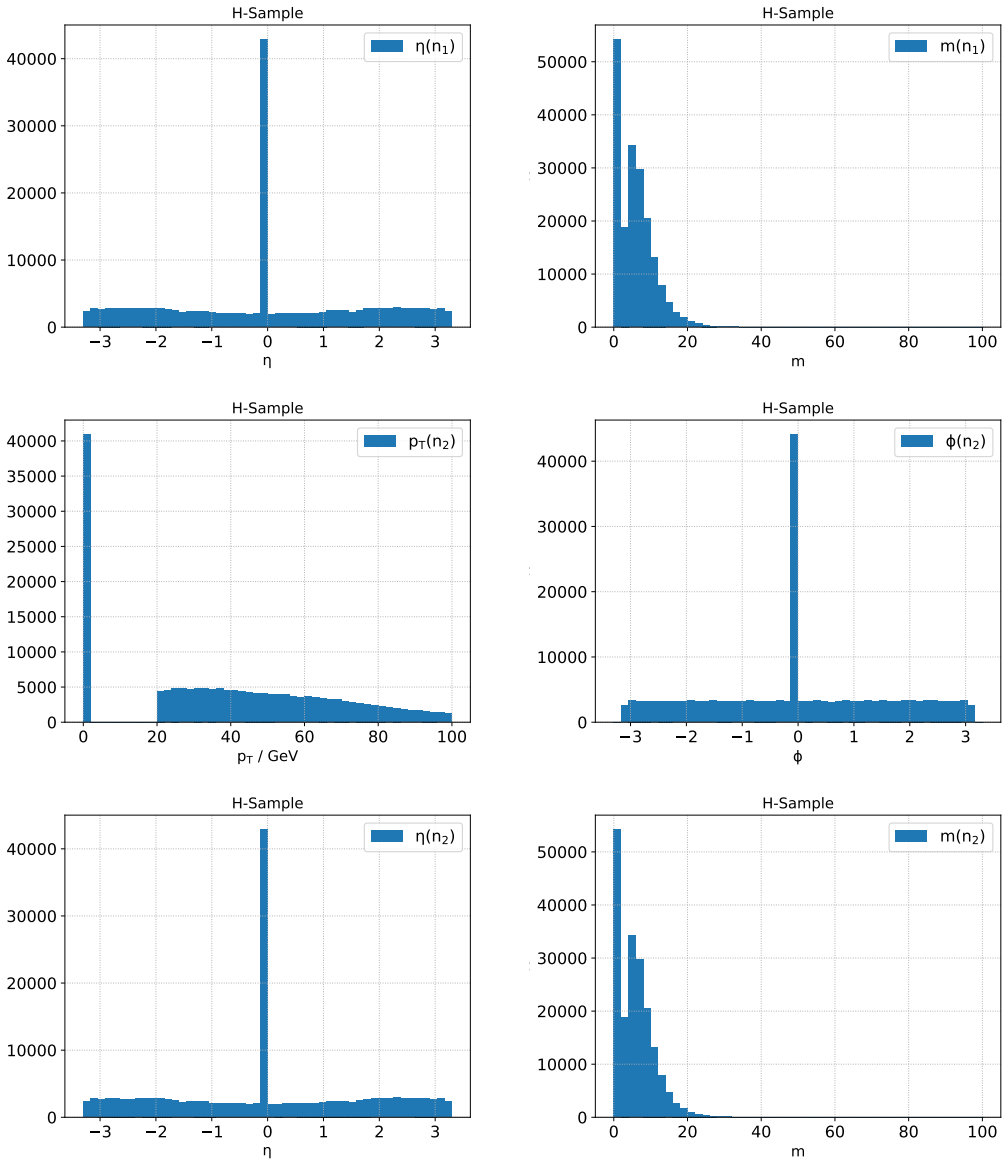


Figure A.28: Distributions of the variables of the H-sample used in the Jets set (2/2).

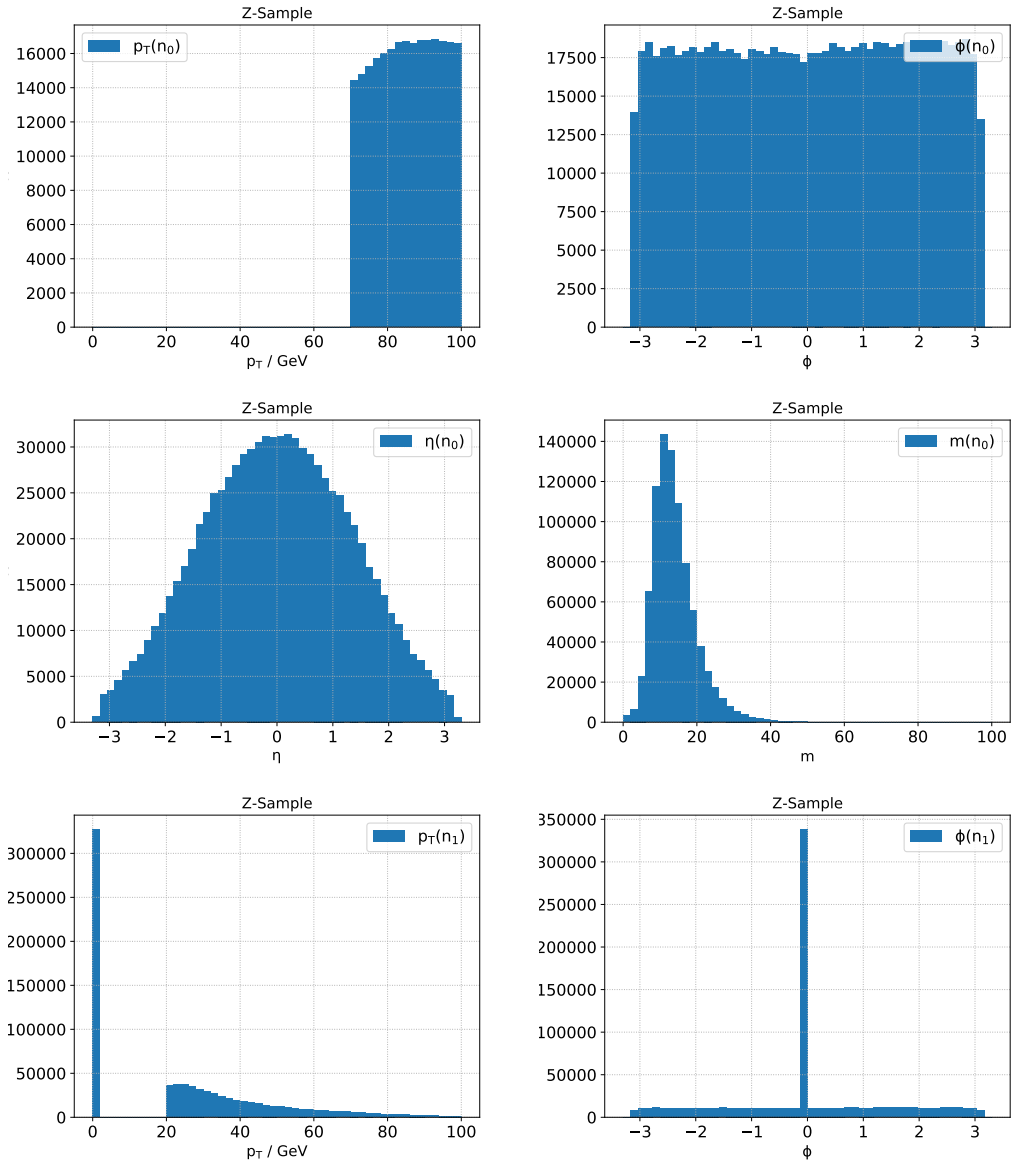


Figure A.29: Distributions of the variables of the Z-sample used in the Jets set (1/2).

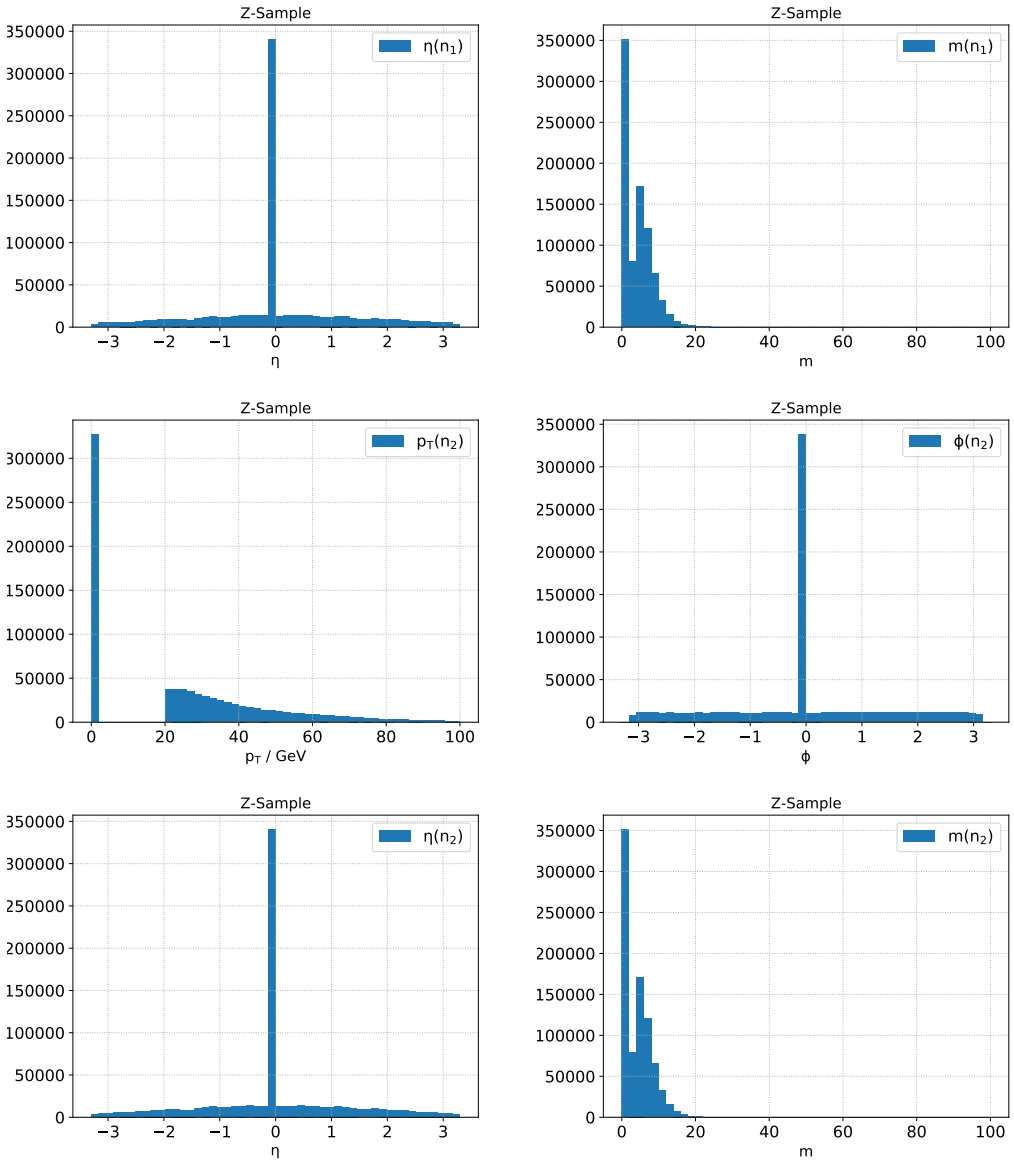


Figure A.30: Distributions of the variables of the Z-sample used in the Jets set (2/2).

List of Figures

List of Tables