Universidade de São Paulo
Instituto de Física

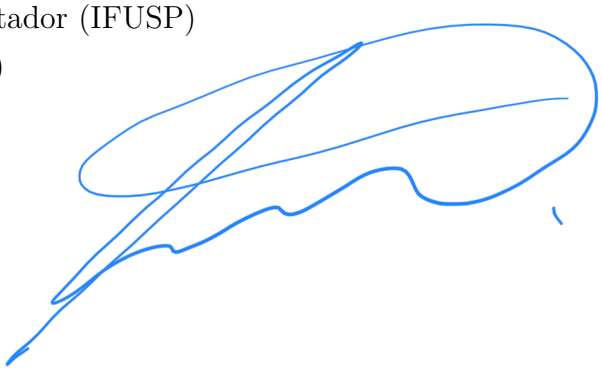
Estudo da reconstrução da massa de bósons Z no processo $Z \to \tau\tau$ em colisões próton-próton a $\sqrt{s} = 13$ TeV no experimento ATLAS

Caio Cesar Daumann

Orientador: Prof. Dr. Marco Aurélio Lisboa Leite

Dissertação de mestrado apresentada ao Instituto de Física da Universidade de São Paulo, como requisito parcial para a obtenção do título de Mestre(a) em Ciências.

Banca Examinadora:
Prof(a). Dr(a). Marco Aurélio Lisboa Leite - Orientador (IFUSP)
Prof(a). Dr(a). André Asevedo Nepomuceno (UFF)
Prof(a). Dr(a). Antonio Vilela Pereira (UERJ)



São Paulo
2023

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Daumann, Caio Cesar

Estudo da reconstrução de massa de bósons Z no processo $Z \rightarrow \tau\tau$ em colisões próton-próton a $\sqrt{s}$ = 13 TeV no experimento ATLAS/ Study of the Z boson mass reconstruction in the $Z \rightarrow \tau\tau$ process in proton-proton collisions at $\sqrt{s}$ = 13 TeV in the ATLAS experiment. São Paulo, 2023.

Dissertação (Mestrado) - Universidade de São Paulo. Instituto de Física. Depto. de Física Nuclear.

Orientador: Prof. Dr. Marco Aurélio Lisboa Leite

Área de Concentração: Física de Altas Energias.

Unitermos: 1. Física de partículas; 2. LHC; 3. ATLAS; 4. Bóson ; 5. Quarks.

USP/IF/SBI-024/2023

University of São Paulo
Physics Institute

Study of the Z boson mass reconstruction in the $Z \to \tau\tau$ process in proton-proton collisions at $\sqrt{s} = 13$ TeV in the ATLAS experiment

Caio Cesar Daumann

Supervisor: Prof. Dr. Marco Aurélio Lisboa Leite

Dissertation submitted to the Physics Institute of the University of São Paulo in partial fulfillment of the requirements for the degree of Master of Science.

Examining Committee:
Prof(a). Dr(a). Marco Aurélio Lisboa Leite - Orientador (IFUSP)
Prof(a). Dr(a). André Asevedo Nepomuceno (UFF)
Prof(a). Dr(a). Antonio Vilela Pereira (UERJ)

São Paulo
2023

Acknowledgments

First, I would like to thank my father, mother and my brothers for supporting me and helping me with everything I needed during my journey until this moment. I would not be here without your help and motivation. Thank you all for loving me.

I would also like to express my gratitude to my supervisor, Marco, for providing me with the opportunity to work in his group under his supervision. Our meetings and conversations were fundamental to the development of this work, as well as my personal growth as a physicist. The knowledge I acquired during this masters will be always useful to me.

This work would not have been possible without the ATLAS di-τ analysis group, whose meetings were very fruitful and helped me improve my work. I would like to give special thanks to Chris Pollard for his patience, insightful conversations and for all of the knowledge transfer.

Furthermore, I express my deep appreciation to my old and dear friends in my hometown Joinville. My sincere thanks go out to Ruan, Fernanda, JP, Wander, Vini, Jéssica, Krankel, Rebeca, Tatiana, Gustavo, Lais, Bia, Evelin, Bruna, Luana, Daniel, Júlio, Hannah, Vitor, Patrick, Alisson, and Bruno. I would also like to extend my gratitude to my friends Gabriel Amancio and Grime for sharing the new experience in São Paulo and supporting each other in this new world. I am thankful for the partnership of each and every one of you, as you are friends that I cherish and intend to keep for life. I consider myself extremely fortunate to have encountered all of you, as you have been supportive during times when I needed it most. Thank you for all the chat and good moments, moments which would become a book if I wrote them here, you have all contributed to making my life better, and for that, I am truly grateful.

I would like to express my gratitude to the friends I made in São Paulo, and in particular, I would like to extend special thanks to my friends from 213: Thais, Mapa, and Lucas. These exceptional individuals are true friends who one does not encounter every day. It was a pleasure to work alongside you and to have your company everyday.

I thank Thais Abelha, who was the most important person in my life over the past year. Our walks home, almost daily calls, and weekends together helped me clear my head, stay motivated, and become a happier person. Her support and kindness enabled me to complete this work. Thank you for your companionship and love

I Thank Mapa, for all our conversations about books, careers, soccer, life, and all the Palmeiras vs. Flamengo soccer games we watched. It was always a pleasure to see my Flamengo beat your Palmeiras.

Also, I extend my thanks to the members of the HEPIC group for the enjoyable happy

hours we spent together. It was a pleasure to have informal conversations over drinks with each one of you.

I would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the scholarship and financial support which made it possible for me to fully dedicate myself to my master's degree.

Abstract

The Standard Model (SM) has demonstrated remarkable success in describing processes observed at the TeV scale at the Large Hadron Collider (LHC), including the experimental confirmation of a particle consistent with the SM Higgs boson by the ATLAS and CMS experiments. However, recent experimental evidence indicating the breaking of lepton universality in B meson decays requires further investigation using both low and high p_T processes. Furthermore, despite a decade of data-taking at the LHC, measurements of Drell-Yan $\tau\tau$ processes are still lacking, which are important to complement existing analyses in the ee and $\mu\mu$ final states. Finally, the fact that physics processes involving third-generation leptons may exhibit a higher sensitivity to physics beyond the Standard Model underscores the need for prioritizing this analysis at the LHC.

From an experimental point of view, several challenges are present in analyses with di-τ final states. The short lifetime of τ and its decay products involving several neutrinos make kinematics reconstruction highly challenging, particularly in the case of Z boson invariant mass reconstruction in the $Z \to \tau\tau$ process. Currently, the ATLAS experiment is undertaking an effort to measure, at the particle level, the $m_{\tau\tau}$ differential cross-section distributions for the fully-hadronic and semi-leptonic decay modes of the tau pairs. These measurements may contribute to testing Leptoquark models.

This study presents the results of mass estimation for τ pairs, wherein several mass estimation methods were evaluated, and their performance was assessed in both the Z resonance region and the high mass region of the spectrum. Additionally, machine learning techniques were developed as an alternative to more traditional mass estimation methods. This study examines the performance of various methods in reconstructing the signal, background, and data. The performance of each method will be presented and discussed in this work.

Keywords: 1. Particle physics, 2. LHC, 3. ATLAS, 4. Boson, 5. Quarks.

Resumo

O Modelo Padrão (MP) tem demonstrado um sucesso notável na descrição de processos observados na escala de TeV no Large Hadron Collider (LHC), incluindo a confirmação experimental de uma partícula consistente com o bóson de Higgs do MP pelos experimentos ATLAS e CMS. No entanto, evidências experimentais recentes indicam uma possível violação do princípio da universalidade leptônica envolvendo decaimentos de mésons B. Esses indícios requerem investigações adicionais usando processos de baixo e alto p_T. Além disso, apesar de uma década de coleta de dados no LHC, existem poucas medidas de processos Drell-Yan com estados finais contendo pares de taus, embora estas medidas sejam nescessárias para completar as análises existentes nos estados finais ee e $\mu\mu$. Por fim, o fato de que processos físicos envolvendo léptons de terceira geração podem apresentar uma sensibilidade maior à física além do Modelo Padrão enfatiza a importância de priorizar essa análise no LHC.

Do ponto de vista experimental, entretanto, várias dificuldades estão presentes nas análises envolvendo estados finais que contêm τ's. O curto tempo de vida do τ e seus decaimentos envolvendo vários neutrinos tornam a reconstrução cinemática altamente desafiadora, especialmente no caso da reconstrução da massa invariante do bóson Z no processo $Z \to \tau\tau$. Atualmente, o experimento ATLAS está desenvolvendo um programa de medidas das distribuições diferenciais de seção de choque de produção $m_{\tau\tau}$ para os modos de decaimento totalmente hadrônicos e semi-leptônicos dos pares de τ. Essas medidas podem contribuir para testar e restringir modelos de Leptoquarks.

Este estudo apresenta os resultados da investigação de diferentes técnicas de estimativa de massa para processos envolvendo pares de τ's . Vários métodos foram avaliados em relação ao seu desempenho tanto na região de ressonancia do bóson Z quanto na região de alta massa do espectro. Além disso, técnicas de aprendizado de máquina foram empregadas para estudar alternativas para a estimativa de massa. O estudo examina o desempenho de vários métodos na reconstrução do sinal, do fundo e dos dados. O desempenho de cada método é apresentado e discutido neste trabalho.

Palavras-chave: 1. Física de partículas, 2. LHC, 3. ATLAS, 4. Bóson, 5. Quarks.

LIST OF TABLES

LIST OF FIGURES

CONTENTS

CHAPTER 1

INTRODUCTION

The Standard Model of particles (SM) has proven to be a remarkably successful theory that describes the fundamental particles and its interactions. In 2012, the experimental observation by the ATLAS and CMS experiments of a particle compatible with the SM Higgs Boson [1, 2] was another triumph of the SM. Figure 1.1 illustrates all of the fundamental particles and force carriers predicted by the SM. The Standard Model of particles are divided into two categories: fermions, which include quarks and leptons, the particles that make up matter, and bosons, which are responsible for mediating interactions between particles [3]. These bosons include the photon, which mediates the electromagnetic force, the gluon, which mediate the strong force, and the W and Z bosons, which mediate the weak force [4]. The exchange of a Z boson transfers momentum, spin, and energy between particles, while leaving their electric charge unaffected. Therefore, this interaction is referred to as a neutral current. The Z boson, as an electrically neutral particle, does not mediate charged currents [5]. In contrast, the exchange of a W boson can transfer both momentum, spin, and energy, as well as electrical charge. Consequently, this interaction is referred to as a charged current interaction. The W boson exchange is responsible for the weak nuclear interaction that governs beta decay, a process where a neutron can decay into a proton, electron, and anti-neutrino by emitting a W boson. It is important to note that the Standard Model does not include gravity as one of the fundamental forces it describes.

The quarks and leptons are separated into three generations or families. The first generation is composed of the up and down quarks, the electron and the electron neutrino. The second generation is composed of the charm and strange quarks, the muon and the muon neutrino. And the third generation is composed of the top and bottom quarks, the tau lepton and the tau neutrino. It is worth noting that each generation of quarks and

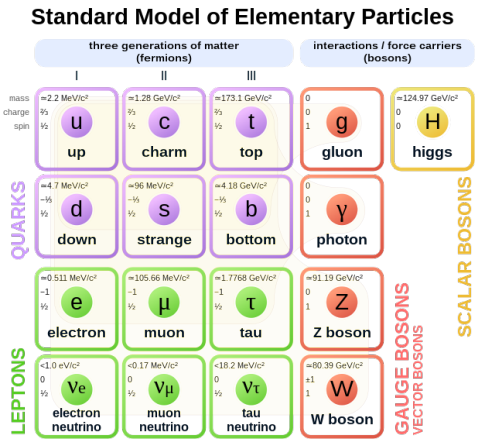


Figure 1.1: The SM elementary particles, and force carriers. [6]

charged leptons becomes increasingly massive as we move to the next one. For example, the tau (τ) lepton is the most massive[1] lepton among the three generations [7], and it is the only one capable of decaying hadronically. The top quark, on the other hand, is the heaviest[2] quark as well as the heaviest elementary particle [8].

While the Standard Model (SM) has been highly successful in describing a wide range of experimental data and making accurate predictions [9, 8, 1], there are both experimental and theoretical results that suggest that it may not be the final and most fundamental theory of nature. For example, the SM does not have a candidate for dark matter [10], it does not explain the matter-antimatter asymmetry [11], and the neutrino masses predicted by the SM are zero, which is not consistent with experimental observations of neutrino oscillations [12]. In light of the Standard Model's inability to address and accurately predict certain phenomena, there are enough reasons to believe that the SM is not a complete theory. As a result, it is plausible that an alternative theory may serve to complete the SM. However, currently, no concrete indication of such a theory has been identified, and therefore,we are exploring various avenues in search of possible leads.

To this end, searches for new physics can be conducted in two distinct ways.

Direct Searches

Direct searches for new particles are commonly carried out in high-energy colliders, where the decay products of a hypothetical new particle, usually a resonance, are explored in the phase space. To illustrate, consider the Higgs boson search performed by the ATLAS experiment. In this study, data distributions were compared with the detector level SM prediction, both with and without the Higgs boson. Advanced statistical methods

[1] 1.777 GeV
[2] 172.9 GeV

were employed to compare the distributions, and the probability that the data could be explained only by the SM without the Higgs boson was found to be extremely low (one chance in 3.5 million) [1], the null hypothesis stating that the SM does not include the Higgs boson was rejected based on an analysis of the final state channels of Higgs decay. This analysis involved a direct examination of the aforementioned channels [1].

Indirect searches

Indirect searches in particle physics are a crucial complementary approach to direct searches, in which we examine the effects of hypothetical new particles or phenomena on other particles. Unlike direct searches, indirect searches do not involve directly observing the BSM particle itself, but rather searching for its indirect effects. For instance, suppose that a hypothetical particle would alter the decay rates of electrons and muons in a certain reaction. In that case, we could measure the decay rates and search for any deviations from the predicted values of the SM. This would provide evidence for the existence of the hypothetical particle, even if it cannot be directly detected.

To date, direct searches have not yielded any conclusive evidence of new particles beyond the SM, and neither have indirect searches revealed any significant deviations from the SM predictions. However, indirect searches have shown indications of results that are not entirely consistent with the SM predictions. For instance, experiments such as B-factories, which are designed to produce and detect a large number of B mesons, have measured possible anomalies in the $b \to c\tau\nu_\tau$ decay process, compared to the expected values of the SM. The BaBar[3] and Belle[4] experiments have measured deviations in the $R_{D^{(*)}}$ and R_D quantities, defined as below:

$$R_{D^{(*)}} = \frac{\beta(\bar{B} \to D^{(*)}\tau\nu_\tau)}{\beta(\bar{B} \to D^{(*)}\ell\nu_\tau} \tag{1.1}$$

and:

$$R_D = \frac{\beta(\bar{B} \to D\tau\nu_\tau)}{\beta(\bar{B} \to D\ell\nu_\tau)} \tag{1.2}$$

The $R_{D^{(*)}}$ and R_D quantities measure the ratio of branching fractions for the $\bar{B} \to D\tau\nu_\tau$ decay compared to the decay into light leptons, $\bar{B} \to D\ell\nu_\ell$. These branching ratios are interesting ways to test the lepton universality principle, which states that all leptons, regardless of their specific type, have the same strength of interaction with

[3]BaBar is located at the SLAC National Accelerator Laboratory, it ceased data collection in 2008 due to budget cuts.

[4]Belle ran until 2010, when it stopped data collection for future upgrades.

other gauge bosons, except for their interactions with the Higgs field, which give rise to differences in lepton masses. Any differences in lepton decays and kinematics can therefore be attributed to the differing masses and interactions with the Higgs field. The BaBar and Belle experiments have measured the $R_{D^{(*)}}$ and R_D ratios, and a global average of these measurements has been performed [13]. The results of these measurements deviate from the Standard Model prediction by 3.3 standard deviations, indicating some tension in the SM [13].

The observed larger production of τ's compared to muons suggests a violation of the lepton flavor universality principle. However, it should be noted that this result is still far from being a definitive discovery and further study of this topic is necessary. In addition to the BaBar and Belle experiments, the LHCb experiment is also currently performing this measurement[14], and the Belle II[5] experiment plans to also perform such measurements.

Other results from the LHCb experiment, based on run 1 data, have highlighted a tension with the lepton universality principle. Specifically, the experiment measured the ratio between the branching ratio of B meson decays into muons and electrons, indicating a higher production of electrons compared to the expected result from the SM [15]. However, more recent LHCb results, utilizing both run 1 and run 2 data, measured the same ratios and demonstrated compatibility with the SM's predicted outcomes [16]. These more recent results surpass the previous LHCb measurements and, therefore, suggest only discrepancies in the tau relative to the other light leptons, rather than between the light leptons themselves.

The deviations from the expected values of the Standard Model (SM) are currently referred to as $c\tau\nu_\tau$ anomalies [17]. Currently, there exist three categories of Beyond the Standard Model (BSM) models that could potentially explain a deviation from the SM [17]. These BSM models include a new W' boson, a charged Higgs, and a Leptoquark (LQ). The diagrams of such process are shown in Figure 1.2.

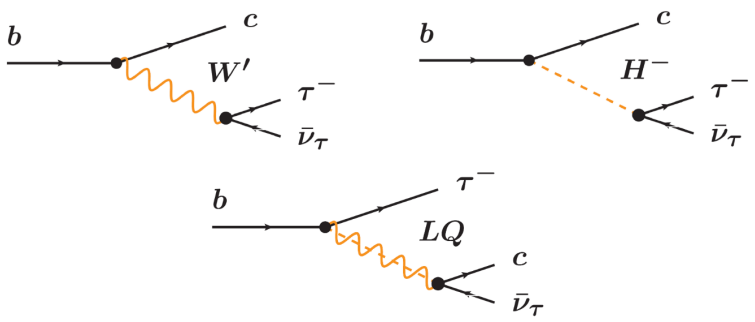


Figure 1.2: BSM models that could explain the $b\tau\nu$ decays, W' (top left), charged-Higgs (top right) and LQ (bottom) . [17]

[5]The upgrade of the Belle experiment, which started data taking at 2018.

Another approach to investigate the $c\tau\nu_\tau$ anomalies is through direct searches of the hypothetical particles that can potentially explain them. Such searches can provide more insight into the problem by setting upper limits on their production cross sections which in turn constrains the kinematical phase space allowed by the current data. Since direct searches require high-energy collisions to probe higher masses, the ATLAS experiment can play a crucial role as a high transverse momentum probe in addressing this issue.

One possible explanation for the observed deviations is the existence of hypothetical bosonic particles called Leptoquarks (LQs), where LQs with a stronger coupling to the third generation of fermions may provide an explanation for the observed anomalies [18]. As a result, processes involving taus or top quarks may be more sensitive to the presence of LQs [19], and hence are interesting to explore experimentally.

Searches for Leptoquarks have been conducted by the CMS and ATLAS experiments [18, 20]. The searches included resonant production of leptoquark pairs and single leptoquarks (processes (a) and (b) in Figure 1.3). Both experiments excluded the existence of leptoquarks with masses below 1 TeV. However, for masses above 1 TeV, the pair production of leptoquarks is suppressed, and the dominant contribution would be non-resonant production (process (c) in Figure 1.3) [18]. Resonant production occurs when a resonance is created in the phase space of a collision. If leptoquarks exist, their decay into a tau and a quark would result in a resonance appearing on the lepton-quark invariant mass distribution. The production rate of a leptoquark resonance is highest at an energy corresponding to the mass of the leptoquark. In contrast, non-resonant decays do not involve the creation of an intermediate resonance and do not result in narrow peak-like structures in the phase space of the decay.

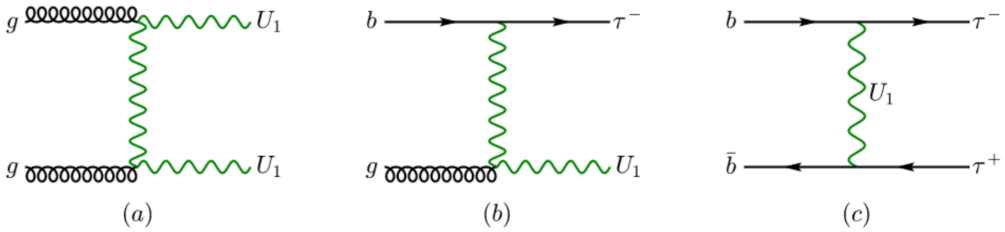


Figure 1.3: Feynman diagrams for vector leptoquark pair production (a) single leptoquark production (b), and t-channel Drell-Yan production [19].

Recent results from the CMS experiment have reported an excess of events with a significance of 3.4 standard deviations from the Standard Model expectation in the $\tau\tau$ + jet channel in a search for leptoquarks. This excess has been observed for a leptoquark with a mass of 2 TeV and a coupling strength of 2.5, indicating the possible existence of leptoquarks [21]. Further investigation and confirmation of these observations are necessary, emphasizing the importance of conducting similar searches by the ATLAS

experiment. This makes searches and measurements in the $\tau\tau$ final state of summary importance, as these leptoquarks would contribute to the SM neutral current Drell-Yan ($Z \to \tau\tau$) production in the high mass region. The Drell-Yan channel is important as it is expected a sizeable effect of interference between the LQ and ($Z \to \tau\tau$) production [18].

Drell-Yan neutral current production is a fundamental process in particle physics that involves the annihilation of a quark-antiquark pair in a hadron collider, followed by the creation of a virtual photon or Z boson, which then decays into a lepton-antilepton pair. The Feynman diagram for the SM neutral current Drell-Yan decay is shown in Figure 1.4.

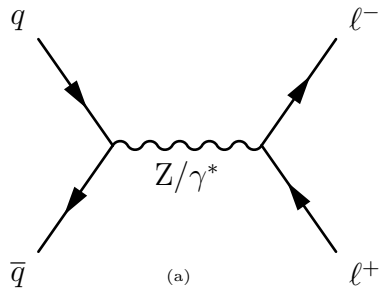


Figure 1.4: Feynman diagrams for the Drell-Yan process. [22].

The study of the neutral current Drell-Yan with taus in the final state is considered a promising avenue for searching for evidence of BSM physics [23]. However, the measurement of tau leptons is faced with several challenges. One of the main challenges is the presence of neutrinos in the tau lepton final state, which results in missing information and the inability to measure the mass of the tau lepton directly. The presence of neutrinos makes it impossible to fully reconstruct the pre-decay τ quantities, and the measured quantities yield underestimated results, as the neutrinos carries away part of the τ momentum. This missing information cannot be measured. Therefore, estimating the mass using visible decay momentum alone will lead to underestimation of the true mass of the event. Mass estimation methods, on the other hand, utilize theoretical knowledge of the tau lepton decay probabilities and constraints to yield a more accurate estimation of the mass.

In addition, the tau lepton is difficult to experimentally identify and distinguish from the backgrounds due to the significant amount of hadronic activity in hadron colliders, which results in large QCD backgrounds. Furthermore, the tau leptons have a short lifetime[6] and decay before reaching any detector apparatus, and hence we have to look for it decay products and reconstruct the tau from them. These challenges make the measurement of the tau leptons a complex task, that requires good identification, mass estimation and background rejection techniques.

[6] $330 \times 10^{-15} s$ [23]

This work purpose is to evaluate several mass estimation methods employed in the resonance region and study their performance in the high mass region. Mass estimation methods are widely used in the resonance region, were searches and measurements like the $H \to \tau\tau$ coupling measurement [24], and inclusive cross section measurment [25] were performed. However, their application and performance in the high mass region have not been thoroughly evaluated, as there is not yet measurements of the $\tau\tau$ cross section in the high mass region above the resonance of the Z boson.

The di-τ cross-section and the leptoquark searches in the high-mass regions are an official ATLAS analysis. The complete team, in which I participate and additional analysis details are available on the ATLAS Analysis Glance page, under the analysis code (ANA-EXOT-2022-42).

1.1 Past results in the field

In the past, several measurements and searches related to the production and decay of the Z boson have been performed by the ATLAS and CMS collaborations. The ATLAS collaboration has measured the inclusive cross-section of the Z boson decaying into τ pairs in the resonance region (66-166 GeV region) using 36 fb^{-1} of $\sqrt{s} = 7$ TeV data. The results were in agreement with the Standard Model (SM) expectations [26, 27]. The CMS collaboration also performed an inclusive cross-section measurement of the $Z \to \tau\tau$ process using $2.3fb^{-1}$ of $\sqrt{s} = 13$ TeV data, which again was in agreement with the SM expectations (computed at next-to-next-to-leading order).

While there have been inclusive cross-section measurements, no differential cross-section measurements of the $Z \to \tau\tau$ process in the resonance have been performed by either experiment. The ATLAS collaboration has performed searches and production cross-section measurements of the $H \to \tau\tau$ process in the resonance [24]. The ATLAS and CMS collaborations has also performed inclusive cross-section measurements of the $H \to \tau\tau$ process [25, 28].

The ATLAS collaboration has previously measured the high-mass production cross-sections of the $Z \to ee$ and $Z \to \mu\mu$ processes during Run 1 [29]. Currently, a measurement with the second run of data collection (Run 2) is being finalized. The measurement of high-mass $\tau\tau$ production would complement those measurements.

Regarding leptoquark searches, as mentioned earlier, the CMS experiment has reported a 3.4σ excess from the SM expectation [21]. In addition, CMS has conducted model-independent searches for heavy neutral Higgs bosons in the di-tau channel [30], as well as searches for minimal supersymmetric particles in the di-tau channel [30]. Similarly, the ATLAS collaboration has performed searches for minimal supersymmetric particles

in the di-tau channel [31]. However, none of these searches have observed any significant deviation above the 3σ level for neutral Higgs or minimal supersymmetric particles.

The current absence of a differential cross section measurement for the $Z \to \tau\tau$ process, both in the resonance and high-mass tails, highlights the need for further study in this area. To address this gap in knowledge, the aforementioned analysis aims to measure differential cross sections for this process, as well as conduct searches for non-resonant leptoquark production in the high-mass region . The work presented in this document supports this analysis.

1.2 Main results of this work

In this work, the performance of several mass estimation methods was studied and evaluated. The studied methods are:

- The visible mass [32]
- Collinear method [33]
- Missing mass calculator [32]
- The $\hat{s}$ minimum method [34]
- Total transverse mass [18]

The Missing Mass Calculator (MMC) method, which is known to be the best-performing method in the Z resonance region [24], was found to be unsuitable for use in the high-mass region. This is because the method relies on derived sets of probability density functions (PDFs) that have not yet been derived for the high-mass region, leading to its failure for the majority of events (see chapter 5). The other available mass estimation methods in the literature were found to produce non-diagonal response matrices due to the methods tendency to estimate lower values of mass in the high mass region. This is because the visible decays of the τ carry a smaller portion of the pre-decay tau p_T for high p_T tau decays, as explained in chapter 5.

To address this limitation, machine learning techniques were investigated to produce more diagonal migration matrices and make these methods suitable for use in the analysis. However, these techniques resulted in lower signal-to-background ratios, as the high-mass region was populated with background events. Also, a bias correction was performed in the mass estimation methods in order to correct for the methods estimation bias and yielding more diagonal response matrices.

The consistency of the methods was demonstrated through their application to different Monte Carlo generators and data, which is especially important for machine learning methods to avoid over-training. Both the neural network and other available methods showed good consistency, with similar performances when applied to data and alternative generators.

This work is structured as follows: first, the ATLAS experiment and its detectors are introduced in Chapter 2. Chapter 3 present the data and Monte Carlo simulations that will be used in the mass estimation studies. Chapter 4 provides an overview of the event selection criteria currently in use. In Chapter 5, existing mass estimation methods in the literature will be introduced. Chapter 6 will evaluate their performance. Chapter 7 will provide an overview of neural networks (NNs) and the NN models used for mass estimation. In Chapter 8, these NN models will be compared with existing mass estimation methods. Finally, Chapter 9 will provide a summary and the concluding remarks.

CHAPTER 2

EXPERIMENTAL SETUP

The present study uses data and simulations from the ATLAS experiment at the Large Hadron Collider (LHC) to conduct the research. The purpose of this Chapter is to provide a brief explanation of the functioning of both the LHC and the ATLAS experiment.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator situated on the Swiss-French border. As the most powerful accelerator built by human beings to date, it is capable of generating proton-proton collisions with energies up to 13.6 TeV. The LHC is an underground accelerator, located approximately 170 meters beneath the surface, and has a circumference of 27 kilometers. Its large circumference enables it to accelerate particles to extremely high energies.

The Large Hadron Collider (LHC) is operated by the European Organization for Nuclear Research (CERN), also known as the Conseil Européen pour la Recherche Nucléaire. In addition to the LHC, CERN also operates a number of other, less powerful particle accelerators that are used to accelerate particles before they are injected into the LHC. These include the Linear Accelerator 2 (LINAC 2), which accelerates particles up to 50MeV, as well as the Proton Synchrotron Booster (Booster), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS), which accelerate particles up to 450 GeV in energy. Once accelerated, the particles are then injected into the LHC for further acceleration.

The LHC is capable of accelerating particles to extremely high energies due to powerful electric and magnetic fields. The electric fields provide the necessary acceleration for the particles, while the superconducting dipole and quadrupole magnets capable of generating magnetic fields up to 8.33 T [35] are responsible for steering and focusing the beams

ensuring that they collide precisely at the desired location and are kept in a defined trajectory along the accelerator.

After the particle beams have attained the desired energy, they are then collided at four strategic points where the LHC experiments are located. The collisions occur at the ALICE[1], the LHCb[2], the CMS[3], and the ATLAS experiment. Additionally, there are two detectors measuring events close to the beam line, namely the LHCf[4] and the TOTEM[5] experiments. A diagram illustrating the LHC complex is available in Figure 2.1.

Each LHC experiment has a different purpose. The LHCb experiment focuses on B meson physics, while the ALICE experiment specializes in analyzing heavy ion collisions to study the properties of the quark-gluon plasma. The ATLAS and CMS experiments are both multi-purpose and aim to study a broad variety of processes in the EWK (Electroweak) sector. With the advantage of being two similar yet independent experiments, one experiment can validate the other's results.

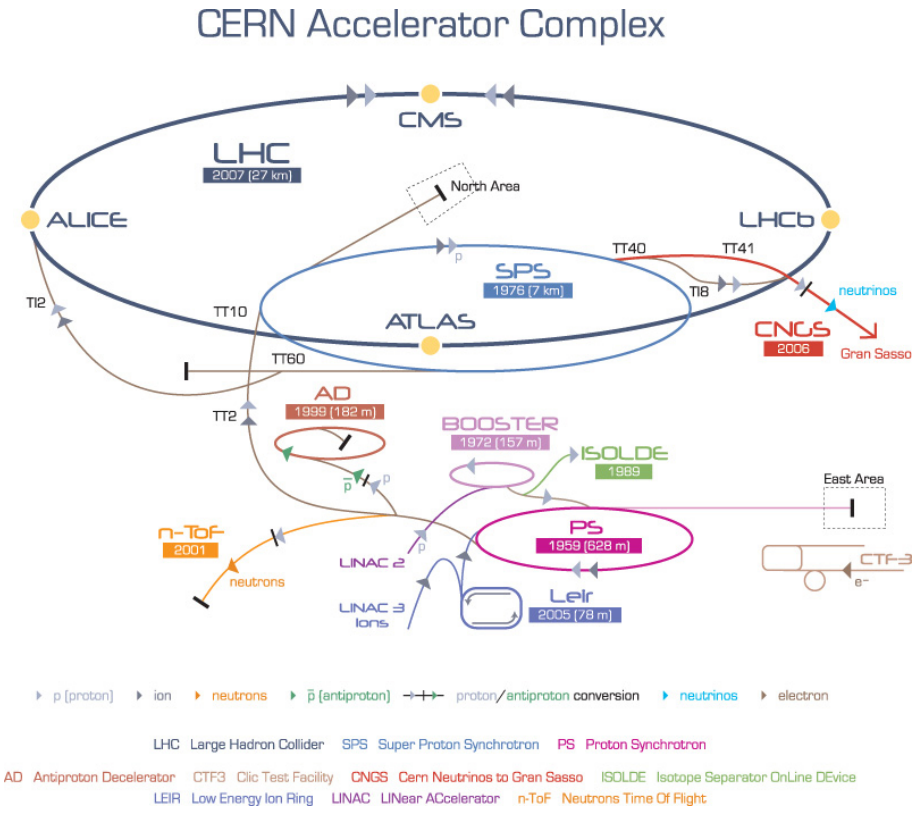


Figure 2.1: Overview of the LHC and its auxiliary accelerators. Taken from: link

The LHC operates in runs. After Run 1, which concluded in 2013, the LHC underwent a two-year period of planned maintenance, repairs, and upgrades to the experiments. In

[1] A Large Ion Collider Experiment
[2] LHC beauty
[3] Compact Muon Solenoid
[4] LHC forward
[5] TOTal Elastic and diffractive cross section Measurement

2015, the LHC resumed operations, accelerating and colliding protons once again, with a center of mass energy of the beam reaching 13 TeV. This marked the beginning of Run 2 of the LHC. This work uses the Run 2 data and simulated MC events. The third run started in July 2022 and is collecting data at $\sqrt{s} = 13.6$ TeV.

During Run 2, the instantaneous luminosity ($\mathcal{L}$) of the LHC, which is proportional to the number of collisions per unit time was increased. The greater the value of $\mathcal{L}$, the higher the number of collisions and, therefore, the number of events. As the LHC accelerates bunches of protons and not single particles to increase the number of collisions and luminosity, the mean number of interactions per bunch crossing (μ) must be increased. Figure 2.2 illustrates the delivered luminosity by the LHC, the recorded luminosity by the ATLAS experiment, and the amount of data that is suitable for use in physics analyses.

In Run 2, the LHC produced an unprecedented amount of data. During this period, the LHC delivered 156 fb^{-1} of data, more than five times the amount delivered in Run 1 (which was 28.26 fb^{-1}). This increase in data volume enables the study of even rarer processes with extremely low cross sections. This was made possible by the increase in μ as the years with the highest delivered luminosities correspond to those with the greatest number of collisions per bunch crossing. The Run 2 dataset corresponds to approximately 4% of the total expected delivered luminosity of the LHC during its lifetime, where the majority of the data is anticipated to be collected during the high luminosity upgrade of the experiment.

During data taking, errors may occur in the experiment apparatus, such as dead pixels or voltage peaks, and data collected concomitant to such errors are not suitable for use [35]. The ATLAS experiment uses a system of good run lists, where the good luminosity blocks (spanning 1-2 minutes of data-taking) are identified for use in physics analysis. This helps to ensure that the data used in the analysis are of high quality and free from any instrumental errors.

2.2 The ATLAS detector

The ATLAS experiment (***A*** ***T***oroidal ***L***HC ***A***pparatu***S***) [35] is the largest multi purpose experiment in the LHC. Its cylindrical design allows for a large coverage around the collision point located at the center of the detector, This broad coverage reflects the wide range of physics processes that the ATLAS experiment is designed to study. The cylindrical design of the detector enables precise measurements of jets, leptons, and photons, with a wide acceptance in a large region of pseudorapidity, and coverage of the full azimuthal angle.

A detailed description of the ATLAS experiment is provided in [35], upon which this

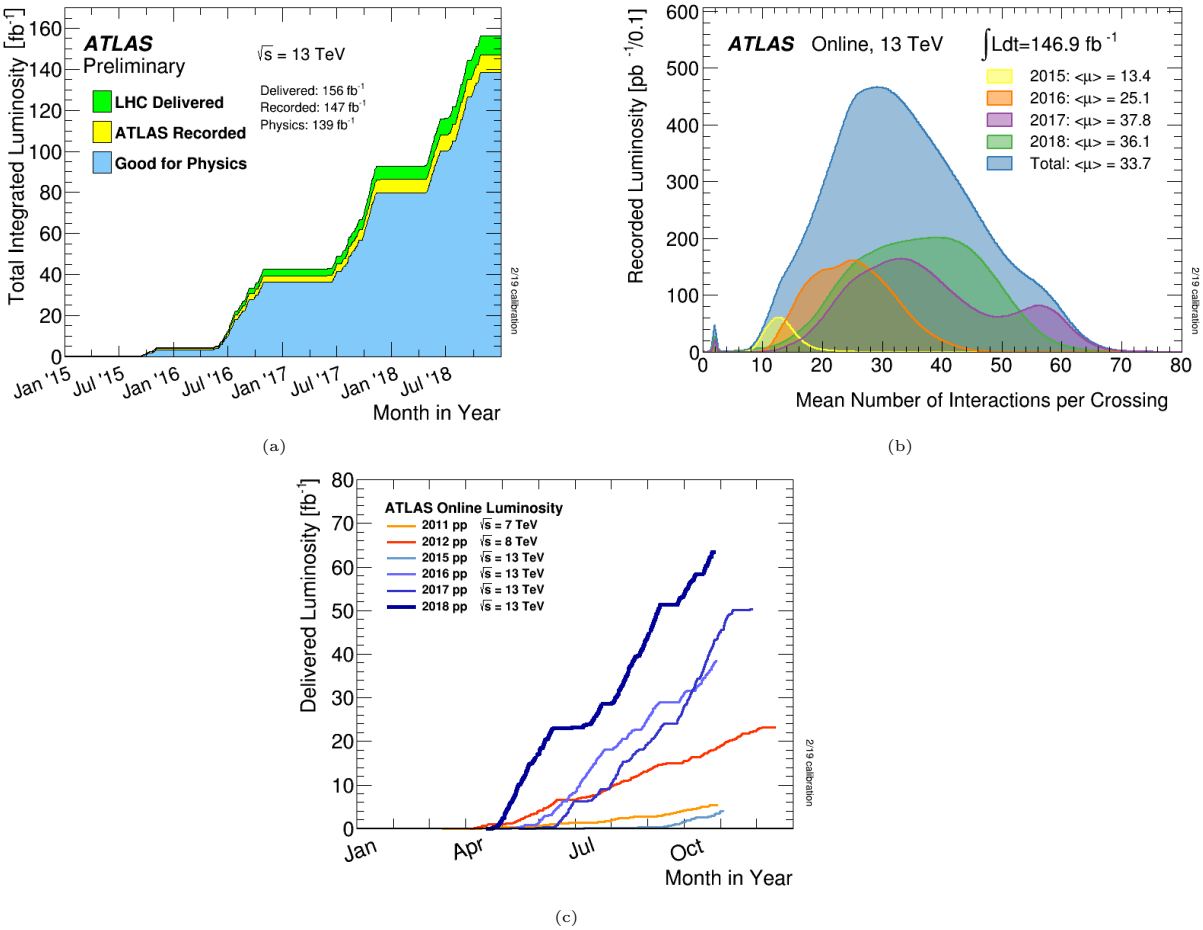


Figure 2.2: Figure (a) shows the total integrated luminosity delivered by the LHC during run 2 (green), the total ATLAS recorded luminosity (yellow) and the amount of recorded luminosity that is good for physics (blue). (b) The mean number of interaction per crossing by run 2 data taking year. (c) shows the delivered luminosity by data taking year of run 2 and run 1. Source: [36].

section is based.

The ATLAS inner detector

The closest ATLAS detector to the beam interaction point is the Inner Detector (ID). The ID is capable of reconstructing the tracks of charged particles that pass through it, and precisely measure it momenta [37]. The ID operates within a 2 T magnetic field, which bends the trajectories of charged particles and enables the measurement of their momenta. However, neutral particles do not leave tracks as they pass through the ID. The ID covers the region $|\eta| < 2.5$ and is composed of multiple layers utilizing various technologies, such as pixels and silicon strips. During collisions, the resulting radiation can damage the detectors and reduce their efficiency. In order to recover from the efficiency loss experienced during Run 1, an additional layer of pixel detectors was inserted into the inner tracker for Run 2. A scheme of the ATLAS inner detector can be found in the Figure [37].

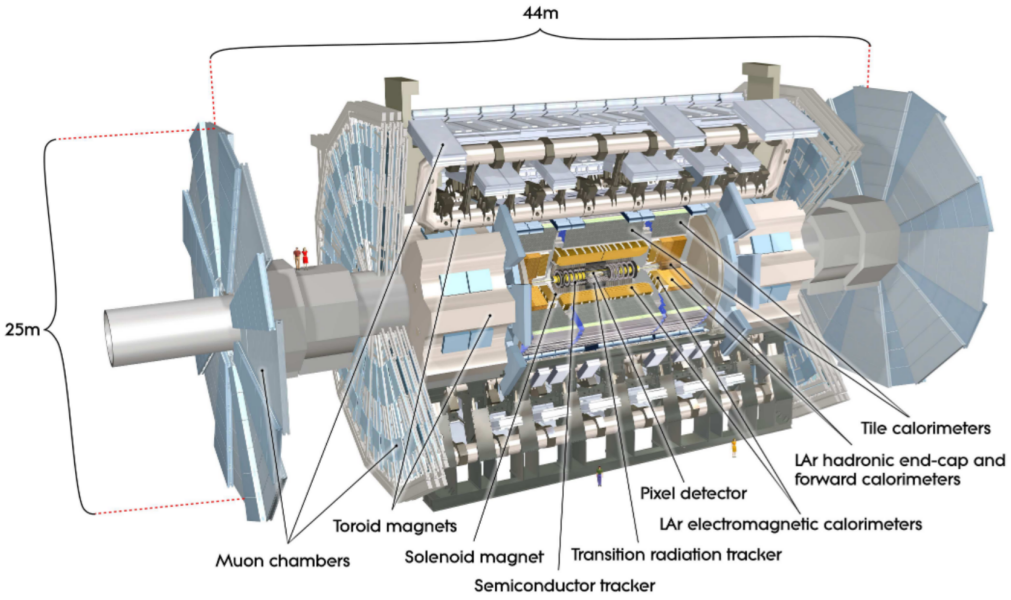


Figure 2.3: Transverse cut of the ATLAS detector. The detector has a height of 25m and a 44m extension. The detector weights around 7000 tons. The image was taken from [35].

The main components of the ATLAS inner detector are the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). The Pixel Detector is the innermost sub-detector, located just 3.3 cm from the LHC beam line. It is used to measure the position of particles with high precision. The Pixel Detector is designed such that a particle traverses three silicon sensors independent of its pseudorapidity.

The Semiconductor Tracker surrounds the Pixel Detector and is used to detect and reconstruct the tracks of charged particles produced during collisions. Its design is such that each particle crosses at least four layers of silicon.

The outermost component of the Inner Detector is the Transition Radiation Tracker (TRT). It is a gaseous detector composed of straw tubes filled a gas mixture of xenon and carbon dioxide. As charged particles pass through the straws, they ionize the gas, creating a detectable electric signal. This is used to reconstruct the tracks of the particles, and due to the phenomenon known as transition radiation [37], it provides information about the particle's flavour.

The inner detector has the capability to accurately measure the transverse momenta (p_T) of charged particles. The ATLAS collaboration performed a measurement of the resolution of muons p_T measurements in the inner detector using the decays $Z \to \mu\mu$ and $W \to \mu\nu$. This measurement was conducted in four distinct η regions, as defined in Equation 2.1, which correspond to relative p_T resolutions. The transverse momentum resolution of the inner detector was measured and parameterized [38] to be approximately:

$$\begin{aligned}
\frac{\sigma(p_T)}{p_T} &= 49\% \cdot p_T \oplus 1.6\% && ,|\eta| < 1.05, && ,\text{barrel} \\
\frac{\sigma(p_T)}{p_T} &= 95\% \cdot p_T \oplus 2.6\% && ,1.05 < |\eta| > 1.7, && ,\text{transition} \\
\frac{\sigma(p_T)}{p_T} &= 139\% \cdot p_T \oplus 3.4\% && ,1.7 < |\eta| > 2.0, && ,\text{end-cap} \\
\frac{\sigma(p_T)}{p_T} &= 14\% \cdot p_T \oplus 4.1\% && ,|\eta| > 2.0 && ,\text{no TRT}
\end{aligned} \tag{2.1}$$

As the transverse momentum of a particle is measured by the amount of its deflection in the magnetic field, a higher p_T value results in a lower curvature and makes it more challenging to measure the particle's p_T accurately. This is the reason why the resolution deteriorates with increasing p_T values.

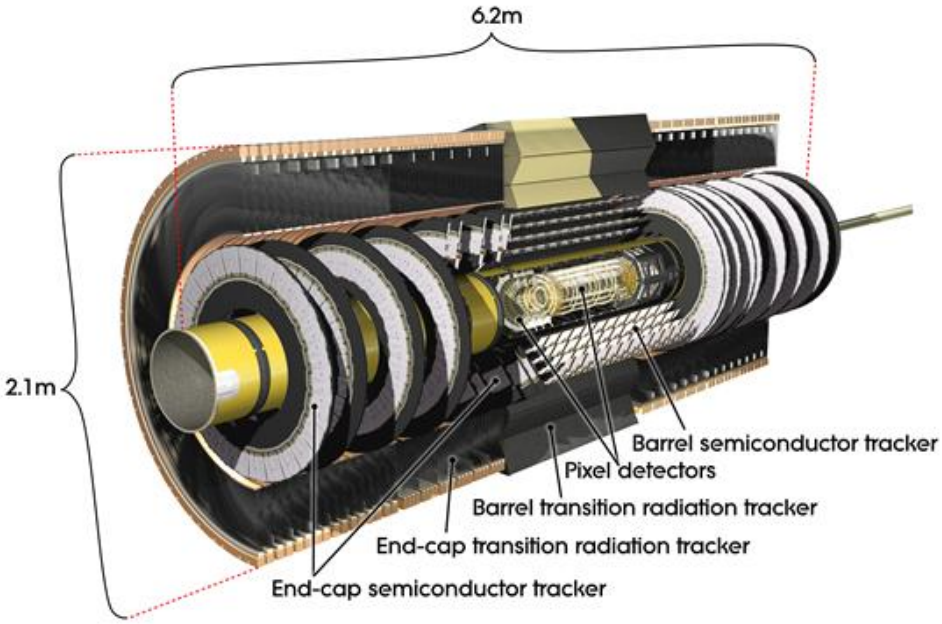


Figure 2.4: The ATLAS inner detector [37]

The calorimeters

The ATLAS Calorimeter system surrounds the Inner Detector, as shown in Figure 2.3. It is responsible for measuring the energy of electromagnetic and strongly interacting particles. While the calorimeter system also provides measurements of the particle's position, these measurements are less precise than those provided by the Inner Detector.

The calorimeters in the ATLAS detector are composed of sampling detectors, where sampling refers to the designs choice of the calorimeters, that utilizes alternating layers of absorbers and active readouts , relying on ionization or scitilation light which is proportional to the energy loss in the medium for energy measurments. The ATLAS calorimeter system is non-compensating. This means that the energy lost by particles due to nuclear collisions is lost to nuclear recoil and dissociation, effects that are not captured by the active readout. As a result, the measured energy requires correction from Monte Carlo simulations to account for these effects

The calorimeter system is separated into an electromagnetic calorimeter (ECal) located closest to the Inner Detector and a hadronic calorimeter (HCAL) surrounding the ECal. Both calorimeters cover the region of $|\eta < 4.9|$ [35], they are composed of high-density materials causing the particles which traverse them to lose all of their energy. Part of the lost energy is captured by readouts and used for energy measurements.

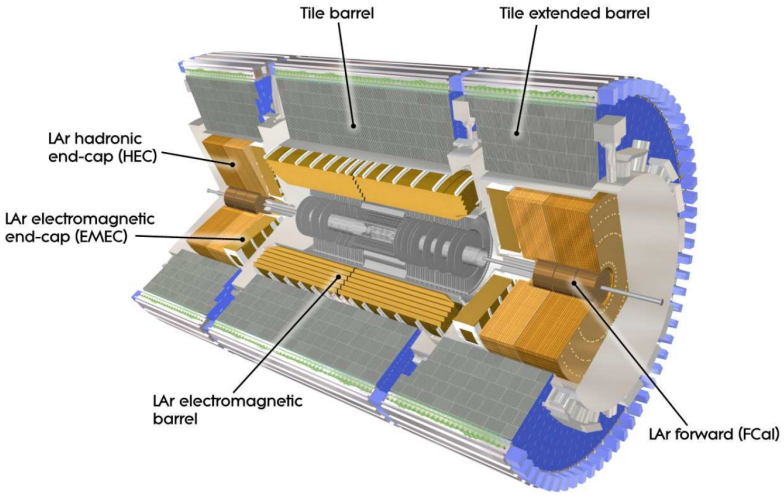


Figure 2.5: The ATLAS calorimeter system [35]

The ECal

The Electromagnetic Calorimeter (Ecal) of the ATLAS experiment measures the energy deposited by electromagnetic interacting particles, such as electrons and photons. Although muons interact via the electromagnetic force, they are relatively heavy compared to other particles, have high penetrability, and can easily pass through the Ecal without

being significantly stopped or absorbed, resulting in only a small amount of energy being deposited in the Ecal and making their measurement challenging. Because of that muons have a specif detector to measure it properties, which will be described later in the text.

The electromagnetic calorimeter (ECal) in the ATLAS detector is divided into three regions: the central LAr electromagnetic barrel, which covers the central region of $|\eta| < 1.475$; the LAr electromagnetic end-cap (EMEC), which covers the region of $1.375 < |\eta| < 3.2$; and the LAr forward calorimeter (FCal), which covers the region of $3.1 < |\eta| < 4.9$, as shown in Figure 2.5.

The Ecal is constructed of alternating layers of lead and liquid argon [35]. Lead, due to its high density and atomic number, acts as an absorber, while the liquid argon serves ass an ionization chamber. Incoming electrons radiate photons when interacting with the detector material. These photons produce electron-positron pairs which then develops into a electromagnetic shower inside the calorimeter [35]. The same happens with photons, the photons deposit their energy through multiple interactions with the atoms of the calorimeter material, producing electromagnetic showers.

The choice of liquid argon for the electromagnetic (EM) and forwardmost calorimeters was based on its radiation hardness and speed. With its relatively high density, liquid argon increases the probability of interaction between particles and the detector material, facilitating the accurate measurement of particle energies. In addition to its density, liquid argon has a high radiation hardness, meaning that it can maintain its functionality in the presence of high levels of ionizing radiation [35]. Furthermore, liquid argon exhibits self-healing properties that enable it to recombine ion-electron pairs upon exposure to ionizing radiation. This property contributes to the low maintenance requirements of the material as it exhibits low degradation when exposed to radiation. Unlike some other materials, liquid argon is stable and does not degrade over time, even when exposed to high levels of ionizing radiation. Therefore, the use of liquid argon ensures that the performance of the ATLAS detector remains consistent over its lifetime [39].

The hadronic calorimeter

The Hcal follows the Ecal and is responsible for measuring the energy of strongly interacting particles like hadrons. The Hcal is divided into a barrel ($|\eta| < 1.7$) and an end-cap region ($1.5 < |\eta| < 3.2$). In the barrel region steel is used as an absorver, and plastic scintillators are used to detect the light produced by the hadronic shower, allowing the energy deposited by the shower to be measured. The end-cap region of the Hcal is composed of liquid argon and layers of cooper, as plastic scintilators are expected to degrated at high luminisity, due to the higher radiation in the forward region.

The foremost component of the calorimeter system is the LAr forward (FCal) calorime-

ter, which covers the region $3.1 < |\eta| < 4.9$. The FCal comprises three separate layers, each with a specific energy measurement function. The first layer is optimized for electromagnetic energy deposition, consisting of a dense stack of lead plates and liquid argon (LAr) electrodes, facilitating the precise measurement of the energy of electromagnetic particles. In contrast, the latter two layers are designed to measure hadronic energy and are composed of a combination of copper and tungsten plates and LAr electrodes.

The pesudo-rapidity region between 1.37 and 1.52 corresponds to the difficult transition region between the barrel and end-cap calorimeters [35]. This "crack" region is not used for photon identification nor for precision measurements, and so, particles within this crack region will not be selected for this analysis.

With respect to the resolution of the calorimeter system, studies were conducted using π and electron test-beams in an independent setup at CERN before the commencement of the LHC to measure the energy resolutions of sub-sections of the ATLAS calorimeter. These resolutions were then parameterized as shown in Equation 2.2 and documented in [40]. In the case of the ECal, the uncertainties in the measurements were similar for both the barrel and end-cap regions [35].

$$\begin{aligned}
\frac{\sigma(E)}{E} &= \frac{10\%}{\sqrt{E}} \oplus 0.7\% && ,|\eta| < 3.2 && ,\text{ECal,} \\
\frac{\sigma(E)}{E} &= \frac{56\%}{\sqrt{E}} \oplus 6\% && ,|\eta| < 1.7 && ,\text{Hcal barrel,} \\
\frac{\sigma(E)}{E} &= \frac{70\%}{\sqrt{E}} \oplus 6\% && ,1.5 < |\eta| < 3.2 && ,\text{Hcal end-cap,} \\
\frac{\sigma(E)}{E} &= \frac{94\%}{\sqrt{E}} \oplus 8\% && ,3.1 < |\eta| < 4.9 && ,\text{FCal}
\end{aligned} \tag{2.2}$$

The muon spectrometer

The measurement of muons in the ATLAS experiment requires a specialized apparatus due to their limited interaction with the Ecal. The muon spectrometer (MS) is at the outermost layer of the ATLAS detector and covers a larger region in pseudorapidity ($|\eta| < 2.7$). This spectrometer uses a combination of gas based detectors such as drift tubes, cathode strip chambers, and resistive plate chambers to detect the muons. When the muons pass through these chambers, they produce ionization in the gas-filled tubes or plates, which is then recorded to determine the muons trajectory and momentum. [35] The muon spectrometer had its resolution measured and parameterized together with the inner detector [38]m and the parameterized results are in equation (2.2).

$$\begin{aligned}
\frac{\sigma(p_T)}{p_T} &= 24\% \cdot p_T \oplus 3.75\% && ,0.0 < |\eta| < 1.05 && \text{, barrel,} \\
\frac{\sigma(p_T)}{p_T} &= 30\% \cdot p_T \oplus 8.8\% && ,1.05 < |\eta| < 1.7 && \text{, transition,} \\
\frac{\sigma(p_T)}{p_T} &= 23\% \cdot p_T \oplus 4.77\% && ,1.7 < |\eta| < 2.0 && \text{, end-caps,} \\
\frac{\sigma(p_T)}{p_T} &= 90\% \cdot p_T \oplus 4.87\% && ,2 < |\eta| < 2.5 && \text{, CSC}
\end{aligned} \tag{2.3}$$

For illustrative purposes, Figure 2.7 shows a simplied sketch of different particle signatures in the ATLAS experiment.

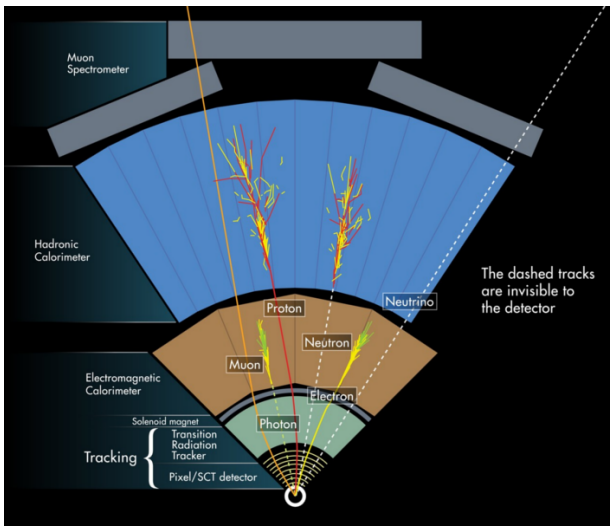


Figure 2.6: Sketch of signatures selected particles leave in the ATLAS detector. [41]

2.3 The ATLAS coordinate system

The ATLAS experiment employs a coordinate system based on the nominal interaction points. The origin of this coordinate system is defined as the nominal interaction point, with the beam direction serving as the z-axis and the transverse x-y plane being perpendicular to it. The direction of particles is usually measured in terms of the azimuthal angle ϕ and the polar angle θ. The ϕ angle is calculated around the beam axis within the transverse plane, while the polar angle is transformed into the pseudorapidity, defined as $\eta = -\ln(\tan(\theta/2))$. Pseudorapidity is favored over polar angle as differences in pseudorapidity are invariant under Lorentz boosts along the z-axis. The angular separation between two particle tracks is then defined as:

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{2.4}$$

The momentum of a reconstructed particle is represented by the magnitude of the transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$, ϕ and η as the inner detector can only measure

transverse momentum.

2.4 Magnetic fields

ATLAS uses three distinct superconducting magnet systems (composed of four large superconducting magnets) to generate high-level magnetic fields, reaching the level of Tesla. These magnetic fields are used to curve the trajectory of charged particles, allowing for accurate determination of their momentum during track reconstruction.

The central solenoid, which surrounds the ATLAS inner detector, is the first magnet that particles encounter within the detector. Its primary function is to bend the trajectories of charged particles produced in collisions along the beam axis. The curvature of the particles path provides valuable information about the particle's momentum, which can be used to determine its charge and mass. The solenoid has been engineered to produce high magnetic fields while also maintaining a minimum material thickness in front of the calorimeter [35].

In addition to the central solenoid, a system comprising three toroidal magnets is employed to measure the momentum of muons. These toroidal magnets are situated outside the central solenoid and encircle the inner detector. By generating a magnetic field perpendicular to the beam axis, they are able to deflect the trajectory of muons.

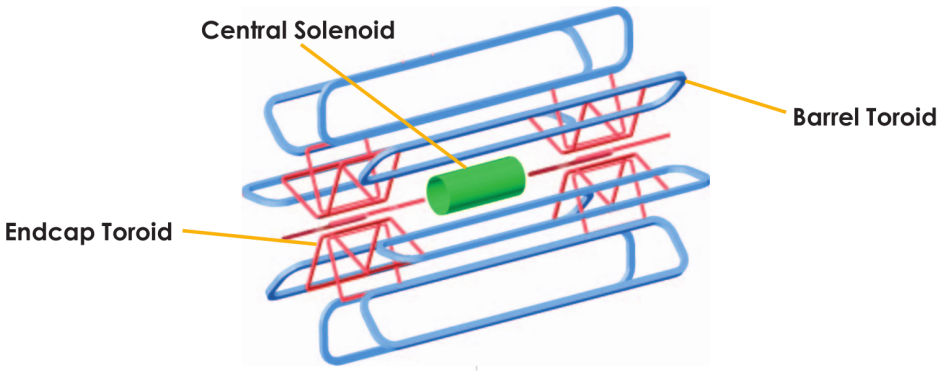


Figure 2.7: Magnetic system of the ATLAS experiment [41].

2.5 The ATLAS trigger system

Due to technical limitations, it is not feasible to record every event generated during LHC collisions, which occur at a rate of 40 MHz. The ATLAS experiment is equipped to store only 1 kHz of data, which is why a trigger system is used to select events of interest for the ATLAS physics program, and dump uninteresting events. The trigger system plays a crucial role in the data acquisition process at the LHC, and its main function is to reduce the data rate to a manageable level. By selecting events of interest, The trigger system ensures the effective and efficient utilization of the limited data processing and transmission capabilities to store the most relevant data for physics analysis.

The Run 2 ATLAS trigger system consists of two stages: the Level-1 (L1) trigger and the High Level Trigger (HLT). The L1 trigger operates on 40 million events per second, and therefore relies on low-level information from energy deposits in the electromagnetic and hadronic calorimeters and also the muon system to make quick decisions [42]. The L1 trigger reduces the initial data rate of 40 MHz to 100 kHz. The HLT, on the other hand, utilizes information from reconstructed particles that have already been processed by the ATLAS algorithms, and allows for additional processing, such as RNN and BDT based particle identifications, and reduces the data rate from 100 kHz to the desired 1kHz.

To fulfill the requirements for the rich physics program of ATLAS, events are selected in parallel by about 1500 unique event selections [42]. It is important that the trigger menu reflects the desired physics objectives for a given data collection period, while also considering the limitations of the ATLAS detector readout, online processing farm, and offline storage, in addition to the instantaneous luminosity of the LHC.

2.5.1 The tau triggers

The tau triggers in the ATLAS experiment are designed to maximize tau acceptance while rejecting background events. This is a challenging task due to the high production of jets that have similar characteristics to tau leptons in the LHC. The trigger algorithms make use of the taus narrow energy deposits and small number of charged tracks, and employ Boosted Decision Trees and Recurrent Neural Networks for tau identification.

To be stored, an event containing a tau must fire a trigger at L1 and HLT level. As different information is used in each trigger level, The different trigger levels have different p_T thresholds, as in the L1 level no energy calibration is performed, the p_T thresholds are lower than in the HLT trigger.

The trigger system has p_T thresholds set to meet the experiment data storage rates. If the p_T and other trigger selection thresholds are too loose, the experiment may not be able to save all events due to computational limitations. Therefore, single tau triggers, which

are activated when one tau in the event meets its conditions, have a high p_T threshold, and di-tau triggers, which require two taus in the event, have a lower threshold to meet the storage rate requirements. As the jets that can be misidentified as taus comes from QCD processes, they tend to have lower p_T, and to maintain the desired output event rate, the single tau trigger must have a higher p_T threshold. In contrast, the di-tau trigger can have lower p_T thresholds, as it is less likely for two QCD jets to be misidentified as two tau leptons. The di-tau trigger is useful for collecting events in the low mass region, where the events have a softer p_T spectrum, while the single tau trigger is more appropriate for collecting events in the high mass region, where the tau p_T is typically higher. For this analysis a combination of single and di-tau trigger is being used. The trigger list is available in appendix A.

Track selection is also utilized, where the number of core[6] and isolation tracks are calculated and only taus candidates with $1 \leq N_{core}^{trk} \leq 3$ and $N_{isolation}^{trk} \leq 1$ are accepted. These track multiplicity requirements achieve a significant rate reduction, enabling the application of more CPU-intensive algorithms on the accepted events [42].

In order to reduce trigger rates by suppressing background events, a new energy-dependent upper threshold on the electromagnetic isolation energy is being applied. Specifically, the tau candidate must meet the condition ($E_T^{EMiso} < E_T/10 + 2$), which takes advantage of the more collimated nature of taus [43].

The HLT tau triggers in ATLAS utilize RNN and BDT based selections for further optimization of the selection process, as stated in [42].

Electron and muon triggers

Due to their clean signature and lower background contamination compared to hadronic taus, low p_T threshold single electron and muon triggers can be used to select $e\tau_{had}$ and $\mu\tau_{had}$ events, as the QCD jets have lower probability to fire a electron or muon trigger. The low p_T thresholds and the requirement of only a light lepton makes e and μ triggers have a higher efficiency in low masses compared to tau triggers, which needs higher p_T thresholds due to the large QCD contamination. The triggers used to collect $e\tau_{had}$ and $\mu\tau_{had}$ events are available in appendix A and consist of single electron and muon triggers.

2.6 Data and MC samples

Modeling the physics phenomena that occur at the LHC is a challenging task. To accomplish this, multiple Monte Carlo (MC) generators are employed to model different phenomena, allowing for a theoretical interpretation of data.

[6] Core and isolation tracks will be defined in the next section

The different processes involved in pp collisions at the LHC are modeled by MC based techniques, which generate random events with distributions following the models at which they were generated. The modeling of the theoretical point of view, up to what we can observe in a experiment is an extremely complicated task, and so there are different generators that are specialized in producting specific types of process.

An overview of a collision process in a proton-proton collider is depicted in Figure 2.8. The hard process is shown in red in the Figure and is the process of interest for measurement. Its perturbative calculation is performed at the matrix element level, considering the probability of a parton carrying a certain momentum fraction of the proton.

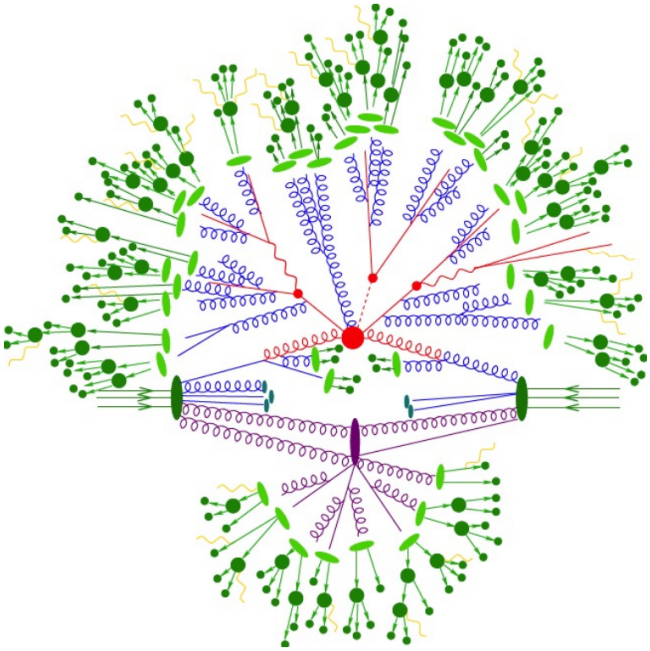
Matrix element generators calculate the probability of a specific hard scatter process occurring, taking into account the energies and momenta of the colliding partons and the properties of the particles produced in the collision. Matrix element generators are specialized for different types of hard scatter processes, depending on the physics phenomenon being studied.

The hard scattering process is accompanied by initial state radiation (ISR) and final state radiation (FSR), which are represented by the blue objects in the Figure. Initial state radiation occurs when a particle emits radiation before colliding with another particle, while final state radiation occurs when a particle emits radiation after undergoing a collision. Both ISR and FSR can have significant effects on the measurements made in high-energy particle physics experiments, as the emitted radiation can alter the kinematics of the particles involved in the collision, leading to inaccuracies in the measurements of the particles energies and momenta.

As the emitted gluons from ISR and FSR undergo hadronization, they result in particle jets, shown in green in Figure 2.8. The hadronization process refers to the conversion of outgoing quarks and gluons produced in high-energy collisions into color-neutral hadrons, such as protons, neutrons, pions, and kaons. This process occurs due to the confinement of quarks and gluons inside hadrons, preventing their direct observation as free particles.

The yellow objects represent additional QED and bremsstrahlung processes. The dark purple processes in the Figure correspond to additional parton interactions, contributing to the underlying event. To simulate the effects of pile-up in Monte Carlo event generators, additional proton-proton collisions are overlaid onto the primary collision of interest. The number of additional collisions to be overlaid is determined by the average number of proton-proton interactions per bunch crossing, which is a parameter that is measured experimentally.

In order to model the full sequence of physics phenomena at the LHC, multiple generators are required. Each generator is specialized in specific steps of the decay chain.



(a)

Figure 2.8: Sequence of processes in pp collisions considered by MC event generators. Taken from [44].

There are integrated frameworks available, such as the SHERPA[7] [45, 46] or the PYTHIA framework [47], which use a combination of perturbative QCD calculations, parton showering, and hadronization to simulate the entire process of particle collisions. On the other hand, there are other frameworks such as PowHeg[8] that are specialized only in the matrix element generation and needs the hadronization to be handled by another generator (e.g POWHEG+PYTHIA combination).

SHERPA Generator

For example, generators like SHERPA [45, 46, 49] are suitable for modeling vector bosons with jets produced alongside them due to their advanced matrix element generation techniques and parton showering algorithms. Recent measurements of the Z boson production associated with b-jets, and W boson production with jets have shown that the SHERPA generator is able to correctly model the various kinematic distributions measured within experimental uncertainties [50, 51].

In this work, all of the $Z \to \ell\ell$ ($\ell = e, \mu, \tau$) and $W \to \ell\nu$ processes were modeled using the SHERPA 2.2.11 generator, as a good description of the production alongside jets (especially b-jets) is desired in order to study some beyond the Standard Model (BSM) models like leptoquarks (LQs) and Z', as searches for these BSM models would be more

[7] **S**imulation of **H**igh-**E**nergy **R**eactions of **PA**rticles
[8] It has this name because of the powheg method it uses [48]

sensitive in the regions where the di-tau pair is produced alongside jets [19]. The matrix elements corresponding to the final states of $\ell\ell$ + jets and $\ell\nu$ + jets (where ℓ represents e, μ , or τ) have been produced with NLO accuracy for up to two extra parton emissions and LO accuracy for up to four extra parton emissions.

PowHeg + PYTHIA Generator

Other generators, such as POWHEG+PYTHIA [48, 47], are better suited for modeling top quark production as they are specifically designed to accurately simulate the production and decay of top quarks. The POWHEG is a next-to-leading-order (NLO) matrix element generator that can calculate the hard scattering process of top quark production at NLO accuracy. The PYTHIA generator, on the other hand, is a leading-order event generator that is used to simulate the fragmentation and hadronization of the final state particles produced by the hard scattering process. Together, these two generators can provide a detailed and accurate simulation of top quark production and decay, making them well-suited for studying top quark physics at the LHC. The $t\bar{t}$ and single top predictions were simulated with the POWHEG + PYTHIA 8 generators.

In order to compare data to Monte Carlo (MC)-generated events, it is necessary to simulate how these generated particles interact with the ATLAS detector. Due to various detector effects, such as trigger, detection efficiency, and background misidentification, it is not possible to directly compare events at the particle level to the data. To account for detector effects, a model of the ATLAS detector built using Geant4 [52] is utilized, and the simulated events are reconstructed with the standard ATLAS reconstruction software [35].

2.6.1 Signal modeling

In order to accurately study the mass estimation in the $Z \to \tau\tau$ process, it is crucial to model the process with precision. In this work, we utilized SHERPA 2.2.11 to generate the primary signal samples, as it has been shown to produce accurate results [50]. Additionally, alternative samples were generated using POWHEG+PYTHIA 8 configuration to compare results.

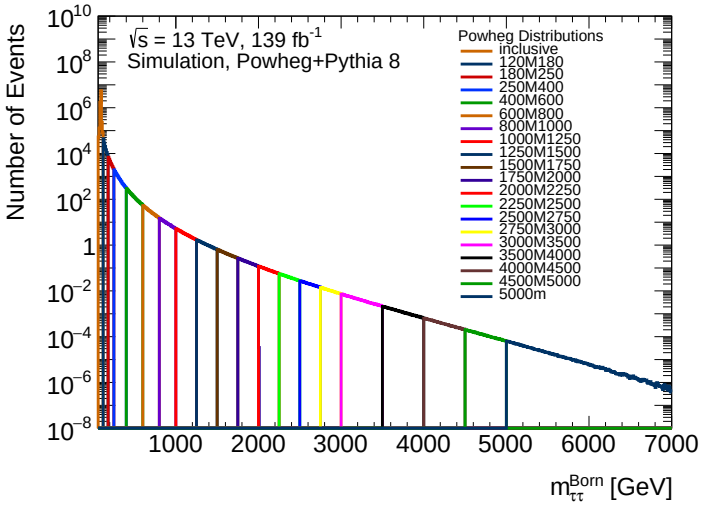
However, it should be noted that the ”bulk” samples configuration is useful for generating events in the vicinity of the Z peak region, the term ”bulk” refers to the fact that these samples are generated to cover a mass range, centered around a specific mass or energy of interest. Nonetheless, this configuration may struggle to produce events in the high mass region, leading to inadequate sample sizes in the region of interest for this study. Therefore, it is imperative to use a configuration that is better suited to generate

events in the high mass region. To achieve this objective, statistical enhanced samples were employed in our analysis.

2.6.2 High Mass enhancement

The tails of the mass distribution are of central importance for a high mass analysis. As the high mass region has a cross-section orders of magnitude lower than the resonance, typical MC techniques are not able to efficiently populate that region and are unable to yield the necessary statistical power needed for the analysis.

To produce a significant and necessary quantity of events in the tails of the distributions, a slicing technique is commonly used [53]. This technique involves the generator producing events only in a specific interval of the mass distribution. For example, the POWHEG generated samples use such a technique to enhance the statistics in the high mass region by slicing the phase space into 20 different regions and generating events only inside each interval. Slicing gives good control of the statistics but often leads to slow event generation times, as many events need to be produced to populate each slice, and inefficient usage of events for the most extreme slices. The distribution of the born mass for of all 20 mass slices of the $Z \to \tau\tau$ process generated with the POWHEG generator is available in Figure 2.9.



(a)

Figure 2.9: POWHEG mass born mass distribution of inclusive and slice samples. The different slices are named according to the interval in which the mass was generated. For example, the sample 180M250 corresponds to a generated mass within the 180 to 250 born mass interval.

In the SHERPA generator, a technique to enhance the statistics of certain observables by biasing the event generation according to a specific kinematic variable is used. This means that a distribution with a higher concentration of events in the high mass region

is produced with a different cross-section than the physical one. However, the physical cross-section can be recovered through the calculation and application of weights [53].

To understand how its done, we can look at the Monte Carlo event generator as a numerical integrator of the matrix element phase space:

$$\hat{\sigma}_{ab\to m} = \int_{\Omega} d\hat{\sigma}_{ab\to m} \tag{2.5}$$

We can also look at an integral as:

$$I_f = \int_{x_1}^{x_2} f(x)dx = (x_2 - x_1)\,\langle f(x)\rangle \tag{2.6}$$

And so, as:

$$\langle f(x)\rangle = \frac{1}{N}\sum_{i}^{\infty} f(x_i) \tag{2.7}$$

We can see that:

$$\int_{x_1}^{x_2} f(x)dx \approx \frac{(x_2 - x_1)}{N}\sum_{i}^{N} f(x_i) \tag{2.8}$$

Therefore, a Monte Carlo sample of N events is obtained via random sampling of a p-dimensional PDF. This is a very simplified view of how the Monte Carlo method works, just for visualization and explanation purposes.

Lets say we want to study a variable whose probability density function (PDF) is $1/(1+x^2)$, but the events of interest to us have higher values of x. Since the majority of the samples drawn from $f(x)$ will be close to zero, only a few will be sampled from higher values of x. One way to increase the number of sampled higher values of x is to multiply $f(x)$ by a function that increases the values of $f(x)$ at higher values and apply weights to correct them back to the original cross section.

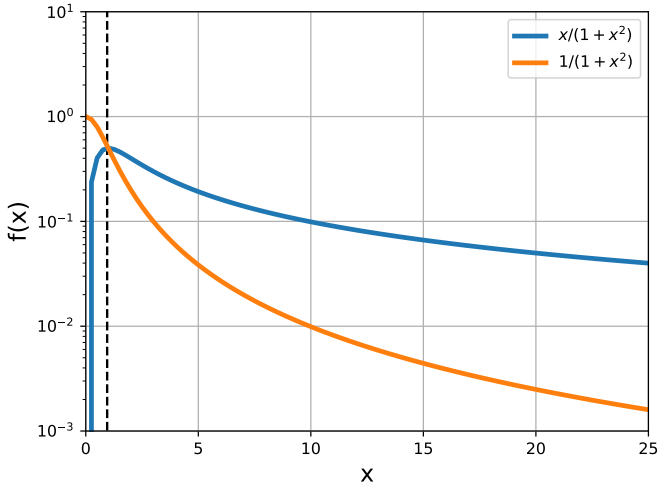
If we use the same line of thinking to the MC sampling, we can modify equation (2.4) and add a $e(\vec{\mathcal{O}})$ with the objective of sampling more events in a given region, we would have:

$$\hat{\sigma}_{ab\to m} \to \hat{\sigma}^{'}_{ab\to m} = \int_{\Omega} d\hat{\sigma}_{ab\to m}\cdot e(\vec{\mathcal{O}}) \tag{2.9}$$

$e(\vec{\mathcal{O}})$ is the user defined function parameterised according to a vector of multiple observable $\vec{\mathcal{O}}$. This function needs to be chosen carefully in order to avoid a large spread of weights. In this work, the $e(\vec{\mathcal{O}})$ function is used to enhance the mass sampling in the high mass region, and the function used is [53]:

$$e(\vec{\mathcal{O}}) = e(m_V) = m_V^2/(20^2) \tag{2.10}$$

Here, m_V refers to the mass of the vector boson, specifically the Z boson. An example of how multiplication by a function can enhance the function on high values of x is illustrated in Figure 2.10 as the plots of the function $1/(1+x^2)$ and $x/(1+x^2)$ are shown. However, it is important to note that the 'physical' cross section must be brought back later by using weights. With the increase in sampled events in the tail, there is potential to increase the statistical power in the region. Care must be taken when multiplying a function by x or x^2 as the distributions at low x or mass may not agree, resulting in large weights needed to bring back the 'physical' cross section, which have a large variance. The issue could be avoided by conducting sampling beyond a certain threshold where the probability distributions exhibit comparable characteristics, a strategy adopted in SHERPA samples. Specifically, statistically improved samples were drawn exclusively from masses exceeding 120 GeV.



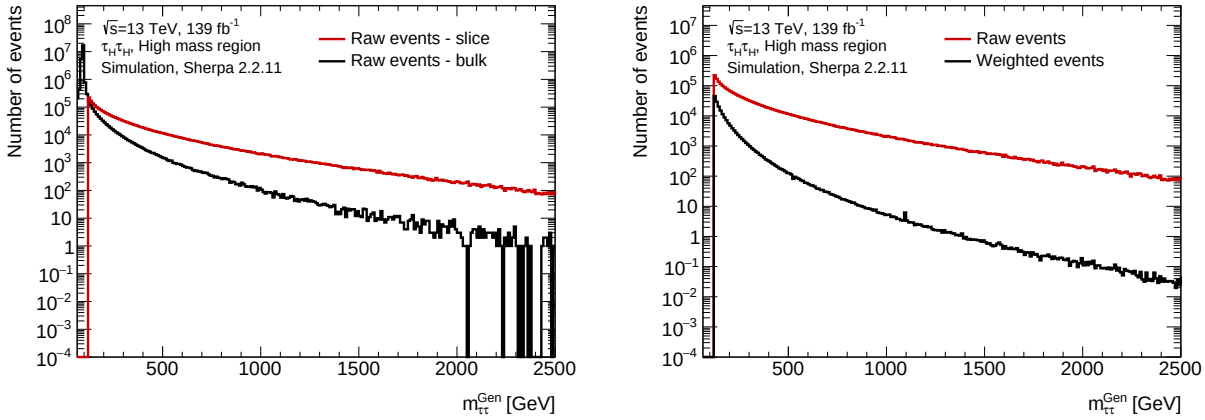
(a)

Figure 2.10: Nominal and enhanced in x function.

The statistical enhancement of a single slice provides a practical way to increase the population of the high mass region and enhance the statistics in that region of phase space. For example, we use a single slice where the event generation only occurs for events with a mass above 120 GeV.

For instance, as depicted in Figure 2.11 (a), the distribution of raw events produced with bulk and the mass enhanced slice configurations is presented. It can be observed that in the high mass region, the slice exhibits a significant increase in the number of generated events compared to the bulk configuration, thereby resulting in a desired enhancement of statistics. Furthermore, Figure 2.11 (b) illustrates the comparison between the raw

events (represented by the red curve) and the weighted number of events (corresponding to the physical number of events expected, represented by the black curve) of the mass enhanced slice, this highlights the how the statistical enhancement technique enhances the event generation w.r.t the physical cross section. Through the implementation of this procedure, we have now obtained larger statistics in the high mass region.



(a) (b)

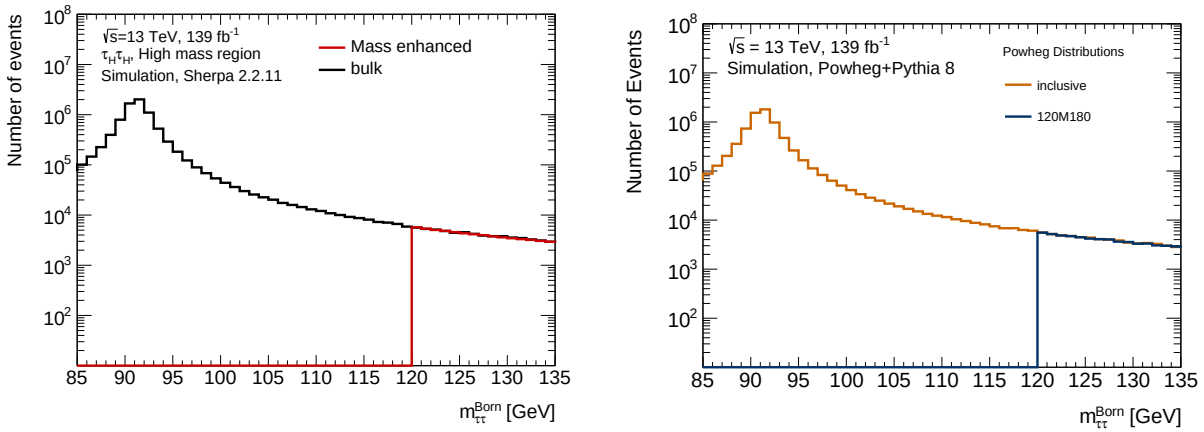
Figure 2.11: (a) Number of raw events of the SHERPA bulk and slice configurations. (b) The red curve shows the number of raw events of the SHERPA mass slice and the weighted distribution in black

In Figure 2.12 we can have a look at the bulk and mass enhanced configuration distributions around the Z resonance. The distributions show the generated mass of event , namely the born mass, and as stated the mass enhanced samples only have events above the value of 120 GeV of mass. To achieve a comprehensive description of the process over the entire mass range, the two samples, namely the bulk and mass-enhanced samples, will be combined into a single segment through a "stitching" process. In this process, events from the bulk sample with masses below 120 GeV will be retained, whereas events beyond this threshold will be discarded. Doing it so, we can achieve a good statistical power in a broader mass range.

If we have a look at a broader mass interval as Figure 2.13 (a) shows, the improvements in statistic of the mass enhanced sample with refernece to the bulk configuration. Additionally, we are able to calculate the statistical precision of a Monte Carlo sample by utilizing the following equation [53]:

$$\delta = \frac{1}{\sqrt{N}} \sqrt{1 + \frac{s_{w_i}^2}{< w >^2}} \tag{2.11}$$

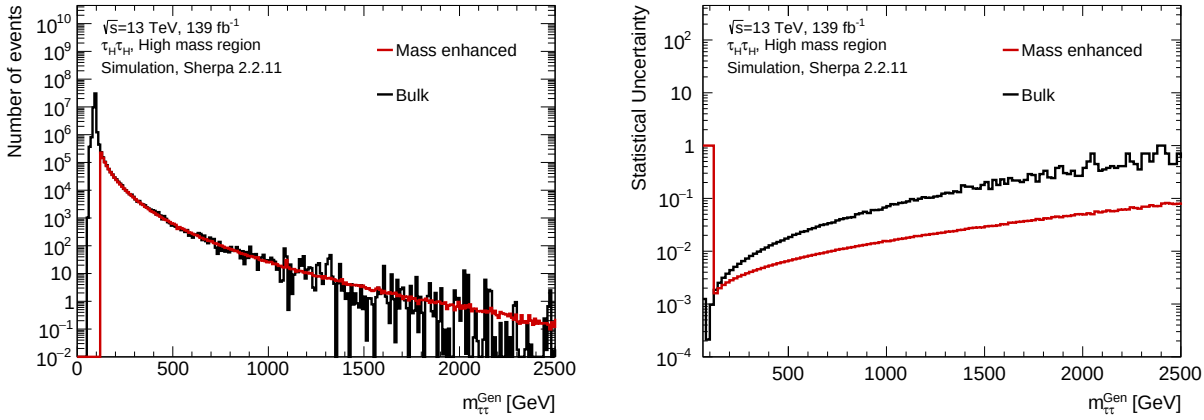
$< w >$ represents the mean value of the weight distribution and $s_{w_i}^2$ represents the variance of the weights. This quantity serves as a method for measuring the statistical uncertainty of the studied samples. As depicted in Figure 2.13 distributions, the weighted distributions of the bulk and slice are presented, in addition to the statistical uncertainty



(a) (b)

Figure 2.12: (a) Bulk and mass enhanced SHERPA 2.2.11 born mass distributions (b) POWHEG inclusive and 180M250 mass slice distributions

for both samples. As anticipated, the mass-enhanced slice exhibits larger statistics in the high mass region and displays less weight variation in comparison to the bulk sample.



(a) (b)

Figure 2.13: (a) Weighted distributions of the SHERPA mass slice and bulk configurations and the (b) statistical uncertainty.

2.6.3 Background samples

Besides the signal, the modeling of the processes that can be misidentified as $Z \to \tau\tau$ is also important, as these events are present in the measured data. There are many processes that can mimic the Drell-Yan $Z \to \tau\tau$ signature in the ATLAS detector, such as $Z + jets$ ($Z \to \ell\ell; \ell = e, \mu$), $W + jets$ ($W \to \ell\nu; \ell = e, \mu, \tau$), and $t\bar{t}$. The leading order diagrams for these processes are shown in Figure 2.14.

In particular, in the $\tau_{lep}\tau_{had}$ channel, the $Z + jets$ background can be misidentified as a $Z \to \tau\tau$ when an electron or muon is misidentified as a hadronic tau, or a jet produced alongside a Z is misidentified as a hadronic tau. To reduce this kind of background, only events containing one light lepton are accepted in the $\tau_{lep}\tau_{had}$ channel. Additionally, events with a misidentified tau, as determined by the RNN ID algorithm, are vetoed.

The $W+jets$ are present both in $\tau_{had}\tau_{had}$ and in the $\tau_{lep}\tau_{had}$ channel, and a jet produced alongside a W boson is misidentified as a hadronic tau. Decays involving top quarks are also important as tops can decay into taus, and top decays are usually accompanied by b-jets, which can also be misidentified as a tau lepton. See the next chapter for a detailed explanation of the τ decay modes.

These backgrounds are simulated using SHERPA 2.2.11 for $Z + jets$ and $W + jets$ backgrounds, and POWHEG+PYTHIA 8 is used for single top and $t\bar{t}$ production.

A more detailed discussion of the background, including the application of mass estimation methods on both backgrounds and data, is presented in Chapter 8.

The QCD multi-jet background is also an important background for the $Z \to \tau\tau$ study, but this background is not well modeled by MC due to extremely low acceptance which leads to poor statistics [51], and data driven methods like the fake factor are needed to correctly model this background [18], and so, they will not be considered in this work.

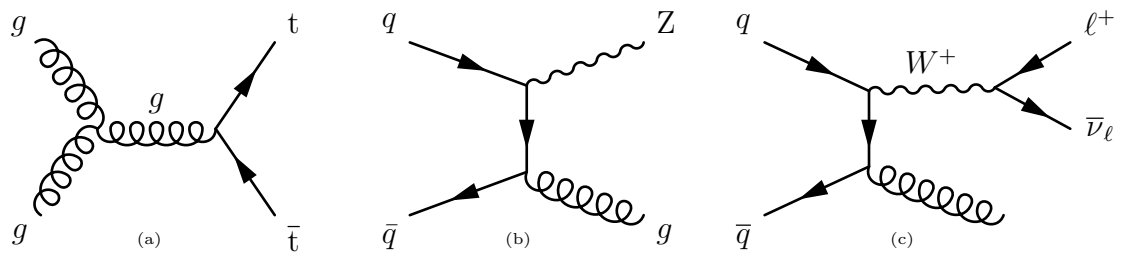


Figure 2.14: Feynman diagrams for the (a) top quark pair (b) Z + jets (c) W + jets process. The diagrams were produced using the Tikz toll [22].

2.6.4 Data samples

The analysis and all of the results presented in this work that contain data samples utilize the full LHC Run 2 data collected by the ATLAS experiment, which is in a suitable state for use. These data correspond to 139 fb^{-1} of pp collisions at 13 TeV. Table 2.1 illustrates the GoodRunsLists utilized.

MC production is divided into campaigns that correspond to specific running periods of the LHC, characterized by specific center-of-mass energy, geometry, and production conditions. In Run 2, the MC campaigns were divided into mc16a (2015 and 2016), mc16d (2017), and mc16e (2018), with each campaign reflecting the specific data-taking conditions of its corresponding year.

Table 2.1: A summary of the GoodRunLists used for the different data years.

Year	GRL
2015	data15_13TeV/20170619/physics_25ns_21.0.19.xml
2016	data16_13TeV/20180129/physics_25ns_21.0.19.xml
2017	data17_13TeV/20180619/physics_25ns_Triggerno17e33prim.xml
2018	data18_13TeV/20190318/physics_25ns_Triggerno17e33prim.xml

CHAPTER 3

SELECTION AND DATA

This section provides a detailed description of all the physics objects used in this analysis, including their reconstruction and selection criteria.

3.1 The tau lepton

The tau lepton is the main object in the analysis. The tau is a third generation lepton with a mass of 1.77 GeV, a very short lifetime and is the only lepton capable if decaying hadronically [23]. Due to its short lifetime, it decays before reaching any detector apparatus, making direct measurement of the tau impossible. In order to study the tau lepton, one must observe its decay products. The tau lepton always decays into a W boson and a tau neutrino ($\tau \rightarrow W\nu_\tau$). As the tau decay chain always includes at least one neutrino, where even more neutrinos can be produced if the W boson decays leptonically, the tau lepton information before it decays cannot be fully reconstructed, resulting in loss of event information due to the presence of neutrinos. Figure 3.1 (b)and (c) show the decay chain of a tau lepton. The tau decays decays hadronically 65% of the times and leptonically 35% [26].

The selection of hadronically decaying taus is of major importance in this analysis, as the hadronic tau is present in all decay channels of interest. The hadronic decay of the tau lepton results in a different multiplicity of charged and neutral particles, which are reconstructed as a jet.

As the tau decay products leave traces of deposited energy in the individual calorimeter cells, these signals are further clustered into a a three-dimensional topological clustering of individual calorimeter cell signals, using the TopoClusters [54] algorithm. These clusters are further classified using a local hadronic calibration algorithm [55]. The cluster clas-

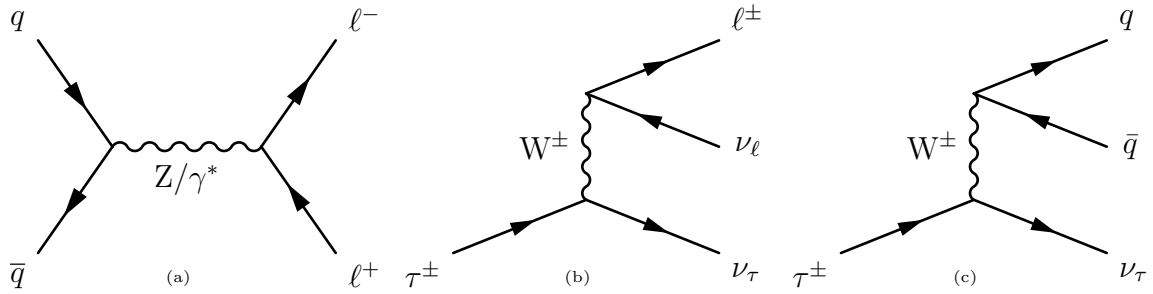


Figure 3.1: Feynman diagrams for the (a) Drell-Yan (b) leptonic tau decay (c) hadronic tau decay process.

sification is made to distinguish between clusters dominated by electromagnetic deposits and those dominated by hadronic interactions. And so, the calibrated clusters serve as input to the $anti-k_t$ jet algorithm [56].

Candidates for hadronic tau decays are built from jets reconstructed using the anti-kt algorithm. The tau leptons hadronic decays have the characteristic that their decays are more collimated than the quark or gluon-initiated jets. A schematic of a tau and a QCD jet can be seen in Figure 3.2. A core region can be defined around the reconstructed jet with < 0.2, and tracks inside this core region satisfying $p_T > 1$ GeV and having at least 7 associated hits in the pixel and SCT detectors are associated with the hadronic tau candidate. The energy of the hadronic tau decay candidate is calculated as the total energy deposited inside the core cone [43].

Another cone with a $\Delta R = 0.4$ is used to collect tracks with $0.2 < \Delta R < 0.4$. This helps to evaluate the isolation of the hadronic tau candidate, as the hadronic tau decays are more collimated than QCD Jets, leaving the isolation cone more empty, as the majority of the energy is in the core cone ($\Delta R = 0.2$).

The energy and four-momentum of the tau calculated by summing the energy inside the $\Delta R < 0.2$ is not the final energy estimate. An energy calibration is further applied into the reconstructed tau , which tries to correct for detection imperfections as energy deposited in dead material or pile-up contributions. The method of further calibrating the energy using the calorimeter only information is called baseline, and has a better performance for high p_T tau leptons [57].

For low p_T tau leptons another algorithm was developed by the ATLAS collaboration called tau particle flow [57]. It classifies the tau into its decays modes, and uses that information to better reconstruct the tau lepton energy. The hadronic decays of the taus occurs trough five dominant decay modes that have one or three charged hadrons (h) neutral pions (π^0) and of course a tau neutrino. The branching fraction of such decay modes is available in table 3.1. The tau decays that contains one charged pion are usually

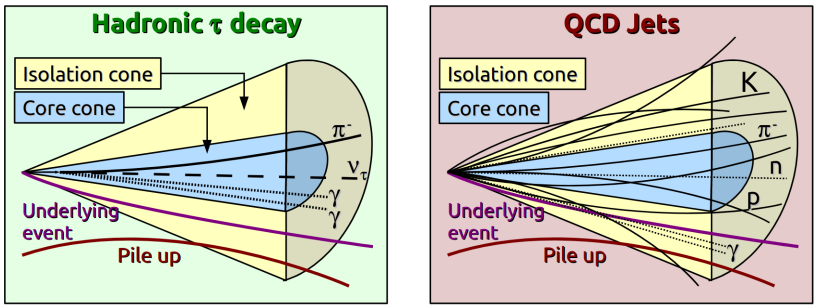


Figure 3.2: Tau and QCD jet decay sketch. Taken from: [58]

called one prong and containing three charged pions decays are called three prong.

Table 3.1: The five dominant tau decay modes. The $\geq$ means that at least there is that number of neutral pions in the tau decay chain [59].

Decay mode	Branching fraction
$h^{\pm}$	11.5
$h^{\pm}\pi^0$	30.0
$h^{\pm} \geq 2\pi^0$	10.6
$3h^{\pm}$	9.5
$3h^{\pm} \geq 1\pi^0$	5.1

The charged hadrons h from tau decays have sufficient lifetimes to interact with the detector before decaying and are thus considered stable in the tau particle flow [57]. The neutral pions decay almost exclusively to a pair of photons, which in turn typically decay into an electron-positron pair if they interact with the beam pipe or inner-detector material. Modes with more neutral pions tend to have wider showers and may be misidentified as quark or gluon-initiated jets, resulting in lower identification efficiency.

Differentiating between the various tau decay modes is important for a number of analyses, and being able to distinguish between one-prong and three-prong taus is crucial for certain mass estimation methods as the kinematics of the decay modes differ. To classify the tau decay mode, a Boosted Decision Tree (BDT) algorithm is utilized, where it classifies the tracks of the tau candidate. The performance of the BDT in classifying the tau decay modes is presented in Figure 3.4. The plot displays the generated decay modes at the particle level on the x-axis and the reconstructed decay modes using the BDT at the detector level in the y axis. The BDT demonstrates a good performance in guessing the number of charged particles in the tau decays, but less so in estimating the number of neutral pions. These decay modes, tau reconstruction and identification algorithms were developed by the tau combined performance group, and are available to physics analysis in the ATLAS to use.

The largest contributions to misclassification arise from events with $h^{\pm} \pm 2\pi^0$, where the energy deposits of the neutral pions are merged into a single cluster or one of the

π^0 fails selection. Discrimination of the $3h^{\pm}$ and the $3h^{\pm} \geq 1\pi^0$ decay modes is also challenging as the energy deposits of the neutral pions are relatively soft in comparison to those of the $h^{\pm}$ deposits.

After the tracks classification additional treatment to reconstruct the neutral pions four-momenta and disentangle the charged hadrons is made. The tau four momentum reconstruction is done by summing up the momentum of all charged hadrons and neutral pions constituents. The resulting four momentum is used to set the $\tau_{had-vis}$ direction. Although the tau particle flow is excellent at low p_T at higher p_T its performance is degraded. At high p_T the calorimeter based technique has a excellent resolution but its degraded at low p_T. Because of that, improved energy calibration has been introduced using the both techniques energy estimation, using a BRT (Boosted Regression Tree). The performance of the calorimeter based (baseline), tau particle flow (substructure) the BRT (combined) and the run 3 proposed energy scale algorithm (Final) are available in Figure 3.3.

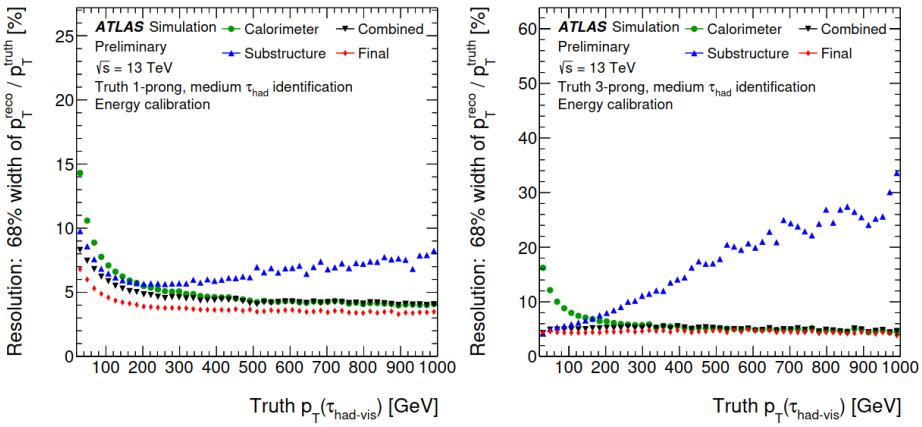


Figure 3.3: Tau energy scale algorithms for one prong (left) and three prong (right) tau decays resolution. [59]

3.1.1 Identification

The tau lepton reconstruction method described above provides limited discrimination against quark and gluon-initiated jets. To enhance the rejection of jet backgrounds ATLAS uses a Neural Network (NN) classifier. The classifier is trained separately for one-prong and three-prong taus, as the decays are significantly different. Information about the impact parameters, as taus typically originate from displaced vertices, and the energy deposited in the calorimeter and the collimation of this energy, are used as inputs for the neural network to reject QCD backgrounds.

For Run 2, a Boosted Decision Tree (BDT) was initially used for tau identification, but later in the run, a Recurrent Neural Network (RNN) based identification network was

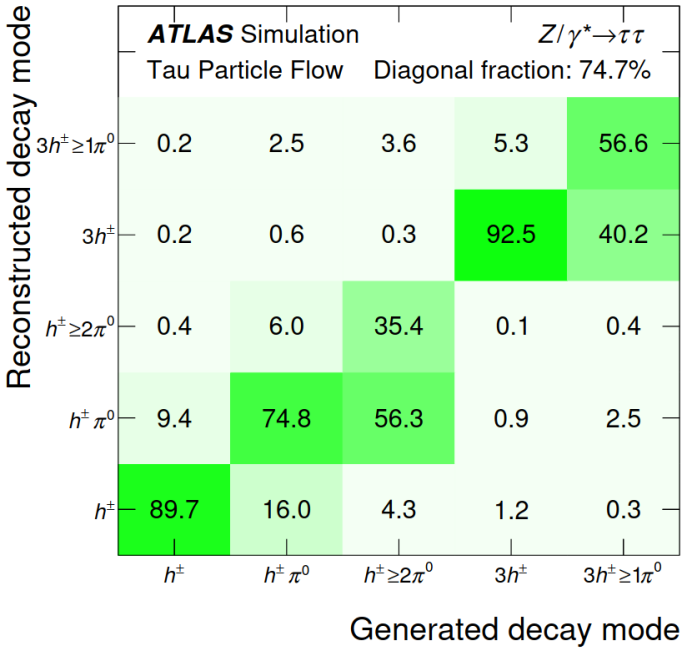


Figure 3.4: Tau hadronic decay mode classification efficiency matrix. [57]

employed. Due to its superior performance in terms of signal acceptance and background rejection, the RNN is used in this analysis. The RNN is trained with signal ($Z/\gamma \to \tau\tau$) and background ($di-jet$) samples. The RNN utilizes as input a combination of individual tracks and cluster variables associated with the τ candidate, as well as high-level observables calculated from track and calorimeter quantities. The RNN is trained to output a value from 0 to 1, where a value of zero indicates that the network is confident that the hadronic tau candidate was produced from a jet, and a value of 1 indicates that the network is confident that the tau candidate is a genuine tau.

The distribution of the RNN tau identification score for a validation sample for one-prong and three-prong taus is shown in Figure 3.5. The RNN score has been transformed so that it corresponds to the fraction of rejected true τ_{had}^{vis}.

The RNN provides a good separation between 'signal' taus and jets. Different analyses may require different signal purity and background rejection, so there are four available working points to be used in tau identification. These working points, labeled as Tight, Medium, Loose and Very Loose, correspond to different background and signal acceptance values. In Figure 3.6, the signal efficiency versus inverse background efficiency is shown. The signal efficiency is the fraction of signal that is kept after a cut in the RNN score is made, and the inverse background efficiency is defined as 1.0/background efficiency, which indicates the fraction of background that is rejected for each RNN score value. All RNN score values are tested, and the signal efficiency and background rejection for each curve is computed, resulting in a ROC curve available in Figure 3.6. The red curve remains

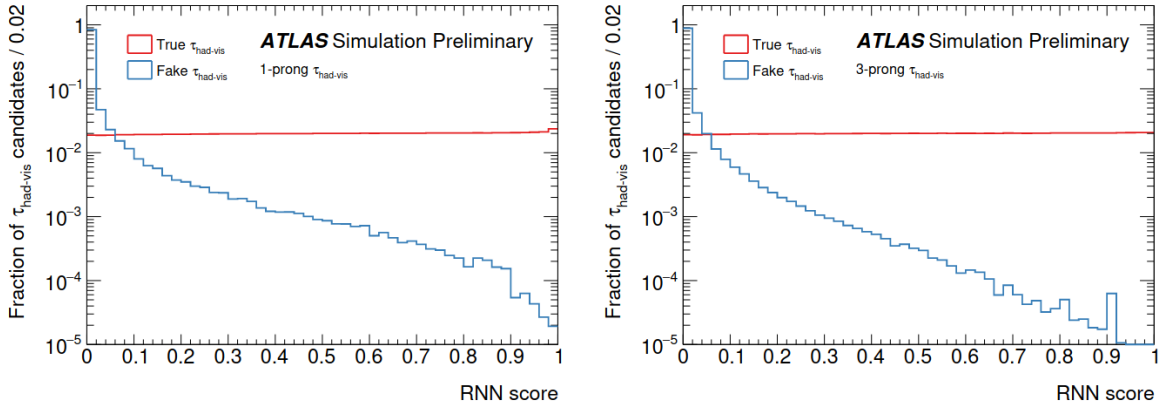


Figure 3.5: Tau RNN ID score curves for the (left) one prong tau deca and (right) three prong tau decay. [58]

constant due to a reweighting process performed with the objective of simplifying the calculation of the signal efficiency.

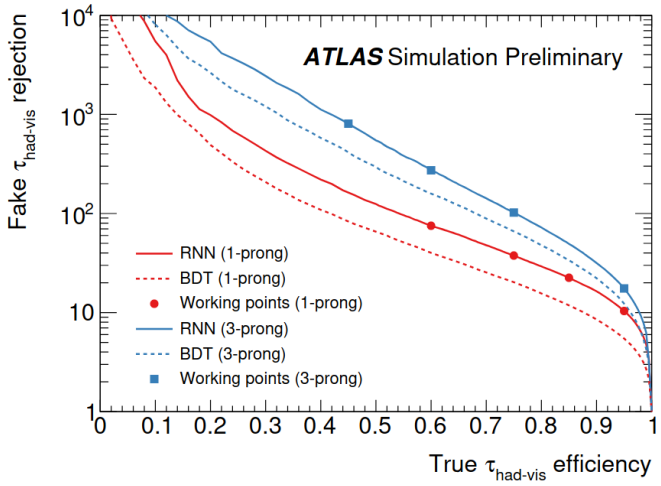


Figure 3.6: ROC curve comparing the RNN and BDT performance for one and three prong taus. [58]

Discrimination against electrons and muons

As electrons can be misidentified as hadronically decaying taus, particularly as one-prong taus, the RNN algorithm described previously may not have enough discrimination power between electrons and one-prong taus because it was trained only on di-jet events. Rejecting electrons as a background is especially important in the $e\tau_{had}$ channel due to the presence of the $Z \to ee$ process. To improve the rejection of electrons, a dedicated BDT is trained. Despite similarities in the signature between one-prong taus and electrons, there are discriminating quantities such as the radiation deposited when the particles traverse the material of the inner detector and the calorimeter shower shape, which can be used to differentiate between the two [43]. These quantities were used to train a BDT for the

purpose of discriminating between taus and electrons. A loose BDT working point is used in this analysis to reject electrons.

As the efficiencies of tau energy scale (TES) corrections and RNN ID can be different when applied in MC and data, these differences must be measured and correction factors called scale factors are measured in data to correct for the differences in efficiencies. An example of such scale factor for the eBDT tool where electron misidentification correction scale factors between data and simulation are measured in data is shown in Figure 3.7 where the scale factor is measured as a function of pseudo rapidity. These scale factors must be applied on MC in order to achieve a good closure between MC and data.

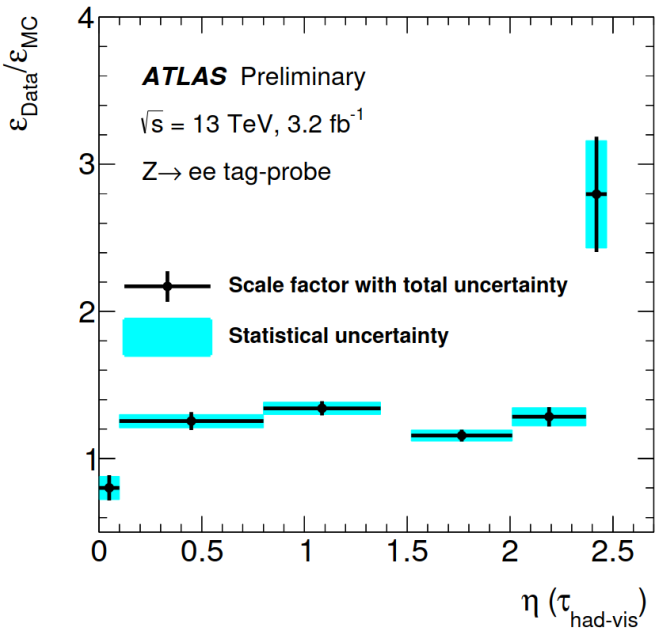


Figure 3.7: Measured tau scale factor as a function of pseudo-rapidity. [43]

In addition to the RNN ID working point, the tau leptons must also have $|\eta| < 2.5$ (excluding the crack region) and have a p_T greater than 35 GeV after the TES correction in order to be consistent with the trigger thresholds.

3.2 b-tagging

The associated production of tau pairs with jets is interesting for searches of BSM models. Leptoquark non-resonant production has larger production cross-sections in the b-jet regions, where the tau pair is produced alongside a b-jet [19], and hence the biggest expected deviation from the SM cross-section is in this region.

Proper identification of jets, and further classification of such jets into b,c or light jets is of utmost importance for leptoquark searches. The algorithms that perform such jet classification tasks are called flavor tagging algorithms. These algorithms exploit the properties of the different jet flavors, such as the long lifetime, high mass and high multiplicity of the different flavored jets. For example, b jets have a significant mean flight length, which generally leads to a displaced vertex and a high decay multiplicity [60].

In this analysis, the ATLAS DL1R b-tagging algorithm [60] is used for the jet flavor tagging. This algorithm is a combination of low and high level algorithms. Low-level algorithms operate on individual jets and output a number of desired variables that can be used to classify the jet flavor. The outputs of the low-level algorithms are then combined as inputs to be used in a fully-connected multi-layer feed-forward neural network (NN), forming the so-called DL1 algorithm series [60]. The DL1R tagger operating point was chosen as the working point which has a signal efficiency of 77%, meaning that 77% of the true b-jets are kept with that cut.

The algorithm performance is characterized by the probability of correctly tagging a signal jet. The background rejection as a function of the b-tagging efficiency can be seen in Figure 3.8. The first panel shows the background rejection as a function of the ϵ_b for three different b-taggers, and the lower two panels show the efficiency of the DL1 and the DL1R algorithms compared to the MV2c10 algorithm (older ID algorithm).

3.2.1 b-jet region

The definition of a b-jet region is crucial because models such as leptoquarks are anticipated to exhibit an elevated production cross-section when the tau pair is co-produced with a b-jet [19]. Additionally, the production of Drell-Yan processes reduces with the number of b-jets, making the regions where the tau pair is accompanied by b-jets the most susceptible to the largest deviations from the Standard Model [18]. Thus, we may establish a b-jet region as a specific region in which tau pairs are exclusively produced along with at least one b-jet candidate. Understating how signal and backgrounds behave in this region is important for the leptoquark searches.

Furthermore, the production of tau pairs in conjunction with jets alters the kinematics of the event. Therefore, it is imperative to verify whether the methods retain their

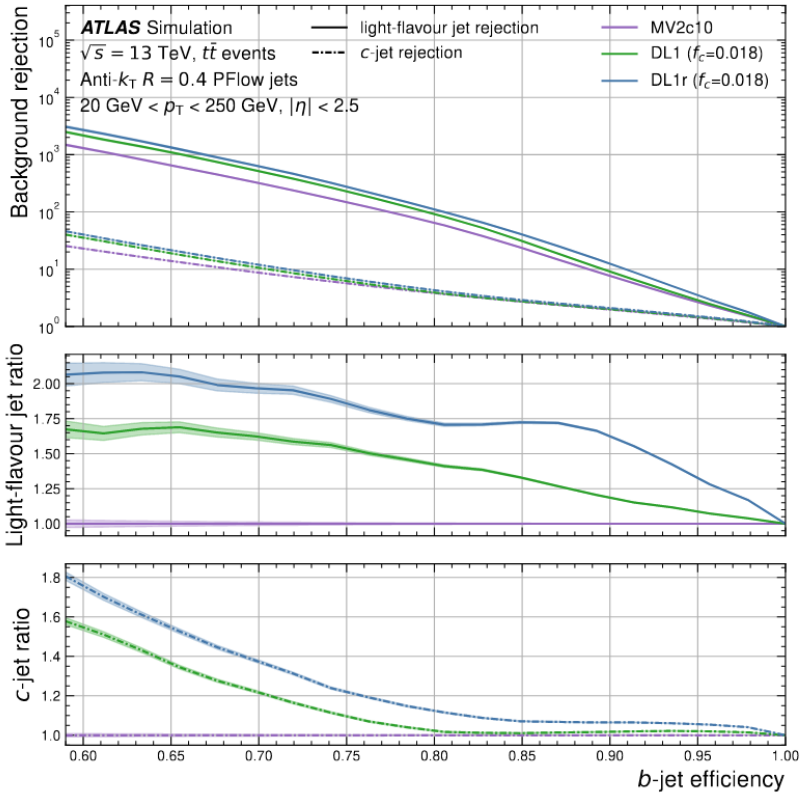


Figure 3.8: Light and c-jet rejections as a function of the b-tagging efficiency for a variety of b-taggers. [60]

performance in reconstructing the mass in this regime.

3.3 Electrons and muons

The semi-leptonic channel is composed of decays where one of the tau leptons decays leptonically, into a light lepton and its neutrino. Thus, identifying such particles is important for the analysis. Electrons typically lose a significant amount of energy due to bremsstrahlung when interacting with the material they traverse. The radiated photons usually decay into electron-positron pairs, which are typically emitted in a collimated way and are normally reconstructed as part of the same electromagnetic cluster. Muons, on the other hand, have a low energy deposited into the calorimeters. They are reconstructed by matching Muon spectrometer (MS) tracks to the Inner detector (ID) tracks and taking into account the energy loss in the calorimeter. After reconstruction, muons must also pass identification criteria based on the number of hits in the ID and MS.

Light leptons that originate from tau decays tend to carry a small fraction of the original tau momentum due to the production of neutrinos in the decay chain. Therefore, light lepton candidates are required to have a $p_T > 7$ GeV.

Electrons must fall within the regions of $0 < |\eta| < 1.37$ and $1.52 < |\eta| < 2.5$, which

are covered by the electromagnetic calorimeter and inner tracker, excluding the electromagnetic calorimeter crack ($1.37 < |\eta| < 1.52$). As the momentum of muons is measured by the muon spectrometer, they must have $|\eta| < 2.5$.

In order to optimize the efficiency of electron identification while controlling the background contamination, various electron isolation working points have been defined in the ATLAS experiment. The isolation criteria are based on the transverse energy sum of the calorimeter cells surrounding the electron or muon within a cone of a certain radius. For instance, for the FCHighPtCaloOnly working point, the energy sum within a cone of $\Delta R < 0.2$ around the electron is required to be less than $max(0.015 * p_T, 3.5GeV)$, while for the FCLoose working point, the energy sum must be lower than 0.2 GeV.[61, 62]. Both muon and electron candidates must pass the medium working point identification criteria, and the electron must pass the FCHighPtCaloOnly [62]isolation while muons must pass the FCLoose isolation working point[61].

3.4 Missing transverse energy

Momentum conservation in the transverse collision plane implies that the sum of the transverse momenta of all particles in a given final state should be zero [63]. A non-zero sum would indicate the presence of invisible particles in the decay chain. The negative sum of the momentum of all particles detected in a given event is called the missing transverse energy, or E_T^{miss}, and is a measurement of the energy imbalance in an event.

In a perfect detector, E_T^{miss} should be equal to the transverse component of the sum of momenta of the invisible particles in the event. However, due to detector finite resolution and imperfections, this may not be the case. Measurement errors of particle transverse momentum and non-detection of decay products can also cause a momentum imbalance, leading to a non-zero sum of transverse momenta in the event.

In this analysis, where signal events contain at least two neutrinos, E_T^{miss} is of fundamental importance, as the neutrinos in the final state will only manifest themselves in the form of E_T^{miss}. This is the only information available about the neutrinos, and a good estimation of this quantity is essential when trying to estimate the full tau pair invariant mass.

E_T^{miss} is reconstructed from the sum of the transverse energy of all detected particles in the final state. Its resolution depends on the resolution of the reconstructed and calibrated "hard objects" (electrons, muons, photons, hadronically decaying taus, and jets). Any other momentum flow that is not attributed to any of the hard objects is reconstructed as the "soft term".

The E_T^{miss} is defined as:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,\mu} + E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau_H} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,soft} \quad (3.1)$$

E_T^{miss} can be separated into the hard term and the soft term. The hard term is composed of muons, electrons, photons, taus and jets, each with their own dedicated reconstruction and calibration. Analysis-specific particle and jet selections must be taken into account in the E_T^{miss} calculation [63].

The soft term comprises all of the detector signals not matched to the reconstructed objects. To reconstruct the soft term, a track-based soft term (TST) is used. The TST includes tracks contributions that are not associated with the hard scatter objects.

In order to do this, the TST needs to provide excellent track-to-vertex matching. This is done by considering tracks based on their longitudinal and transverse impact parameters d_0 and z_0, defined relative to the hard scatter vertex position. d_0 is defined as the shortest distance in the transverse plane from a track to the beam line, and z_0 is the z-coordinate of the track point closest to the beam line.

The track selection criteria for the TST are:

- $p_T > 0.5$ GeV,
- $|\eta| < 2.5$,
- $|z_0 sin(\theta)| < 2.0$ mm

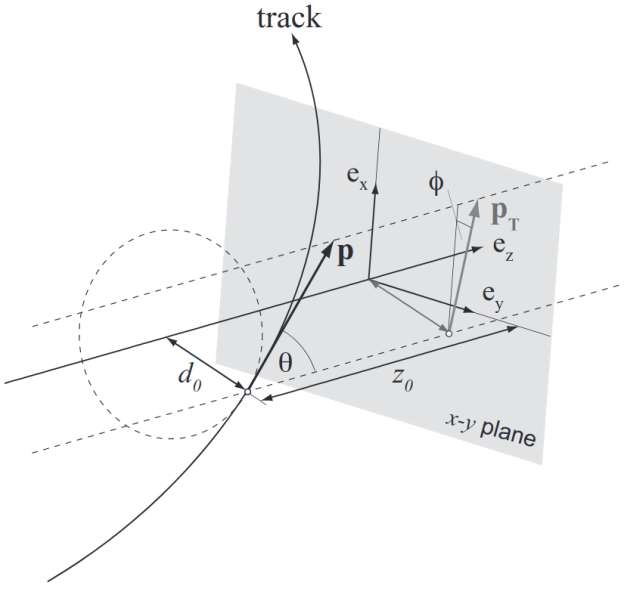


Figure 3.9: Illustration of the perigee parameters of a track. the point of closest approach is given by the signed transverse impact parameter d_0 and the longitudinal impact parameter z_0. [64]

Tracks that are not associated with any hard scatter objects and pass the above requirements are considered in the E_T^{miss} soft term. The sum of the tracks that passed

the TST selection and the hard objects in the event composes the final estimation of E_T^{miss}.

3.5 High mass region

As this analysis targets high mass measurements and searches, a high mass region is defined such as $m_{ll}^{vis} > 100$ GeV. This selection targets to remove the majority of the events from the large Drell-Yan productions at low invariant mass.

3.6 Efficiency

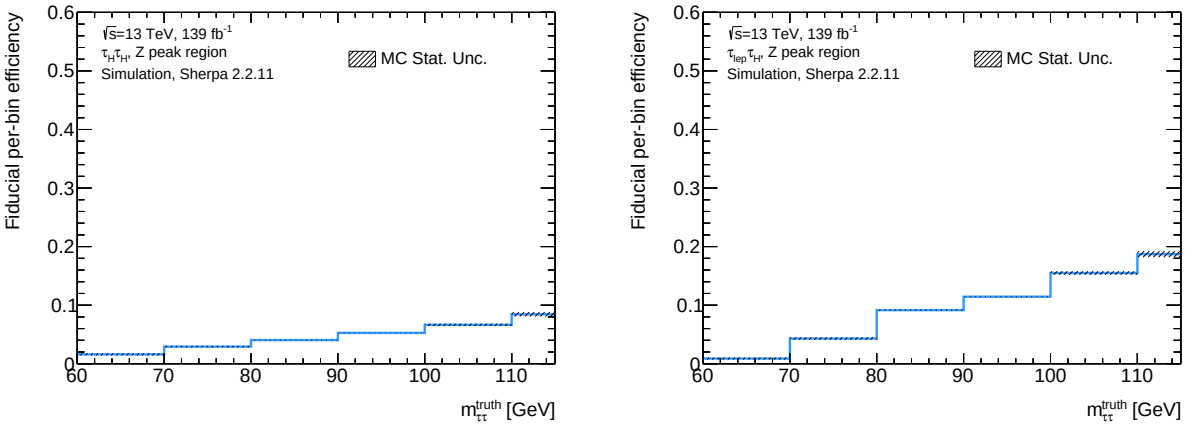
The selection process has a significant impact on the number of events that will be analyzed by the analysis. A set of strict or tight selection criteria will allow only a few events to pass through it, whereas more lenient or loose selection criteria will allow more signal events to be selected. In the context of signal detection, we can define a parameter called the selection efficiency or simply efficiency, which quantifies the ability of the current selection criteria to enable the selection of signal events.

In Monte Carlo simulations, the expected total number of events can be predicted, and hence, we can compare the total expected number of signal events to the number of signal events that passed through our selection. By calculating the efficiency in different bins, such as mass bins, we can determine the percentage of events that passed through the selection criteria for each bin. This is achieved by computing the number of events that passed both the ”truth” and ”reco” selection criteria and dividing it by the number of events that passed only the ”truth” selection criteria for each bin. This ratio represents the efficiency for that particular bin.

Figures 3.10 and 3.11 display the efficiencies per bin of true mass for the Z peak region and the high mass region, respectively. In the Z-peak efficiency plots, particularly in the $\tau_{had}\tau_{had}$ channel, the efficiency is observed to be extremely low. This can be attributed to the fact that the tau triggers have a higher p_T threshold, compared to the single muon and electron trigger. As lower invariant mass events produce lower p_T taus, the taus are not able to fire the triggers.

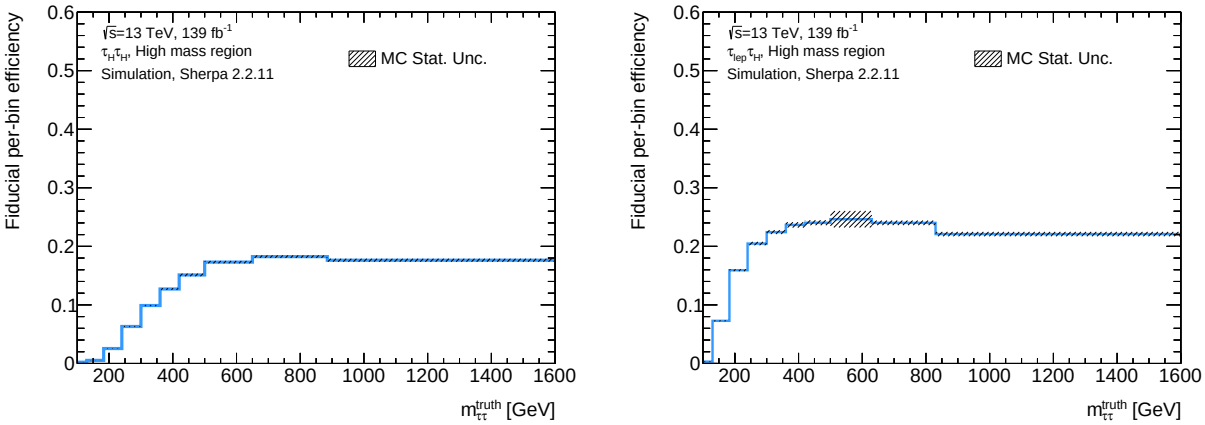
On the other hand, the efficiency in the high mass region, as illustrated in Figure 3.11, is higher due to the trigger being capable of selecting more events resulting from the high p_T production of taus in the high mass region. Once again, the low efficiency in the first bins of the plots is attributed to the cut in the visible mass, which is utilized to define the high mass region. The efficiency in the $\tau_{lep}\tau_{had}$ channel is higher in both Z peak and high mass regions due to the higher electron and muon reconstruction, identification, and

trigger efficiencies compared to those of tau leptons [65].



(a) (b)

Figure 3.10: Plots of the efficiency in the Z peak region for the (a) $\tau_{had}\tau_{had}$ and (b) $\tau_{lep}\tau_{had}$ channels.



(a) (b)

Figure 3.11: Plots of the efficiency in the high mass region for the (a) $\tau_{had}\tau_{had}$ and (b) $\tau_{lep}\tau_{had}$ channels.

3.7 Computation

This section provides a overview of the data generation and processing methodology employed in the ATLAS experiment data framework. The ATLAS workflow, as illustrated in Figure 3.12, is different for raw data and MC simulations that necessitate different handling procedures. These procedures are described in this section.

3.7.1 Simulation

In Monte Carlo simulations, the first step involves event generation, which involves simulating interactions between quarks and gluons produced by colliding protons. This process is followed by the simulation of parton showering, hadronization, and decay processes.

To effectively interpret the data obtained from the ATLAS experiment, it is crucial to have a comprehensive understanding of how the detector responds to various types of particles. To this end, the ATLAS simulation process is employed, which allows for the modeling of particle behavior within the detector and the prediction of expected detector responses.

In the case of real data, the output from the detector is in the form of raw data, which comprise electrical signals that require processing and analysis to obtain relevant insights about the collisions. In this analysis chain, the digitization process is a critical step that involves transforming the raw detector signals generated from the Geant simulation into digital data. This conversion is essential, as it allows for further processing and analysis of the electrical signals, effectively converting the simulated energy deposits into a detector response that resembles the raw data from the real detector.

Following the digitization process, the resulting digital data is processed through a series of algorithms to reconstruct the energy, position, and identity of the particles that produced the signals. This reconstruction process involves accounting for various detector effects, such as energy loss and resolution, as well as applying pile-up reweighting techniques. Additionally, the trajectories of charged particles are reconstructed through the use of tracking algorithms.

The algorithms used for particle reconstruction apply various corrections to the digitized data to account for the non-linear response of the detector and other detector-specific effects. For example, corrections may be applied to account for the dependence of the energy response on the position and angle of the particle relative to the detector, as well as the effects of detector material on the energy measurement. Similarly, pile-up reweighting techniques are applied to account for the additional particle interactions that occur in the same bunch crossing as the primary collision.

In summary, once the digitization process is complete, the reconstructed events follow

the same analysis chain as the real data. As a result, they are subject to the same physics analysis procedures that enable the extraction of significant information regarding the underlying particle interactions.

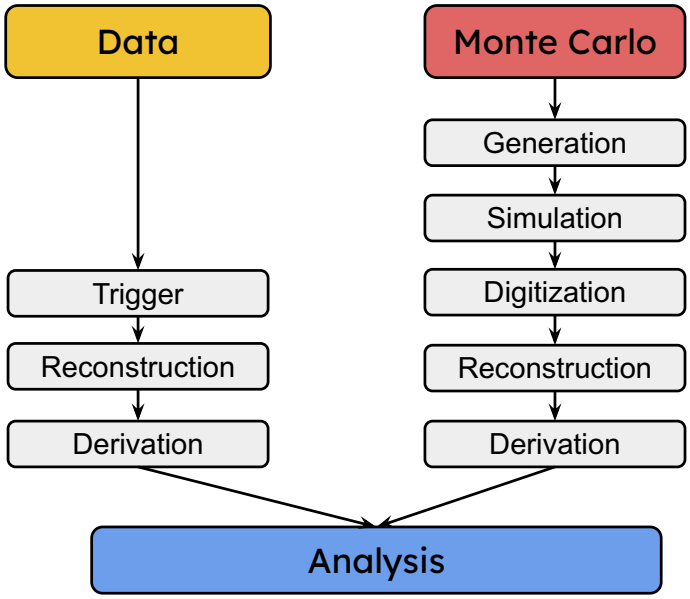


Figure 3.12: The ATLAS data processing chain. Based on [66].

Derivation

Reconstructed real and simulated data are stored in Analysis Object Data (AOD) files. These files are too large for individual analyzers to process independently. In order to enable timely analysis of the ATLAS data and reduce the amount of CPU resources utilized, smaller formats, known as Derived Analysis Object Data (DAOD) files, are produced centrally [67].

The derivation process plays a crucial role in the data analysis pipeline of the ATLAS experiment, where it involves the selection and preparation of data samples for further analysis. This process generally entails applying a set of selection criteria to the raw data, thereby identifying events that are pertinent to a specific physics analysis. Subsequently, the selected events are stored in a more compressed format referred to as a derived dataset, which is easily accessible and amenable to analysis by the researchers conducting the analysis.

The derivation process commences with the acquisition of raw data from the ATLAS detector. The raw data is subjected to the detectors trigger system, which identifies a subset of events that are relevant for subsequent analysis. Subsequently, the events that

pass through the trigger system are saved on a disk for later processing.

Following the identification of the relevant events, they are archived in a derived dataset that contains only the pertinent information for the particular physics analysis. The derived dataset is generally stored in a more compressed format than the raw data, and comprises reduced reconstruction data that is customized for specific analysis and performance groups. This approach simplifies data accessibility and analysis for researchers, as they do not have to process the entire raw dataset, leading to a reduction in the overall CPU usage by the collaboration.

3.7.2 Analysis framework

To effectively treat data and utilize the necessary tools for identifying and calibrating physics objects, an analysis framework is essential. The framework enables the centralized use of recommended tools, including pile-up reweighing, RNN ID, TES, and b-tagging, in physics analysis.

The DTA (Di-Tau Analysis framework) is the high level code used by the analysis team, and is actively being developed and validated by the analysis group[1]. This framework is based on the AthAnalysisBase software package used by the ATLAS experiment for data analysis. AthAnalysisBase is based on the ATLAS software framework Athena and provides physicists with a set of tools and templates to perform analyses consistently and compatibly with the ATLAS software framework. Its purpose is to facilitate researchers in writing their analysis code and to ensure the consistency and compatibility of all analyses. It is designed to work seamlessly with the ATLAS grid computing infrastructure, enabling researchers to conduct their analyses on distributed computing resources.

Table 3.2 summarizes the selections, and table 3.3 summarizes the MC samples and its generators that will be used throughout this work. Having all the triggers, reconstruction, selection and processing framework setup we can apply all of this infrastructure in the signal, background and data to produce the desired results for the analysis.

[1]Including this student

Table 3.2: Summary of detector-level cuts used throughout this analysis.

Cut	definition
di-lepton	at least two reconstructed leptons are required in all selected events.
Identification WP	all hadronic decay tau decay candidates need to pass the Tight RNN ID WP.
lepton flavors	$\tau_{had}\tau_{had}$, $\tau_{lep}\tau_{had}$ regions are defined based on the flavors of the reconstructed leptons in an event
trigger	both data and MC detector level events must fire at least one trigger in order to be selected
high-mass	the visible di-lepton invariant mass must be $m_{\tau\tau}^{Vis} > 100GeV$
lepton signs	two regions are used: the opposite-sign leptons signal region and the same-sign leptons validation region.
b-jet multiplicity	events are divided into 0 and at least 1 b-jet regions based on the number of b tagged jets.
$\tau_{had}\tau_{had}$ p_T cuts	leading tau $p_T > 35$ GeV and subleading tau $p_T > 25$ GeV.

Table 3.3: Summary of monte carlo samples and its generators.

Sample	Generator
$Z \to \tau\tau$	Sherpa 2.2.11
$Z \to \tau\tau$ (alternative samples)	Powheg+Pythia 8
$Z \to \ell\ell$ ($\ell = e, \mu$)	Sherpa 2.2.11
$W \to \ell\nu_\ell$ ($\ell = e, \mu, \tau$)	Sherpa 2.2.11
$t\bar{t}$	Powheg+ Pythia 8
Single top	Powheg + Pythia 8

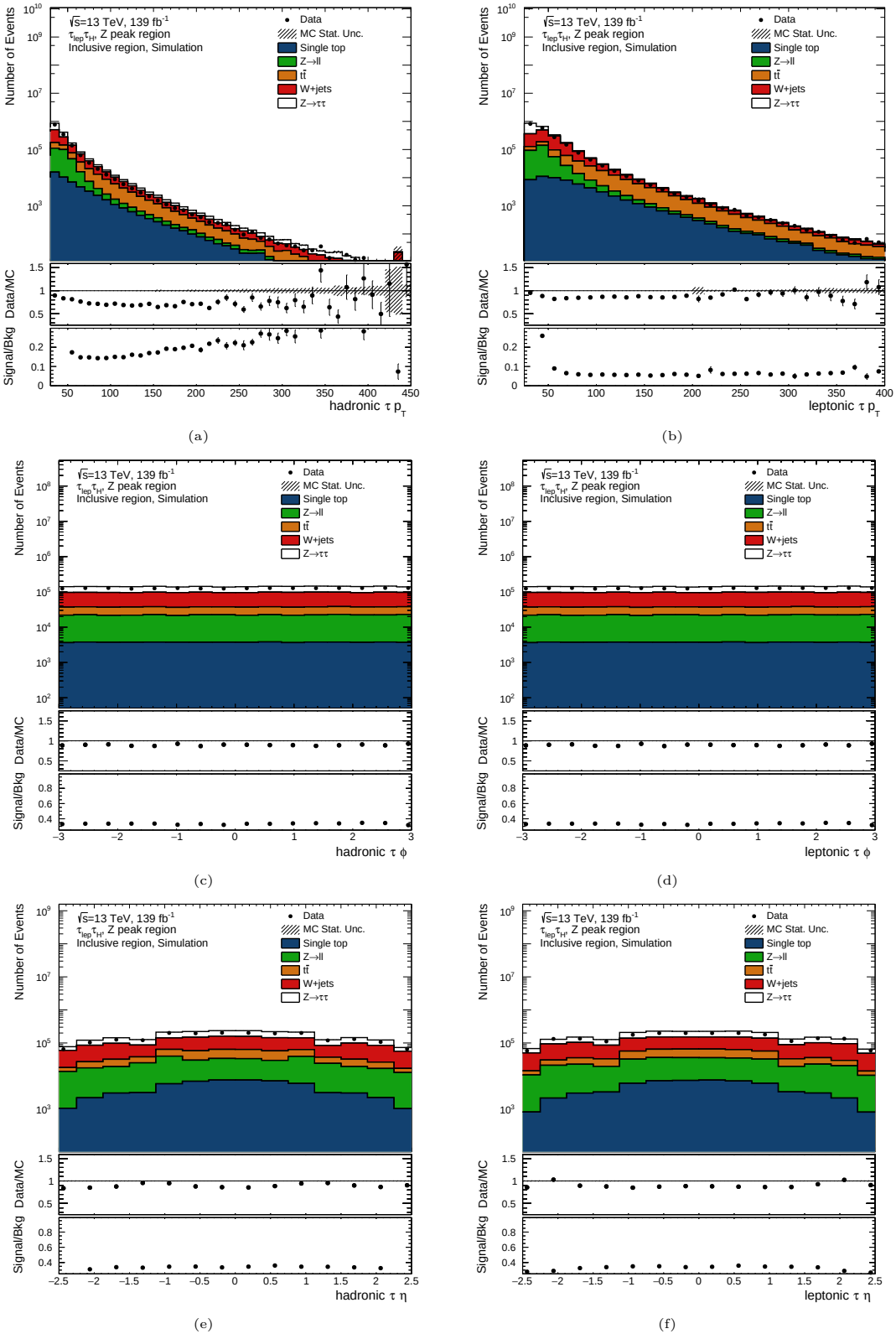
3.8 Kinematics distributions

By utilizing the selection criteria and infrastructure outlined in this chapter, we can now process both the data and Monte Carlo simulations to generate and analyze the distributions of interest. The kinematic plots of the hadronic and leptonic decaying taus in the $\tau_{lep}\tau_{had}$ channel can be found in Figure 3.13. Similarly, Figure 3.14 illustrates the same distributions for the leading and subleading taus in the $\tau_{had}\tau_{had}$ channels.

The p_T plots exhibit interesting behavior, particularly in the $\tau_{had}\tau_{had}$ channel where three peaks are observed. The peaks in the leading tau p_T distributions are attributed to the 35, 80, and 160 GeV p_T thresholds of the trigger used. The p_T distributions in the $\tau_{lep}\tau_{had}$ channel exhibit a steep decline due to the low threshold of the single muon and electron triggers. In the $\tau_{had}\tau_{had}$ channel, a discrepancy between the data and Monte Carlo simulations arises because the QCD multi-jet background has not been estimated, as this kind of background (fakes) is usually estimated from data-driven estimations due to the incapacity of the MC methods to correctly model those backgrounds [68]. No data driven background estimation are used in this work, only MC estimation of backgrounds.

The ϕ distributions exhibit a uniform distribution, as expected, since there is no reason for the distribution to be biased on the azimuthal angle. In the pseudo-rapidity distributions, a lower number of events occur in the region $1.37 < |\eta| < 1.52$ due to the presence of a calorimeter crack, that region which should not have any events, is populated since the bin edges do not coincide with the calorimeter crack region, there are

some events in that region



(a) (b) (c) (d) (e) (f)

Figure 3.13: The p_T distributions of (a) hadronic and (b) leptonic tau decays in the $\tau_{lep}\tau_{had}$ channel are presented. Plots (c) and (d) display the azimuthal angle distribution of the hadronic and leptonic tau decays, respectively. Additionally, plots (e) and (f) depict the distributions of the pseudo-rapidities for both tau decays.

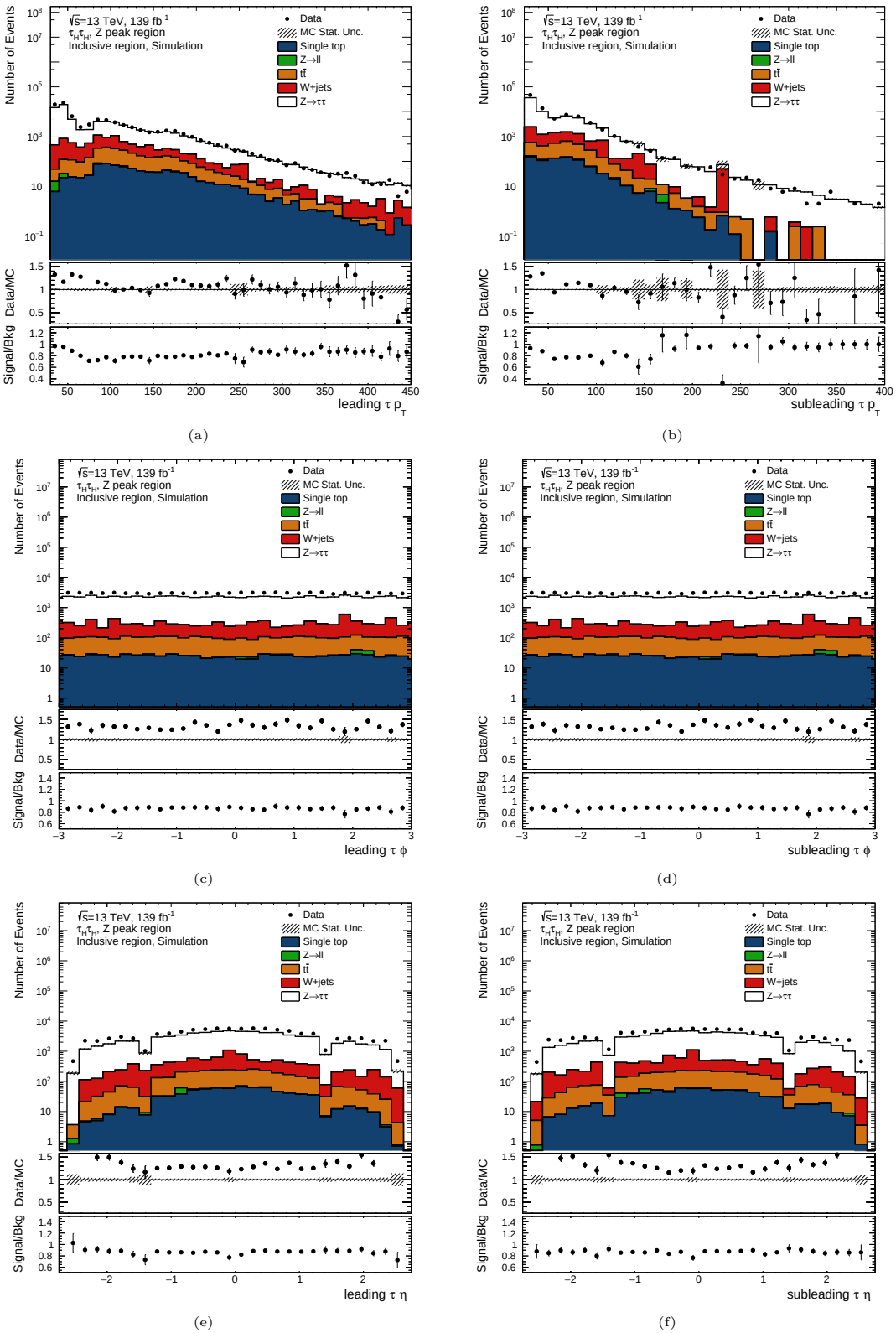


Figure 3.14: The p_T distributions of (a) leading and (b) subleading tau decays in the $\tau_{had}\tau_{had}$ channel are presented. Plots (c) and (d) display the azimuthal angle distribution of the leading and subleading tau decays, respectively. Additionally, plots (e) and (f) depict the distributions of the pseudo-rapidities for both tau decays

CHAPTER 4

MASS ESTIMATION METHODS

As briefly motivated in the introduction, the goal of this work is to study several mass estimation methods performance in the Z peak and in the High mass region. As one of the objectives of the analysis is to unfold the mass distributions at the high mass region and to perform searches for BSM models, such as leptoquarks, mass estimation methods have a central role, as mass estimation methods can increase the searches sensitivity and reduce bin migration, a behavior desired for the unfolding method. Theorically, we can reconstruct the mass of a given event using four momentum algebra:

$$m_{\tau\tau}^2 = (p_{\tau_1} + p_{\tau_2})^2 = (p_{\tau_1^{vis}} + p_{\tau_1^{\nu}} + p_{\tau_2^{vis}} + p_{\tau_2^{\nu}})^2 \tag{4.1}$$

As the neutrinos do not interact with the ATLAS detector it is not possible to measure them and hence it is not possible to analytically calculate $m_{\tau\tau}$. Alternative methods to estimate the invariant mass of the system must be pursued. Some of these methods are:

- The visible mass [32]
- Collinear method [33]
- Missing mass calculator [32]
- The $\hat{s}$ minimum method [34]
- Total transverse mass [18]

The methods are introduced in this chapter, and their performance will be evaluated in the next chapter.

4.1 The visible mass

The easiest approach to estimate the invariant mass of the tau pair system is to neglect the neutrinos information and to calculate the invariant mass only of the visible decays of the tau lepton, as:

$$m_{\tau\tau}^{vis} = (p_{\tau_1^{Vis}} + p_{\tau_2^{Vis}})^2 \tag{4.2}$$

This estimate leads to mass reconstruction with lower values than the generated value of the mass, due to the missing neutrinos information. The mass distributions of the $m_{\tau\tau}^{vis}$ at detector level, with the selections described in chapter 3 are shown in Figure 4.1.

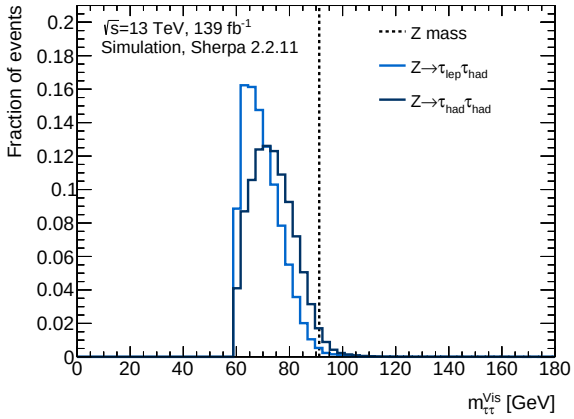


Figure 4.1: Distribution of the visible tau pair mass in the Z peak region in the dark blue $\tau_{had}\tau_{had}$ and light blue $\tau_{lep}\tau_{had}$ channel. The Z boson mass values was taken from the PDG (91.1876 GeV) [3] .

As the $m_{\tau\tau}^{vis}$ uses only visible momentum information of the tau leptons, the visible tau momentum resolution defines the $m_{\tau\tau}^{vis}$ resolution. As momentum and angular resolution of light leptons are better than for hadronic decaying taus, the resolution of the distribution in the $\tau_{lep}\tau_{had}$ channel is better. And as there is more lost information in the $\tau_{lep}\tau_{had}$ decays, as three neutrinos are present in the decay chain, the $\tau_{lep}\tau_{had}$ distribution has it peak at a lower value. The low number of events at low masses is due to tau p_T cuts in order to be compatible with the trigger thresholds.

Due to the lower mass values and lack of neutrino information, the $m_{\tau\tau}^{vis}$ will not be considered for further results in this analysis.

4.2 The collinear method

The collinear approximation technique was the first technique to approach a full reconstruction of di-tau event mass [32]. Alongside the visible decay information, it also includes

the E_T^{miss} information. It was proposed as a method to reconstruct the mass of a Higgs boson produced alongside a high energetic jet, and the Higgs later decays into a pair of taus [33, 69].

The method assumes that the tau and its decay products are collinear, and that the direction of the neutrinos in the tau decays is the same as the visible tau decay direction. This assumption is based on the fact that the mother particle (Z or Higgs boson) has a much larger mass than the tau, making the tau decay products highly collimated. The method also assumes that all of the missing transverse energy in the event is only due to neutrinos. With these assumptions, a system of equations can be constructed and solved [24]:

$$\begin{cases} \not{E}_{Tx} = p_{mis1}\sin(\theta_{vis1})\cos(\phi_{vis1}) + p_{mis2}\sin(\theta_{vis2})\cos(\phi_{vis2}) \\ \not{E}_{Ty} = p_{mis1}\sin(\theta_{vis1})\sin(\phi_{vis1}) + p_{mis2}\sin(\theta_{vis2})\sin(\phi_{vis2}) \end{cases} \tag{4.3}$$

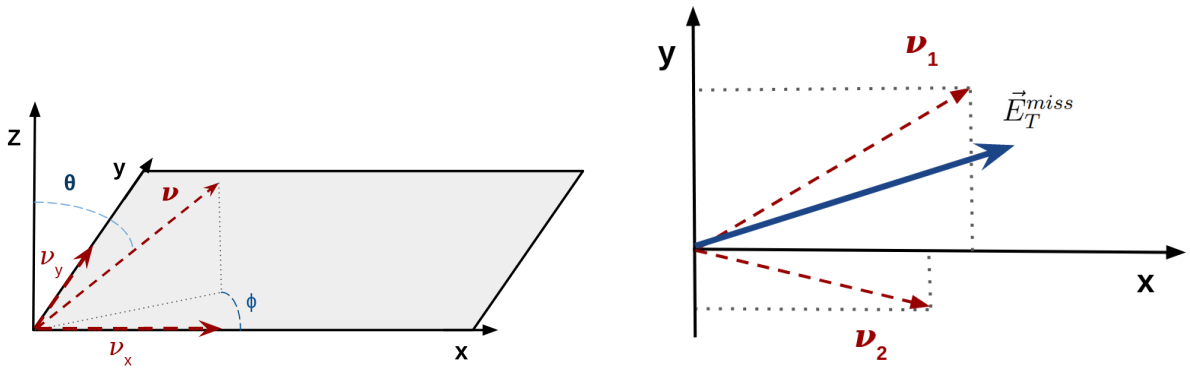


Figure 4.2: (a) projection of a 3d vector into the x-y plane (b) x-y neutrinos projections adding up to the $\vec{E}_T^{miss}$ vector.

In the above equations and in the following results, the two neutrinos in a leptonic tau decay will be considered as a single neutrino system, where the neutrino system is the sum of the two neutrinos momenta vectors. This approach has the advantage that no distinction between semi-leptonic and fully-hadronic expressions must be made. Additionally, if we treat the neutrinos from the leptonic decay separately, this would increase the dimensionality of an already under-constrained problem.

The set of equations (4.3) can be obtained by projecting the neutrino momenta of both taus onto the x and y axes, as shown in Figure 4.2. $\not{E}_{Tx}$ and $\not{E}_{Ty}$ are the components of the $\vec{E}_T^{miss}$ vector, p_{mis1} and p_{mis2} are the combined invisible momenta of the neutrinos from each tau decay, and the respective θ and ϕ are the combined neutrinos polar and azimuthal angles. These angles are assumed to be the same as the angles of the visible decay products, as stated in the earlier assumptions. In the system of equations 4.3, with the collinear method assumptions there are only two unknown quantities, the combined

neutrinos momenta.

While the collinear approximation technique has the advantage of fully reconstructing the di-tau mass, it also has several drawbacks. The assumptions made by the method are strong and not always accurate, as there can be multiple sources of missing transverse energy, such as momentum measurement errors. Additionally, the tau decay products are not always collinear. They tend to be collinear for high transverse momentum tau decays, but this may not be the case for all events in the Z peak region.

The collinear approximation technique also has the issue of diverging if the taus are produced back-to-back. The system of equations solutions diverges as $|\phi_{vis1} - \phi_{vis2}| \to \pi$, and as Z and Higgs decays tend to have a large separation between the taus, this causes the collinear method to fail in many of the events, resulting in a significant loss of statistical power in the analysis. This issue is reduced if one is solely interested in tau pairs produced with high p_T jets, where in such topologies the tau pair tends to have a smaller angular separation [32] .

The performance of the collinear method and all the methods that follows will studied in the next chater.

4.3 Missing Mass Calculator(MMC)

The missing mass calculator (MMC) is an advanced likelihood-based technique [24] devoloped by the CDF experiment at Tevatron [32] and was later adopted by the ATLAS collaboration. The MMC also allows for a full mass estimation and incorporates the information abou the $\vec{E}_T^{miss}$. But what makes the MMC method a robust method, is that it does not rely on the same assumption as the collinear method, and does not suffer from the same drawbacks. The method is able to reconstruct the di-tau mass on the resonance with a much better resolution than the collinear method and with much lower failure rate.

The MMC method is based on the solution of a extension of the system of equations used in the collinear method. The system of equations is presented below:

$$\begin{cases} \not{E}_{Tx} = p_{mis1} sin(\theta_{mis1}) cos(\phi_{mis1}) + p_{mis2} sin(\theta_{mis2}) cos(\phi_{mis2}) \\ \not{E}_{Ty} = p_{mis1} sin(\theta_{mis1}) sin(\phi_{mis1}) + p_{mis2} sin(\theta_{mis2}) sin(\phi_{mis2}) \\ m_{\tau 1}^2 = m_{mis1}^2 + m_{vis1}^2 + 2\sqrt{p_{vis1}^2 + m_{vis1}^2}\sqrt{p_{mis1}^2 + m_{mis1}^2} - 2p_{vis1}p_{mis1}cos(\Delta\theta_{\nu m_1}) \\ m_{\tau 2}^2 = m_{mis2}^2 + m_{vis2}^2 + 2\sqrt{p_{vis2}^2 + m_{vis2}^2}\sqrt{p_{mis2}^2 + m_{mis2}^2} - 2p_{vis2}p_{mis2}cos(\Delta\theta_{\nu m_2}) \end{cases} \tag{4.4}$$

This system of equations takes into account the projection of the neutrinos system onto the x and y axes as the collinear method, but also has two additional equations that

constrain the mass of the taus. These additional equations can be obtained by stating that the tau mass is equal to the sum of the visible and invisible tau decay product four momenta squared for each tau ($m_\tau^2 = (p_{vis} + p_{miss})^2$) .

Since the MMC method does not makes assumptions about the collinearity of the taus, the above system cannot be directly solved as there are more unknowns than equations. There are 8 unknowns in the MMC system (both neutrinos system three momentum and invariant masses) and only four equations. In the $\tau_{had}\tau_{had}$ channel, there are actually 6 unknowns because the invisible mass of the neutrino system is zero, as there is only one neutrino in the tau decay chain.

To be able to solve the system of equations, the method samples possible kinematical configurations of the event. For example, we know that the visible tau decay products and the neutrinos, although they are collimated they are not collinear and there is a angular separation between them. If we guess a value of the neutrinos azimutal direction ϕ we could in principle solve the MMC system of equations for the $\tau_{had}\tau_{had}$ channel, since we guessed the ϕ of both neutrinos there would be only four unknows, exactly the number of equations.

The MMC method is based on that, the method samples for possible configurations of neutrinos and also $\vec{E}_T^{miss}$, as the MMC method allows for possible mismeasurments in the $\vec{E}_T^{miss}$. The MMC samples points around the mean value of the measured $\vec{E}_T^{miss}$ while simultaneously sampling the neutrinos ϕ values and tries to solve the system of equations for each sampled point.

In order to solve the system of equations, a valid physical configuration must be sampled, which does not always happens, and sometimes the system cannot be solved for all sampled points. In the current configuration of the MMC method, used in the Higgs coupling measurement [24], the method samples 50k possible configurations and tries to solve the MMC system. The points where a solution was possible yield a mass estimation.

However not all sampled neutrino and E_T^{miss} configurations are equally likely, and hence theoretical knowledge of the tau decays is used. A set of probability density functions (PDFs) is derived in order to weight the different solutions and sampled points accordingly to their likelihood. An example of such PDF is in the Figure 4.3, where the probability of the angular distribution of the visible and invisible tau decay products is shown. The set of PDFs were derived using SHERPA $Z \rightarrow \tau\tau$ samples, under run 2 luminosity and center of mass energy conditions, and were derived by the ATLAS Higgs group.

Recapitulating, a possible kinematical configuration of sampled neutrinos ϕ angle and missing E_{T_x} and E_{T_y} are sampled and the system of equations is solved (if possible) for that point. If a solution is found, this point yields a mass estimation, and the likelihood

of this point based on detector measured and the sampled points is evaluated, and a histogram is filled with each mass estimation, weighted by its likelihood. This is done for all 50k points, and the final mass estimation is taken to be the point where the maximum of the histogram is located.

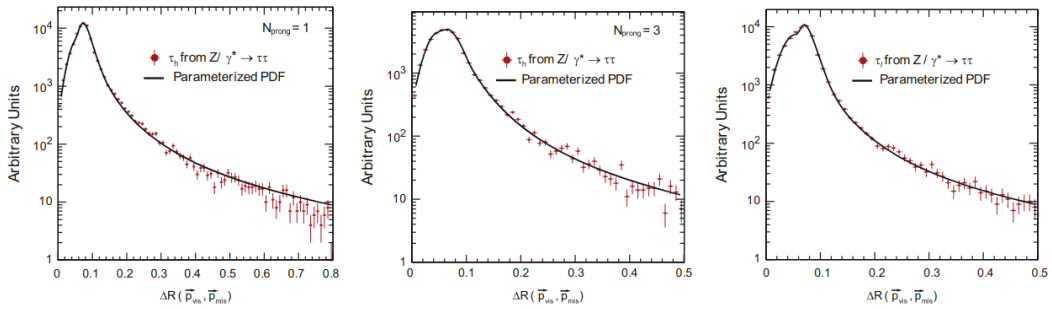


Figure 4.3: Probability distribution of the visible tau decay and tau neutrino ΔR for (left) one prong (middle) three prong (right) leptonic tau decays. [32]

For the $\tau_{lep}\tau_{had}$ (lep-lep) channel we have one (two) more unknowns, the mass of the invisible particles system. For that end another PDF is derived in order to sample possible masses, together with the missing E_T and neutrino angles.

The MMC method is capable of reconstructing the invariant mass of a di-tau pair around the Z peak with high performance and efficiency, as will be shown in the next chapter. This method does not result in mass tails and improves the separation of the Z and Higgs peaks. As a result, it decreases the Z background and enhances the sensitivity of Higgs measurements and searches [24].

4.4 The total transverse mass

The total transverse method (m_T^{total}) uses the information of the transverse mass of the event. The transverse mass of a two particle system is defined as:

$$m_T(p_T^i, p_T^j) = \sqrt{2p_T^i p_T^j (1 - cos\Delta\phi)} \tag{4.5}$$

The m_T^{total} mass is defined as:

$$m_T^{total} = \sqrt{m_T^2(\vec{p}_T^{\tau_1}, \vec{p}_T^{\tau_2}) + m_T^2(\vec{p}_T^{\tau_1}, \not{E}_T^{miss}) + m_T^2(\vec{p}_T^{\tau_2}, \not{E}_T^{miss})} \tag{4.6}$$

The method uses the transverse information of both taus and missing E_T, and has the property that it reconstructs QCD background with lower masses than the di-tau spectrum [20], giving a cleaner sample in the high mass region. This method was also recently used by the CMS collaboration for high mass Higgs and leptoquark searches [30].

4.5 $\sqrt{s}_{mim}$ method

The S_{mim} is a method that can estimate the invariant mass of the event and the event full kinematics. The method objective is to derive the minimal center of mass energy needed to reproduce the measured momenta and $\not{E}_T$ of an event. The authors were able to derive a simple analytical expression to the center of mass energy of the event visible quantities. We start from the expression of s for a system with n_{inv} invisible particles, such expression can be written as:

$$s = \left(E + \sum_{i=1}^{n_{inv}} \sqrt{m_i^2 + \vec{p}_i^2}\right)^2 - \left(\vec{P} + \sum_{i=1}^{n_{inv}} \vec{p}_i\right)^2 \tag{4.7}$$

The function s has a global minimal when considered as a function of the unknown variables $\vec{p}_i$ subject to the constrain $\sum \vec{p}_i = \not{P}_T$. Performing the minimization, one can obtain:

$$\vec{p}_{iT} = \frac{1}{2}\vec{P}_T \tag{4.8}$$

$$p_{iz} = \frac{1}{2}\frac{p_z}{\sqrt{E^2 - P_z^2}}\not{P}_T \tag{4.9}$$

The results above are already considering only SM neutrinos in the final state. The general expression for other kinds of invisible BSM particle can be found at[34]. The E and P_z in the above expression are the total energy and longitudinal momenta of the

event, which means that all of the neutrinos in the final state will have the same four momenta. This configuration in general does not represent the kinematics of the event, and can sometimes lead to poor performance. Substituting this expression into the s expression we can finally obtain:

$$\hat{s}_{min}^{1/2} = \sqrt{E^2 - p_Z^2} + \not{E}_T \tag{4.10}$$

Where E is to total measured energy of the event, the p_z is the total longitudinal momentum, $\not{E}_T$ is the total transverse missing energy. The quantities that will be used in the above expression are the reconstructed quantities at a detector level.

The method is simple and easy to apply, as it does have analytical equations and does not require any expensive computational operation. However the method presents some drawbacks. As the method estimates the minimum center of mass energy, and all of the visible quantities from the event must be considered, in regions where the ditau pair is produced alongside jets the method may overestimate the mass. Not only that the method is also sensible to initial state radiation and underlying events, which would increase the minimum center of mass estimation.

CHAPTER 5

PERFORMANCE OF MASS ESTIMATION METHODS

In this section, we will examine the performance of various methods for reconstructing the mass of the $Z \to \tau\tau$ process in the Z resonance region. SHERPA 2.2.11 samples are used as signal samples, and the trigger and selection criteria listed in Chapter 3 are applied.

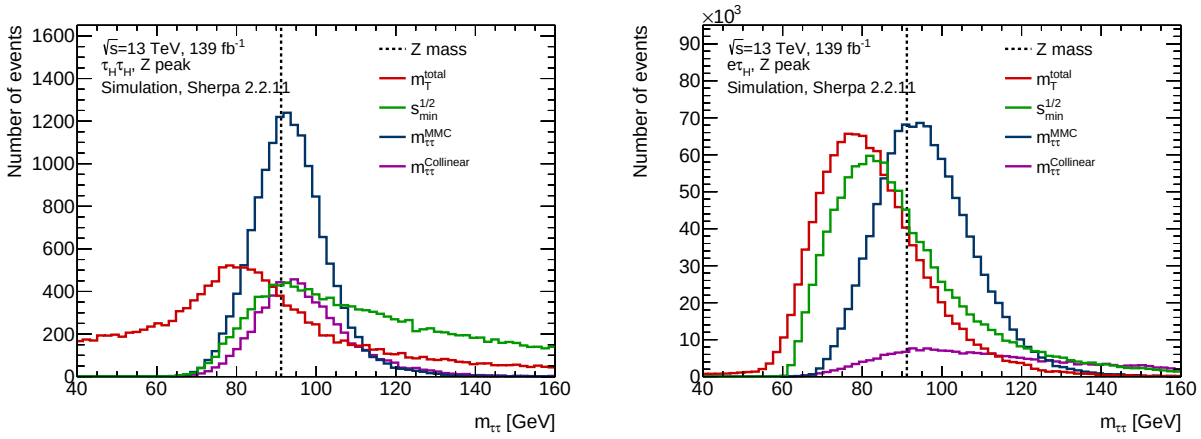
5.1 Methods performance in the Z peak

The Z peak region is extremely important as it allows us to study and measure several properties of the Z and Higgs bosons. A good mass estimation technique is crucial in order to study the Higgs boson, as both particles are are electrically neutral bosons with similar masses, their signature is very alike. In the $\tau\tau$ channel, the biggest discriminant between the Higgs and its much larger Z background is the invariant mass.

A good mass estimation method would be one that reconstructs the mass around the Z peak value and has the best resolution, in order to avoid larger tails that would bury any Higgs signal in this channel. Let's first analyze how the methods studied in the previous section behave in reconstructing events in the resonance. Figure 5.1 shows the different mass estimation methods' distributions at detector level in the resonance.

In order to generate the plots depicted in Figure 5.1, the mass estimation methods were integrated into the analysis framework. For each event that meets the current selection criteria outlined in Chapter 2, the MC detector level information is utilized to calculate the mass of the event using each of the available mass estimation methods. Subsequently, the resulting events are binned into a histogram, and the resulting distribution is displayed in the Figure.

The MMC method has the best performance in the resonance when compared to other estimation methods. It reconstructs the mass peak around the Z resonance value and its



(a) (b)

Figure 5.1: Mass distributions at detector level of the studied mass estimation methods in the Z peak region in the (a) $\tau_{had}\tau_{had}$ and (b) $\tau_{lep}\tau_{had}$ channels. The Z boson mass is taken from the PDG value [3]

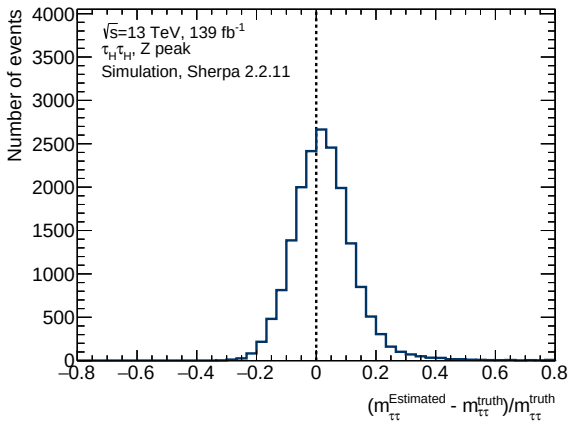
distribution has the lowest resolution of all the methods. We can quantify the bias and resolution of each method. To do this, let's define the relative mass difference per event as:

$$(m_{\tau\tau}^{Estimated} - m_{\tau\tau}^{truth})/m_{\tau\tau}^{truth} \tag{5.1}$$

Where $m_{\tau\tau}^{Estimated}$ is the mass estimated by any of the mass estimation methods at the detector level, and $m_{\tau\tau}^{truth}$ is the tau pair mass at particle level. We can define the relative bias and resolution with the relative mass difference per event. We can calculate the value for all events in a given sample and store them in a histogram, like the one in Figure 5.2. The mean of this histogram shows the systematic shift of the estimation method w.r.t truth masses at particle level, and it is defined as the relative bias and the RMS of such histogram indicates how spreaded are the mass estimations and hence, we can define the relative resolution based on that.

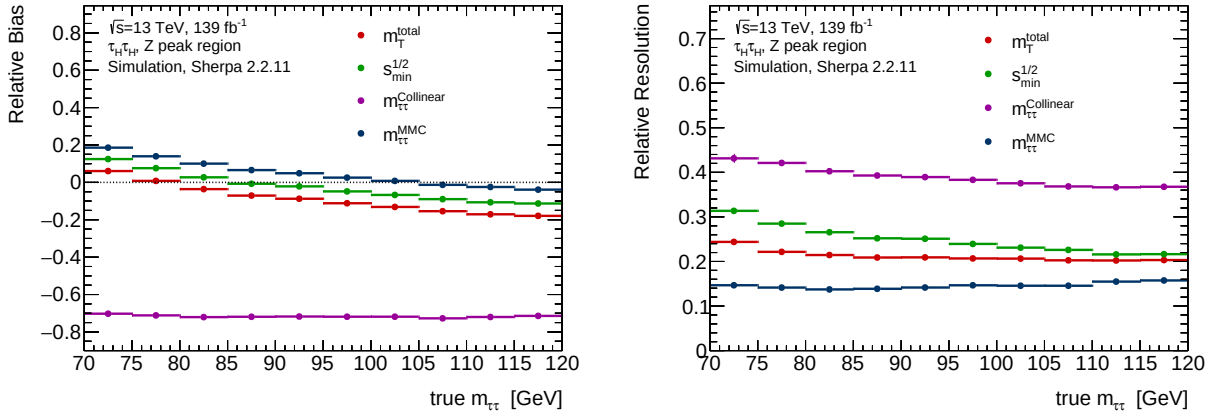
We can also study how the relative bias and resolution changes with some variable. In this case, particularly in the high mass range, we would like to understand how the bias and resolution evolve as a function of the mass at the particle level. We want to know the behavior of a method with events generated in a given mass bin, for example if the method tends to overestimate or underestimate the events generated in that region. Figure 5.3 shows the plot of the relative bias and resolution in given bins of truth mass at the particle level.

It is clear that both the $s_{min}^{1/2}$ and MMC methods have low bias around the resonance and both distributions have their peak close to the Z boson mass in the HH channel. However, the $s_{min}^{1/2}$ method has a worse resolution, which results in a long tail towards high masses. The total transverse mass also has a high resolution and as a result, it has



(a)

Figure 5.2: Example distribution of the relative mass difference per event for reconstructed MMC events in the Z peak.



(a) (b)

Figure 5.3: Relative bias (a) and resolution (b) around the Z boson resonance for various mass estimation methods.

large mass tails. The resolutions of the methods are also different in the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channel, particularly for the $s_{min}^{1/2}$ and m_T^{total} methods. This is also the case for the $m_{\tau\tau}^{Vis}$ mass. As the methods use the visible taus momenta and missing Et as inputs, their resolution depends on the resolution of these quantities, which is worse for hadronic tau decays than for leptonic decays, and as they also depend on the missing Et, the loss of resolution is bigger compared to the $m_{\tau\tau}^{Vis}$. Although the $\tau_{lep}\tau_{had}$ channel the methods have increase resolution, but now suffer from a worse bias, due to more neutrinos in the event.

The MMC method exhibits an inverse behavior in terms of its mass estimation, where it performs better in the $\tau_H\tau_H$ channel as compared to the leptonic-hadronic channel. This can be attributed to the way in which the MMC method accounts for missing energy and potential mismeasurements. The inferior resolution of the MMC method in the semi-leptonic case can be attributed to the weaker constraints present in the system of

equations and the requirement of sampling an additional variable, namely the neutrino system mass, as stated in [32].

Lastly, the performance of the collinear method is noteworthy, as it appears to have the poorest normalization compared to the other methods. However, it is able to accurately locate the peak of the distribution. This odd behavior is further highlighted when analyzing the method's resolution and relative bias. The bias plot does not align with this observation, as the method tends to overestimate the mass of the system, resulting in long tails in the distribution. The reasons for this peculiar behavior is the high failure rate of the method.

If we examine the survival probability of the methods, which is defined as the number of events that yielded a mass estimation divided by the total number of events in a given mass bin, we can quantify the proportion of events in which the method was able to reconstruct the mass of the event. Figure 5.4 illustrates the survival probability plot around the Z peak. It can be observed that both the Missing Mass Calculator (MMC) and collinear methods do not exhibit a 100% efficiency, far from that, the collinear method fails for approximately half of the events. On the other hand, the MMC method has a failure rate ranging from $1-3\%$ of events in which the system of equations cannot be solved [24]. This behavior sheds light on the collinear method resolution and bias behavior.

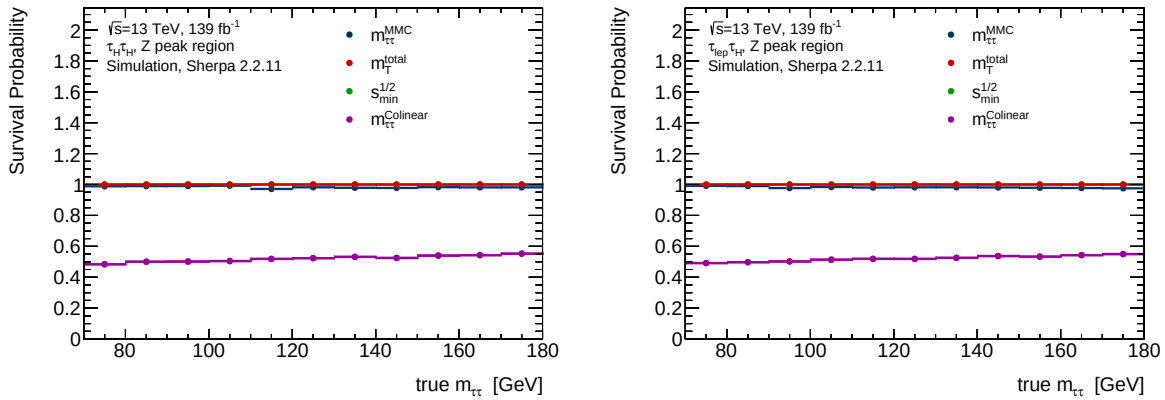


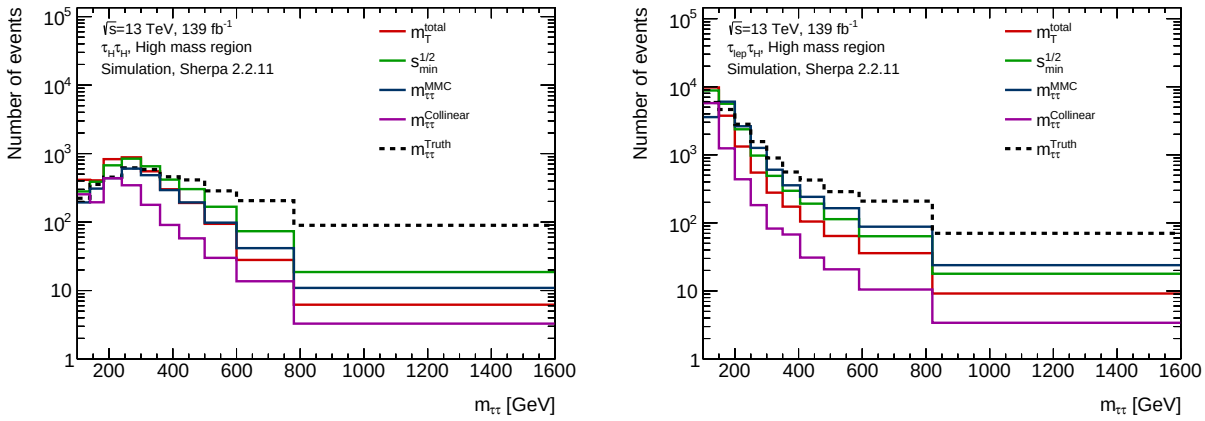
Figure 5.4: Survival probability of the studied mass estimation methods in the Z peak region for the (a)$\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels.

Based on the observations presented, the MMC method appears to be the most effective for reconstructing the mass in the Z peak region. It accurately predicts the Z boson mass and does not exhibit high mass tails towards the Higgs peak, making it an appropriate choice for estimating the Z boson mass in the resonance.

5.2 Mass estimation in the high mass region

We aim to analyze the performance of mass estimation methods in the high mass region, as defined in Chapter 3. This region comprises events with a visible di-tau mass, $m_{\tau\tau}^{vis}$, greater than 100 GeV. The cut in the visible mass implemented in order to exclude the significant presence of Z peak events and prevent contamination of low mass Drell-Yan events into the high mass region.

Lets then study the distributions of the methods presented in the past section. In this section, we will examine the distributions of these methods using SHERPA 2.2.11 samples and the selection and triggers outlined in Section 3. Figure 5.5 illustrates the distributions for the $\tau_{had}\tau_{had}$, $\tau_{lep}\tau_{had}$ channels. Additionally, the plots in this section also display the truth mass distribution at the particle level, allowing for a comparison of the shapes of the distributions in the particle and detector level. In the resonance plots presented earlier, the truth distribution was not included as the width of the distribution in the resonance is negligible in comparison to the experimental uncertainty. Therefore, only the position of the resonance was depicted.

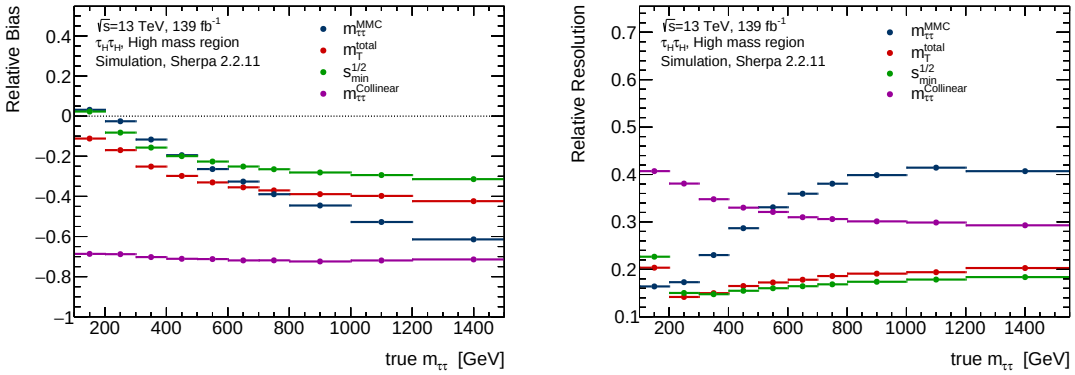


(a) (b)

Figure 5.5: Truth and detector level reconstructed tau pair mass by multiple methods in the (a) $\tau_H\tau_H$ and $\tau_{lep}\tau H$ channels.

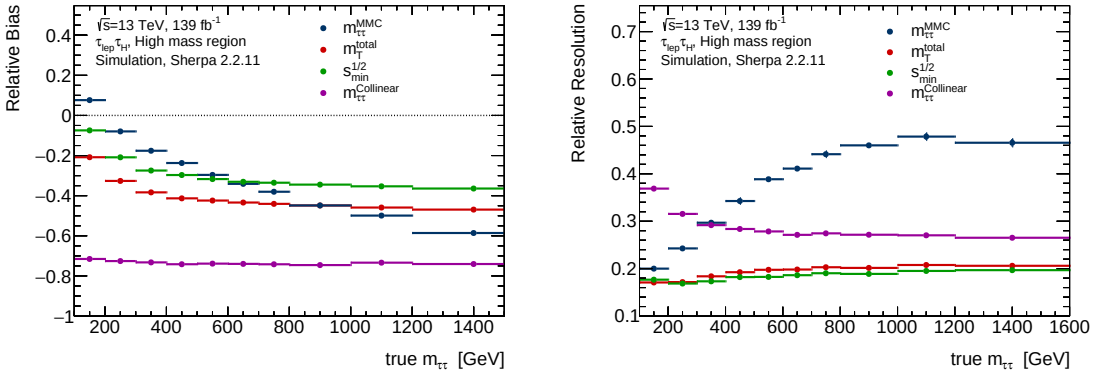
We have a commom behavior between all method in the high mass region. All of the methods have fewer events in the last bins of the distribution compared to the truth mass distribution. To study why this happens, lets first analyse the relative bias and resolution of the mass estimation methods. Figures 5.6 and 5.7 shows the relative bias and resolution of the methods in the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channel respectively.

With the exception of the collinear method, all of the methods exhibit a similar behavior in their bias distributions, characterized by a high negative bias towards the high mass region. This behavior can be explained by examining the relationship between the momentum of the tau leptons and the truth momentum at the particle level. Figure 5.8



(a) (b)

Figure 5.6: Relative bias (a) and resolution (b) in the $\tau_{had}\tau_{had}$ channel of several mass estimation methods in the high mass region in bins of the $m_{\tau\tau}^{truth}$.

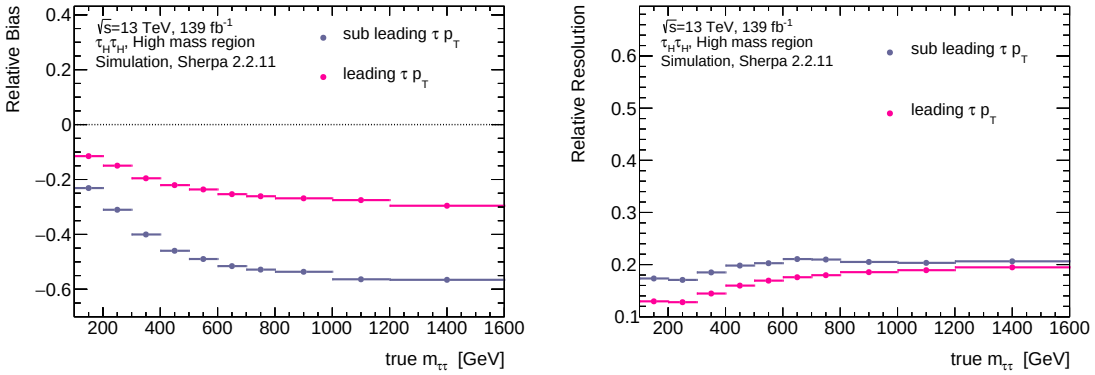


(a) (b)

Figure 5.7: Relative bias (a) and resolution (b) in the $\tau_{lep}\tau_{had}$ channel of several mass estimation methods in the high mass region in bins of the $m_{\tau\tau}^{truth}$.

illustrates the bias and resolution of the reconstructed visible tau p_T at the detector level and the particle level-generated total tau p_T. The plots depict the bias and resolution for both the leading and subleading taus. The results indicate that as we progress into the high mass region, the tau p_T at the detector level tends to be reconstructed with lower values.

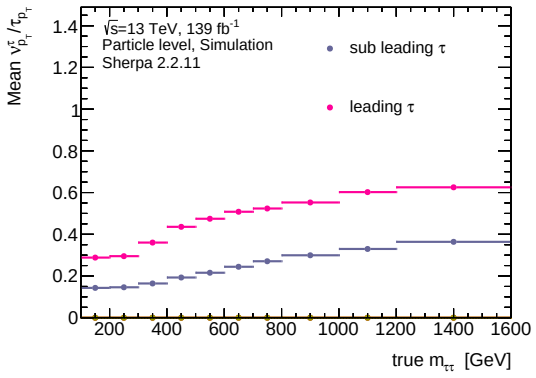
This occurs because the fraction of the momentum carried by the tau neutrinos is larger for taus generated with higher masses. To confirm this, a plot of the $\nu_{p_T}^{\tau}/\tau_{p_T}$, which is the ratio of the tau neutrino p_T divided by the total tau momentum, is presented at the particle level. Figure 5.9 illustrates the mean of the $\nu_{p_T}^{\tau}/\tau_{p_T}$ ratio for each mass bin in bins of the generated mass. It demonstrates that the tau neutrinos carry a larger fraction of the tau momentum for taus generated in the high mass region, resulting in more lost information in this region, which explains the high negative biases observed. It is worth noting that the energy scale corrections serve to adjust the tau momentum to the visible momentum, while the scale factors do not incorporate information regarding



(a) (b)

Figure 5.8: (a) Relative bias and (b) resolution of the leading and sub-leading reconstructed visible tau decays when compared to the generated tau momenta in bins of the truth tau pair mass.

the neutrinos.



(a)

Figure 5.9: Momentum fraction carried by the tau neutrino mean for the leading (pink) and subleading (grey) taus, in bins of bins of truth mass.

Although the reconstructed taus at detector level can explain the bias and resolution of the $s_{min}^{1/2}$ and m_T^{total} methods, the collinear and MMC methods exhibit an unusual behavior, with the MMC method showing an increasing resolution towards the high mass region. Although the MMC method performs well in the resonance region, it is currently only optimized for that region, as the PDFs used by the method were derived specifically for it. If we analyze the survival probability of the methods in the high mass region, defined as the number of events in which the method was able to yield a mass estimation divided by the total number of events in a mass bin, we will observe that both the MMC and collinear methods have a large number of failures in the high mass region, as shown in Figure 5.10. The Figure shows the survival probability of the different mass estimation methods in the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels.

The reason why the MMC method has a high failure rate in the high mass regions, is due to the fact the methods PDFs in which the neutrinos kinematics are sampled

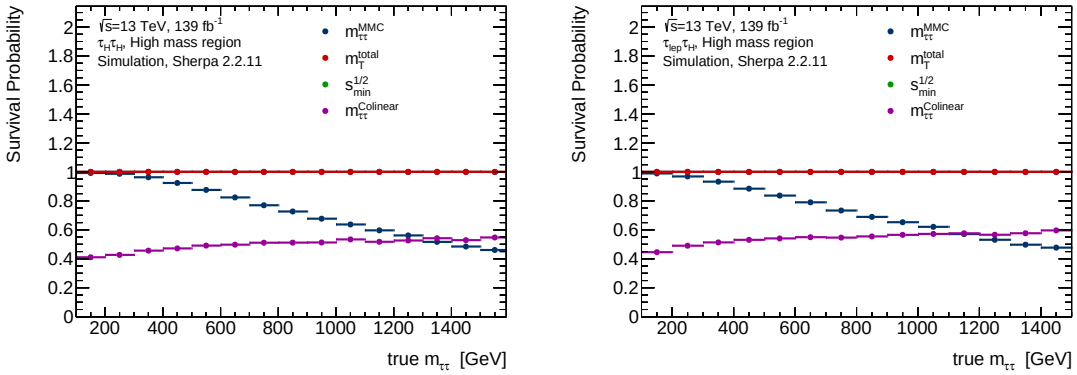


Figure 5.10: Survival probability of the studied mass estimation methods in the (a) $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels.

from are derived for the resonance region [24], and as in the high region the kinematical configurations of the tau decays are diferent, due to the high p_T of the taus, the system of equations sometimes cannot be solved for the sampled points from resonance derived PDFs. If the MMC system of equations does not have solution for any of the 50k sampled points, the method fails and does not return a mass estimation.

In principle, the MMC problem could be solved by the derivation of new sets of PDFs specifically tailored for the high mass region. However, given that this analysis aims to cover a large region of the phase space, not just the Z and Higgs peak regions, multiple sets of PDFs may be required, as the kinematics of the decay changes in different phase space region. This solution, however, would require a significant amount of time to produce the PDFs, conduct tests, and validate everything, and it does not fit within the current analysis and this work scope constraints. Thus, this solution was not pursued in this work.

None of the evaluated methods demonstrated an adequate performance in the high mass region. Although it is possible to modify the MMC method to enhance its efficacy in this region, the amount of work is far from trivial. This calls for a new approach to the problem, and it will be discussed in the next chapters.

CHAPTER 6

MACHINE LEARNING TECHNIQUES

Machine learning (ML) is a rapidly growing field that has revolutionized the way we process, analyze, and make decisions based on data. It allows computers to learn from data and make predictions, decisions, or take actions without being explicitly programmed. ML algorithms can identify patterns in data and use them to make predictions or take actions that would be difficult or impossible for a human to do.

Machine learning is today widely used for data analysis in the ATLAS collaboration. It can be applied to tasks like classification (b-tagging)[60], regression (energy scale corrections) [59], and clustering (calorimeter clustering) [70]. During the first run of data analysis, ML techniques such decision trees, boosted decision trees, and boosted decision forests were utilized for various purposes. However, during the second run, deep neural networks (DNNs) started replacing the decision trees in many tasks.

In the past, DNNs were complex and expensive to train, so other ML techniques like decision trees were used for many classification and regression tasks. However, since 2012, efficient training techniques for deep networks have become available, leading to a rapid increase in the use of DNNs in High Energy Physics (HEP). The widespread use of DNNs is due in part to the availability of python frameworks like TensorFlow, Keras, and PyTorch [71], which make it easier to train and employ DNN algorithms.

As the networks and training techniques have improved, DNNs have become more complex and efficient. They now perform better on more complex problems and have replaced other machine learning techniques in most applications. In this study, we explored the application of DNNs for mass estimation, using various neural network architectures for the task.

The aim of this chapter is to give a general understanding of the basic concepts of machine learning and introduce the neural network structures that will be utilized for

mass estimation.

6.1 A brief introduction to machine learning

The ML algorithms can be taught to carry out a specific task, in this case, the estimation of the particle level mass of an event based on the quantities measured at the detector level. The training and learning process is a crucial aspect of ML techniques, and the performance of the network will be determined by it. There are three distinct training approaches:

Supervised learning

Supervised learning trains a network by providing it with desired target outputs. In this work, the target output is the tau pair mass at the particle level. This training technique is used for both classification and regression tasks, where the network learns from labeled data to make predictions or classifications based on input patterns.

Unsupervised learning

In this training technique, there are no defined targets and the network attempts to discover hidden patterns in the training data on its own. This training method is often used in clustering applications, where the goal is to group similar data points together without prior knowledge of their grouping, like clustering.

Reinforcement learning

Reinforcement learning is a type of machine learning in which an agent learns to make decisions by performing actions within an environment to maximize a reward signal. The agent is provided with feedback in the form of rewards or penalties for each action taken, and it continually adjusts its strategy to maximize the overall reward.

The objective of this work is to regress the quantities measured at the detector level back to the undecayed (particle level) tau quantities. To achieve this, we will employ a supervised learning technique. In order to train the neural network for this regression task, MC samples will be utilized.

Supervised learning involves providing the network with an input vector, upon which it makes a prediction. The network is then evaluated based on the deviation between its prediction and the target values, with the weights of the network being adjusted accordingly to reduce this discrepancy.

The training samples used in this study are simulated using SHERPA 2.2.11, as previously described. The training samples and methodology will be further elaborated upon later in the chapter. Having established the training methodology and selected the training samples, the training of the neural network can commence. The network operates as a function, receiving an input vector $\vec{x}$ and transforming it through a combination of linear and non-linear operations. The objective of a neural network is to approximate a function $g(\vec{x};\theta)$ in order to model the relationship between the inputs $\vec{x}$ and the desired outputs $\mathbf{y}$, such that:

$$\mathbf{y} \approx \hat{y} = g(\mathbf{x};\theta) \tag{6.1}$$

The parameters θ represent the set of free parameters in a neural network, including its bias and weights, which will be trained to approximate $g(\mathbf{x};\theta)$, $\hat{y}$ is the prediction of the neural network and $\mathbf{y}$ is the target output vector.

The main feature of neural networks is their ability to learn and adapt. This is achieved by emulating the functioning of the human brain through the use of artificial neurons, which are inspired by the structure and behavior of biological neurons. Each artificial neuron, referred to as a neuron from this point forward, is associated with a weight that is adjusted during the network's training and a bias, which represents a constant offset. Neurons are the building blocks of neural networks and their behavior can be visualized in Figure 6.1.

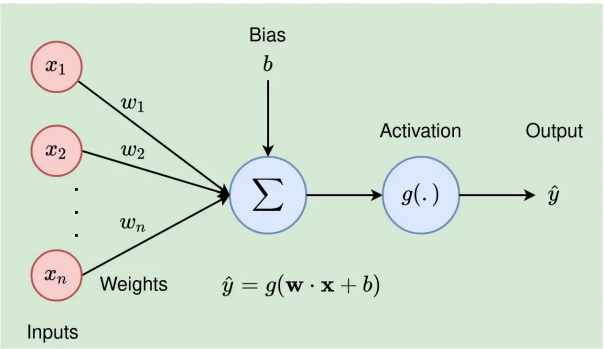


Figure 6.1: Illustration of how a artificial neuron works [72].

A single neuron can receive several inputs (x_i) and a bias. Each component of the input is connected to the neuron through the corresponding weights (w_i). The inputs (x_i) are multiplied by the respective weights (w_i), and the resulting products are summed. The bias is then added to the sum and the neuron is activated by an activation function. Analogous to the human brain, the activation function provides a rule that determines the extent to which and how the input information should be transmitted to subsequent neurons. The activation function will be discussed in more detail later in this section.

This process can be summarized by the following equation:

$$\hat{y} = g(\mathbf{w} \cdot \mathbf{x} + b) \tag{6.2}$$

Where $\mathbf{w}$ represents the weight vector that connects the inputs to the neuron, $\mathbf{x}$ is the input vector, and b is the bias. During network training, these weights connected to the neurons will be adjusted, in order to approximate the network outputs to the target values.

To build a neural network several neurons are needed. These neurons will be distributed across different layers, and the output of a neuron in a layer i will be the input of another neuron in a layer $(i+1)$. A deep neural network can be observed in Figure 6.2. A neural network is composed of an input layer, a hidden layer and an output layer. A hidden layer in a neural network refers to a layer of nodes or neurons between the input layer and the output layer. The data flowing through the hidden layers is transformed and processed in such a way that it is not directly observable, hence the term "hidden". A Deep Neural Network is when the network have more than one hidden layer. The network shown is also said to be fully connected, where each neuron is connected with each other.

Now we have multiple neurons connected to multiple inputs, in the Figure 6.2 network. The question is now, given an input $\mathbf{x}$ how can we make operation with this input until an output is produced by the network?

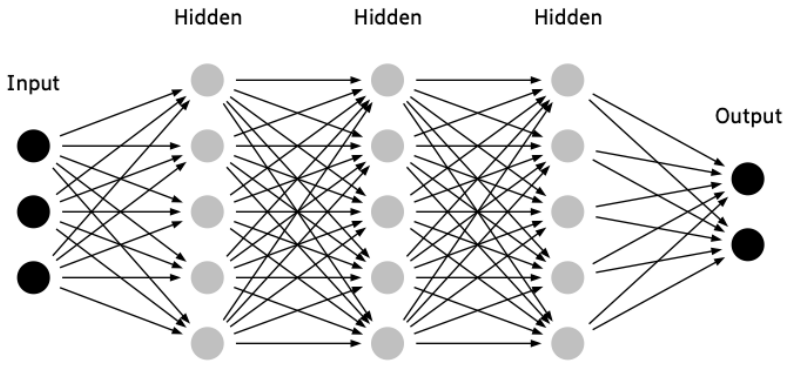


Figure 6.2: Representation of a deep neural network [73].

Now that we have multiple neurons connected to the inputs, the $\mathbf{w}$ is no longer a vector but a matrix. The term $w_{i,j}$ represents the matrix elements. Lets build the matrix $\mathbf{w}$ connecting the input layer to the first hidden layer of the network 6.2. The index i will denote the i the network input and the index j will denote the j the neuron of the first hidden layer. Then the matrix $\mathbf{w}$ connecting the inputs to the first hidden layers is:

$$\mathbf{w} = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ w_{2,1} & w_{2,2} & w_{2,3} \\ w_{3,1} & w_{3,2} & w_{3,3} \\ w_{4,1} & w_{4,2} & w_{4,3} \\ w_{5,1} & w_{5,2} & w_{5,3} \end{bmatrix} \tag{6.3}$$

With the weight matrix we can perform the multiplication operations and obtain the outputs of the neurons of the first layer of the network:

$$\hat{y}^{[1]} = g(\mathbf{w}^{[1]} \cdot \mathbf{x} + b^{[1]}) \tag{6.4}$$

where $\hat{y}^{[1]}$ stands for each neuron output of the first layer, which in this case has dimension 5, $\mathbf{w}^{[1]}$ stands for the weight matrix which in this case has dimension 5x3, $\mathbf{x}$ is the vector of inputs, whth dimension 3 and $b^{[1]}$ is the bias vector of that layer.

Next, we repeat this process using the outputs of the first layer as input to the second layer neurons. The input is fed to the second layer as follows:

$$\hat{y}^{[2]} = g(\mathbf{w}^{[2]} \cdot \hat{y}^{[1]} + b^{[2]}) \tag{6.5}$$

But now the $\mathbf{w}^{[2]}$ would have dimension 5x5. This operation can be done recursively until the final output, and a network result is achieved for that set of inputs. The act of feeding the network an input and work out the weights multiplications and non-linear operations is called forward propagation.

6.2 Network training

After conducting the forward propagation, the outputs of the network can be corrected towards the desired values using a loss function. This loss function serves to quantify the discrepancy between the network's prediction and the desired target. In other words, after feeding the network an input vector, it outputs an estimated value, $\hat{y}$, but the goal is for the network to output the true value, $\mathbf{y}$. This discrepancy can be quantified using a loss term as:

$$L = \ell(\mathbf{y}, \hat{y}) \tag{6.6}$$

ℓ represents a selected loss function. When training a neural network, it is common practice to divide the training data into smaller groups, known as $mini - batches$, as this makes the computation more efficient. For each $mini - batch$, the back-propagation

process is carried out to obtain the output for each element and calculate the loss for each. If the $mini - batch$ size is m, the loss function is calculated as follows:

$$L = \frac{1}{m} \sum_{i=0}^{i=m} \ell(\mathbf{y}_i, \hat{y}_i) \tag{6.7}$$

The larger the values of L, the greater the correction required in the network biases and weights. The goal of the network training is to minimize this loss or sometimes called $cost function$. Commonly used loss functions in the field include Mean Square Error (MSE), it is defined for a single event as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{6.8}$$

The MSE is a widely used loss function in regression problems as it is sensitive to large differences between the predicted values and the actual values. It scales the difference squared, emphasizing even more on predictions that are further from the target. It will be used in the DNN and DeepSets models training, which will be explained later in this chapter.

After the backpropagation, once we have established the discrepancy between the predicted and desired values through the loss function, we can proceed to train the network, based on the loss function.

Backpropagation and optimization

To train a neural network, it is necessary to update its weights so that the results minimize the cost function. The training process objective is to minimize the loss function, and the goal is to find the weights that minimize the error for a specific problem. To update the weights in a manner that reduces the error, it is necessary to calculate the gradients of the loss function with respect to each weight. This information can be used to determine how changes in each weight will impact the loss function.

One effective method for calculating these gradients is the backpropagation algorithm. The purpose of backpropagation is to adjust the weights and biases of the neural network based on the calculated loss. It works by propagating the error through the network, starting from the output layer and moving backwards through the hidden layers. As the error is propagated, the weights and biases of each neuron are updated.

To understand how backpropagation works, take two hidden layers of a neural network and suppose that the first layer contains an i-th neuron and the subsequent layer contains a j-th neuron. To calculate the derivative of the loss function (L) with respect to the weight $w_{i,j}$ that connects these two neurons, we use the following expression for the derivative:

$$\frac{\partial L}{\partial w_{i,j}} = \frac{\partial L}{\partial o_j}\frac{\partial o_j}{\partial z_j}\frac{\partial z_j}{\partial w_{i,j}} \tag{6.9}$$

Here, the chain rule of calculus is applied. As the loss function depends on the output o_j of the j-th node, which in turn depends on z_j, the chain rule can be used to obtain the expression on the right-hand side. The final two terms in the right-hand side are known to us and are expressed as follows:

$$\frac{\partial o_j}{\partial z_j} = f^{'}(z_j) \leftrightarrow \frac{\partial z_j}{\partial w_{i,j}} = o_i \tag{6.10}$$

Now the equation becomes:

$$\frac{\partial L}{\partial w_{i,j}} = \frac{\partial L}{\partial o_j} f^{'}(z_j) o_i \tag{6.11}$$

We now need to calculate the derivative of the loss function with respect to the output of neuron j, which can be a more complex task. This is because all of the neurons in the subsequent layer receive the output o_j of neuron j as input. Considering all neurons N in the next layer that receive the output o_j:

$$\frac{\partial L}{\partial o_j} = \sum_{\ell \epsilon N} \left(\frac{\partial L}{\partial o_l}\frac{\partial o_l}{\partial z_l}\frac{\partial z_l}{\partial o_j} \right) \tag{6.12}$$

Where ℓ is a neuron of the N neurons in the layer following neuron j. Equation (6.12) implies that in order to calculate the gradients of $w_{i,j}$, it is necessary to have knowledge of the gradients in subsequent layers. And in order to calculate the derivatives in the ℓ layer, we would need to know the derivatives in the layer after that, making this a recurrent calculation.

In order to be able to calculate all of the derivatives, one must first calculate the expressions in the output layer, and backpropagate it into the first layers of the network, hence the name "backpropagation". Backpropagation is an efficient method for calculating the gradients of the loss function with respect to the weights. Once these gradients are calculated, the network weights can be updated using various algorithms, such as gradient descent. The gradient descent method involves updating the weights as follows:

$$w_{i,j} = w_{i,j} - \alpha \frac{\partial L}{\partial w_{i,j}} \tag{6.13}$$

The term α is referred to as the learning rate and determines the magnitude of the steps taken. It is a hyperparameter of the neural network, meaning that it is not learned during the training process and must be chosen prior to training. The learning rate is a crucial factor, as a value that is too large can result in divergence, while a value that is

too small will result in an excessive number of steps. The goal is to minimize the loss function by iteratively adjusting the model parameters in the direction of the negative gradient of the loss with respect to the parameters.

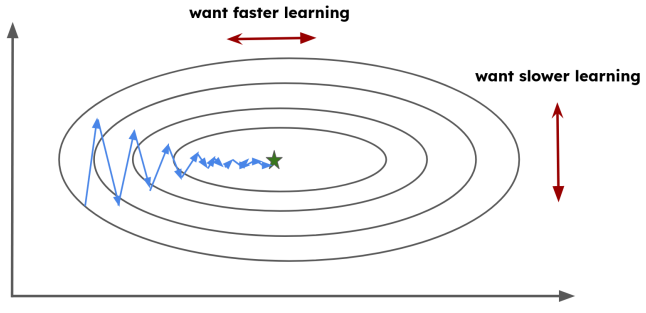
Optimizers

The gradient descent is the simplest and first used optimization method. It consists of a forward pass of the training data and calculation of the loss for these data. Then the loss value is back propagated and as the gradients are calculated the weights can be updated for the next forward propagation. The objective is to reach the minimum of the loss function. The gradient descent algorithm updates the weights as:

$$\theta = \theta - \alpha \frac{\partial L}{\partial \theta} \tag{6.14}$$

where θ is any weight of the neural network. But the method presents a lot of drawbacks that lead to the development of other optimization algorithms. The gradient descent has a problem of slow convergence. This can be illustrated in Figure 6.3. The Figure shows contour plots of a loss function that depends only on two parameters. The objective is to update the weights (x and y labels) to minimize the loss function and reach the minimum of the loss function, the green star at the center of the image. The optimizers objectives is to be able to reach the minimum, and do it so with the lowest number of iterations possible.

In Figure 6.3 the network ends up taking more steps than it should because the steps are too big vertically, where the minimum point is located horizontally direction. This leads to poor optimization and the necessity of using small learning rates in order to the algorithm to be able to converge.



(a)

Figure 6.3: The contour plot of a loss function that depends only of two parameters. The blue arrows shows the direction of the gradients and the green star shows the minimum of the function.

This issue can be addressed by the addition of momentum. By allowing the optimizer to take into account the past gradients when updating the parameters, rather than just

the current gradient, it can understand the the trend of the network is horizontal as the vertical oscillations cancel out and so the network could reach the minimum faster. The way it is done is by adding the term:

$$v_t = \beta v_{t-1} + (1 - \beta) d\theta \tag{6.15}$$

where β is the momentum term and $d\theta$ is defined as:

$$d\theta = -\alpha \frac{\partial L}{\partial \theta} \tag{6.16}$$

The network weights can be updated as follows:

$$\theta_t = \theta_{t-1} + v_t \tag{6.17}$$

The average of past values is now encoded in the v_t term, which represents an exponentially weighted average over past gradient values [74]. The advantage of using momentum in this manner, as opposed to storing all past values and computing their average, is computational efficiency, as only the value of v_{t-1} needs to be retained.

RMS prop

The Root Mean Squared Propagation (RMSprop) algorithm is another approach for mitigating oscillations in the undesired direction. Unlike traditional methods which rely on an average of the parameter updates, RMSprop utilizes an exponentially weighted average of the squares of the parameter updates. This approach provides a more robust estimation of the gradient, helping to dampen oscillations and improving the overall convergence of the optimization process.

The momentum is replaced by:

$$s_t = \beta s_{t-1} + (1 - \beta) d\theta^2 \tag{6.18}$$

and the weight θ can be updated as:

$$\theta_t = \theta_{t-1} + \frac{d\theta}{\sqrt{s_t + \epsilon}} \tag{6.19}$$

In this formulation, ϵ is a positive parameter that ensures the division is never by zero. The parameter updates are dampened by the squared average of their past interactions.

The ADAM optimizer

The momentum and RMS prop algorithms can be used in synergy to create the ADAM [1] optimizer [75]. ADAM computes adaptive learning rates for each parameter by considering the historical gradient information. This allows ADAM to perform well even with sparse gradients and noisy objective functions, making it suitable for a wide range of applications.

The equations that are used for the weights updates in the ADAM method are [75]:

$$\begin{aligned} v_t &= (1-\beta_1)d\theta + \beta_1 v_{t-1} \\ \hat{v}_t &= \tfrac{v_t}{1-\beta_1^t} \\ s_t &= \beta_2 s_{t-1} + (1-\beta_2)d\theta^2 \\ \hat{s}_t &= \tfrac{s_t}{1-(\beta_2)^t} \\ \theta_t &= \theta_{t-1} - \tfrac{\alpha}{\sqrt{\hat{s}_t+\epsilon}}\hat{v}_t \end{aligned} \tag{6.20}$$

The terms v_t and s_t in ADAM are subject to bias correction [76]. One of the key benefits of ADAM is its rapid convergence, which allows it to perform well in a variety of situations. The combination of momentum and RMS prop helps ADAM adapt to changing gradients and ensures efficient optimization, making it a go-to choice for training ML models.

The introduction of β_1, β_2, and ϵ is in practice considered as hyperparameters. The values for these parameters will be adopted as proposed and recommended by the authors of ADAM [75] ($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1e^{-8}$).

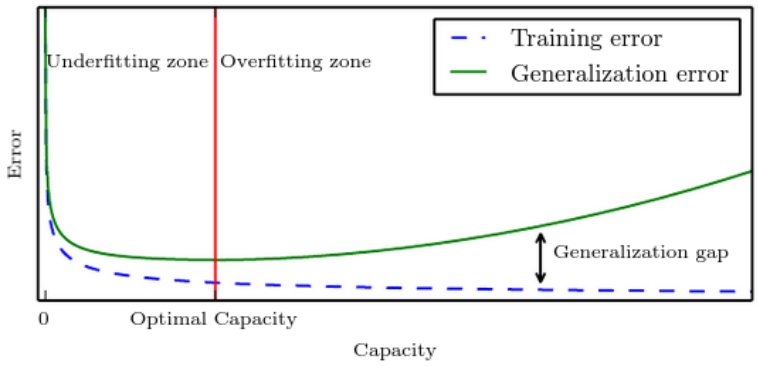
Generalization

A neural network must be able to generalize the learned information's from the training dataset to the unknown data. The ability to transfer the knowledge gained from the training dataset to unseen data is referred to as generalization. This ability is closely tied to a models capacity, which is its capability to fit a diverse range of functions. The capacity of a model increases as the number of nodes and layers in the network increases.

Having a low number of nodes and layers may result in underfitting, where the network lacks the capacity to learn the key features of the training data. On the other hand, having a high capacity, with many free parameters, may cause the network to overfit the training data by memorizing the properties of the training dataset. This behavior is depicted in Figure 6.4.

The behaviors of both underfitting and overfitting are undesirable as they significantly impact a network's ability to generalize. To reduce these problems, generalization and regularization techniques can be employed. These techniques aim to improve the gener-

[1] Adaptive Moment Estimation



(a)

Figure 6.4: Typical relationship between capacity and error. The red line signalizes the optimal capacity of the network. [76]

alizability of neural network models. Generalization is crucial for the model to perform well on completely independent MC generators and data, as training will be performed using only one type of MC samples.

One way to prevent overfitting is to consider the size of the training dataset. If the dataset is too large, it is less likely for the network to overfit it due to its size. This is noted in [76], which states that:

> "For very large datasets, regularization confers little reduction in generalization error. In these cases, the computational cost of using dropout and larger models may outweigh the benefit of regularization."

Although we will be utilizing a large training dataset, some regularization techniques will be employed to mitigate the risk of overfitting. This is necessary due to the significant number of nodes and layers present in the networks.

One such technique is the dropout probability method, which is used to improve network generalization. Dropout probability refers to the likelihood of a node from a network layer being set to zero [77]. This has the effect of reducing the network's sensitivity to individual neuron weights, which results in a network that is capable of better generalization and less prone to overfitting the training data. Instead, it encourages the network to learn more generalized features [77]. In all NN models, a dropout probability of 30% was assigned to each of the hidden layers.

Gradient clipping

Gradient clipping is another simple technique that can enhance the efficiency of network training [78]. This method is designed to alleviate the exploding gradient problem. The principle of gradient clipping involves establishing a maximum magnitude for the gradients, and any gradient that exceeds this threshold is scaled back to the threshold value.

This ensures that the model parameters are updated in a stable manner and that the training process remains consistent. The threshold value used during the training was of $1e^{-3}$.

Activation function

Given their remarkable capability to learn patterns from the data, neural networks are said to be universal function approximate, as stated by the universal approximation theorem:

> "Standard multi-layer feedforward networks with as few as one hidden layer using arbitrary squashing functions are capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficiently many hidden units are available. In this sense, multilayer feedforward networks are a class of universal approximators". [79]

One of the main aspects that enables networks to learn non-trivial patterns in data are the non-linear transformations that happen in the algorithms trough the activation (squashing)functions. Without a non-linear activation function a neural networks is just a linear regression model [76].

The universal approximation theorem was first proven for the sigmoid or also called logistic equation activation function [79]:

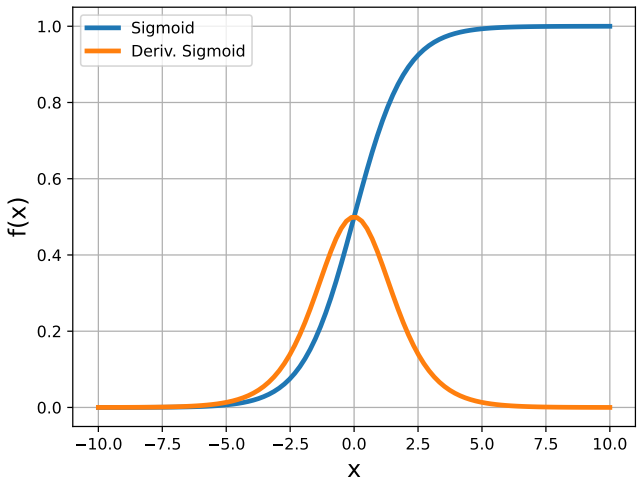
$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{6.21}$$

The sigmoid activation function was widely utilized in the past, but it has several limitations that hinder its effectiveness, including saturation (refer to Figure 6.5 for a plot of the function and its derivative). The logistic equation saturates to one for positive values of x and saturates to 0 for large negative values of x. This results in small gradients being calculated in those regions, which can lead to the vanishing gradient problem. In back-propagation, small gradients in one layer result in even smaller gradients in the previous layers, making it difficult to train the earlier layers of the network [80].

One approach to address this issue is through the utilization of activation functions such as the ReLU (Rectified Linear Unit) and Leaky ReLU, which are commonly used in contemporary network trainings. The ReLU function is defined as follows:

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.22}$$

The ReLU and Leaky ReLU activation functions have been acknowledged as significant advancements in the field of deep learning, enabling the training of complex and deep

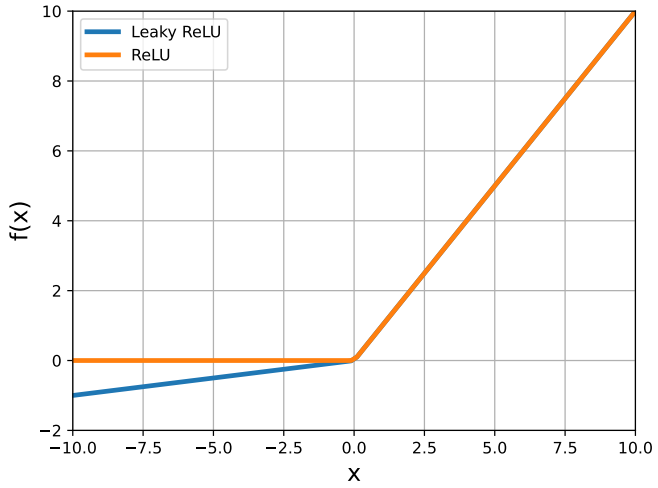


(a)

Figure 6.5: Sigmoid and its derivative behavior.

networks [76]. This can be attributed to their proficiency in accelerating the learning process and promoting the development of deeper networks, without being hindered by the gradient vanishing problem that affects other types of activation functions..

The ReLU activation function does not saturate for positive inputs and only becomes saturated when the input is negative. This results in large gradients for positive inputs, leading to faster training in deep neural networks. The behavior of the ReLU and Leaky ReLU functions is illustrated in Figure 6.6.



(a)

Figure 6.6: ReLU and Leaky ReLU functions behaviors.

The difference between the ReLU and the Leaky ReLU activation functions lies in the handling of the ”dying ReLU” problem. The ”dying ReLU” problem arises when a

ReLU activation function in a neural network is updated such that the weighted sum of its inputs is consistently negative, causing the activation function to output 0 and effectively rendering the neuron "dead". To address this issue, a variant called the Leaky ReLU function was developed [76]. The definition of the Leaky ReLU function is as follows:

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0.01, & \text{otherwise} \end{cases} \tag{6.23}$$

with a simple derivative

$$f(x)' = \begin{cases} 1, & \text{if } x > 0 \\ 0.01, & \text{otherwise} \end{cases} \tag{6.24}$$

Technically, the derivative at input 0.0 cannot be calculated. However, it is commonly assumed to be zero. In this work, we will utilize the Leaky ReLU activation function.

Weights initialization

One question that remains is how do we initialize the weights matrix once we start the training of the neural network. The weight initialization is important because it sets the starting point, and it can in some cases lead to different results based on the initialized weights. If the weights are initialized with inappropriate values, it may result in slow convergence, vanishing/exploding gradients, or sub-optimal performance. Therefore, a proper weight initialization scheme can help the model find a good local minimum, speeding up convergence, and improve the overall accuracy of the model.

For the ReLU activation functions, there are random weight initialization technqiques that are developed specifically for those activation functions, they are the He initiazlition methods [81]:

- He normal: $N(0, \frac{2}{n_{in}})$
- He uniform: $U(-\sqrt{\frac{6}{n_{in}}}, \sqrt{\frac{6}{n_{in}}})$

Learning rate scheduler

The implementation of a learning rate scheduler in deep learning models is a common technique to optimize the training process. The purpose of this technique is to adjust the learning rate throughout the training process so that the network converges smoothly to the optimal solution. The reasoning behind this is that in the early stages of training, the model parameters change rapidly, and a high learning rate is appropriate. As the model

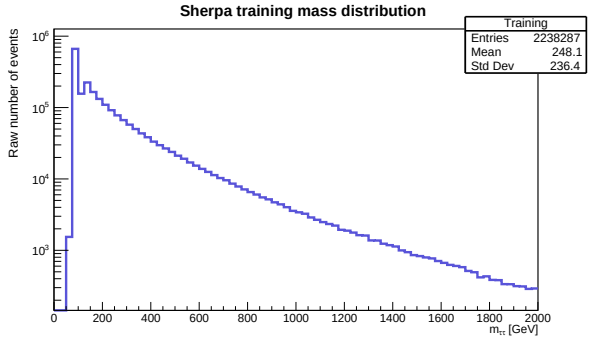
approaches the optimal solution, the changes to the parameters become smaller, and a lower learning rate is necessary to avoid overshooting it.

One example of a learning rate scheduler is the "Reduce on Plateau" scheduler, which in this work reduces the learning rate by 0.5 if the validation loss does not improve within 15 epochs, which is a complete pass trough the training data. This method is designed to help the network converge to the minimum of the loss function by gradually reducing the learning rate as training progresses.

6.3 Training sample

The careful selection of a suitable training sample is crucial for the training of a neural network. One must ensure that the training sample has good statistics, so the network has a lot of training events and so it can learn the needed characteristics from the training data. We also want to make the mass estimation as unbiased as possible. In this application, the challenge of correctly estimating mass over a wide range requires the provision of good statistics throughout the entire training range. Fortunately, the use of mass enhanced samples is able to provide the necessary statistics across the entire training range.

The training was conducted using SHERPA 2.2.11 samples, which comprised bulk and mass-enhanced samples to ensure adequate statistical coverage across the entire mass range. The target mass distributions of training samples are presented in Figure 6.7. The networks are trained with $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ so they are able to estimate the mass in both channels.



(a)

Figure 6.7: Mass target distributions of the training sample.

The SHERPA samples were divided into training and validation sets. Generation was as outlined in section 3, but the selection criteria varied. Events were not required to pass triggers, and the ID working point was set to medium. The goal of this selection was to be more inclusive and increase statistical power in training.

6.4 Machine learning architectures

We will now describe the NN architectures used in this study for mass estimation. Three different neural network architectures were evaluated: a deep feed-forward network (DNN), mixture density network (MDN), and a deep sets (DS) architecture. The implementation details of each architecture are described below.

6.4.1 Deep Feed-Forward Network

A DNN is a type of fully connected neural network that consists of multiple hidden layers, with information only flowing in a forward direction, and no cyclic connections between the layers. Figure 6.2 is a example of DNN.

In this work, the DNN architecture consisted of four layers, each containing 256 nodes. Fewer than four layers resulted in inferior biases and resolutions, and the addition of more layers did not improve the results. The Mean Squared Error (MSE) was chosen as the loss function for this study, as it is widely used in regression applications. MSE punishes large deviations from the desired output value, which can lead to lower resolutions by penalizing events that deviate significantly from the target value more severely, as the loss is calculated as the difference squared.

The network inputs and outputs used are the same for the three models used. So it will be described only once here. As the network is trained to perform the mass estimation in the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ decay channels, the input vector has room for two taus, one electron and a muon. If the event is from the $\tau_{had}\tau_{had}$ channel, the muons and electrons entries are zero. And so the network can identify the decay type by the position of the non zeroes in the input vector. The network inputs are each taus Cartesian coordinates of the three momentum, number of charged tracks if the tau decay is hadronic and the Cartesian coordinates of the missing E_T. The input vector of the neural network is the one below:

$$(p_x^{\tau_1,vis}, p_y^{\tau_1,vis}, p_z^{\tau_1,vis}, p_x^{\tau_2,vis}, p_y^{\tau_2,vis}, p_z^{\tau_2,vis}, N_{tracks}^{\tau_1}, N_{tracks}^{\tau_2}, p_x^e, p_y^e, p_z^e, p_x^\mu, p_y^\mu, p_z^\mu, E_{T_x}^{miss}, E_{T_y}^{miss}) \tag{6.25}$$

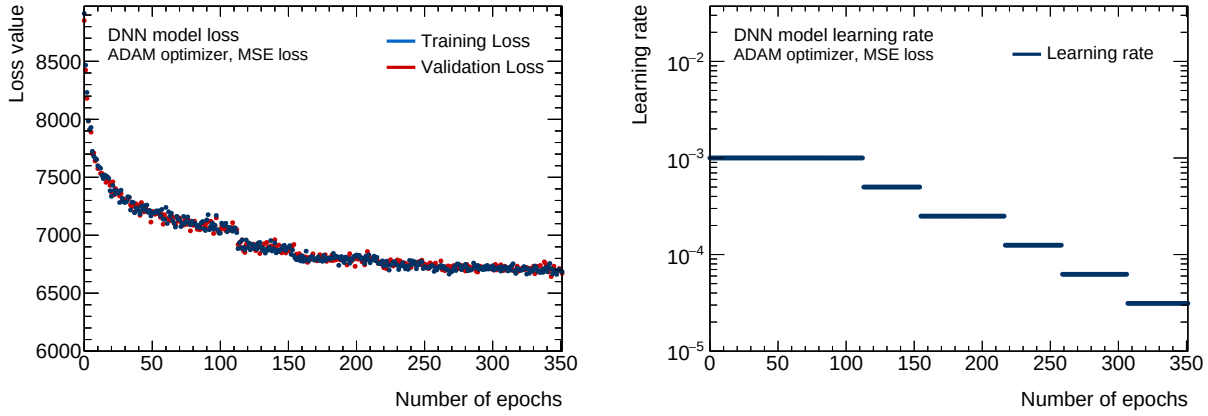
For the outputs of the NN, the mass plus other quantities of interest are outputted. Since the analysis wants to measure more than one quantity, we wanted a network that could regress all of the quantities of interest, so the full output vector is the one below:

$$(true\ E_{T_x}^{miss}, true\ E_{T_y}^{miss}, p_x^{\tau_1}, p_y^{\tau_1}, p_z^{\tau_1}, p_x^{\tau_2}, p_y^{\tau_2}, p_z^{\tau_2}, m_{\tau\tau}) \tag{6.26}$$

Tests performed with only mass as outputs had a similar performance, and the network being trained to also regress quantities like the missing transverse energy could help the network in the mass regression task, as now it also trained to regress the true missing E_T. These inputs and outputs are common to all network models (except for the mixture density networks, that has these plus a full covariance matrix as output).

The optimizer used for all network training is the Adam [75], and the framework used to build and train the all NN models in this work, is the PyTorch [71].

Regarding the training of the neural networks, Figure 6.8 shows the loss value for the training and validation datasets as a function of the number of epochs and also the learning rate as the number of epochs. The decrease in the loss function is attributed to the change in the networks learning rate, which is a result of the learning rate scheduler adjusting the rate in response to a period of stagnant loss improvement.



(a) (b)

Figure 6.8: Loss (a) and learning rate (b) as a function of epochs in the DNN model training.

6.4.2 Deep sets network

The Deep Sets network is a type of neural network that allows the jet information to be incorporated as input. The primary challenge in utilizing jet information lies in the variable number of jets produced in di-tau pairs, requiring the network to handle non-constant inputs. The Deep Sets network addresses this issue by enabling the use of variable input lengths and also boasts the advantage of being permutation invariant under the permutation of its inputs [82].

This network has already been used for b-tagging in the ATLAS collaboration [83] and for tau decay mode classification [59]. Due to its ability to handle variable input lengths and its permutation invariance, as well as its faster training time compared to RNNs, the Deep Sets network is increasingly being adopted in the collaboration for its advantageous properties.

In this work, the Deep Sets network will be applied to a regression task. The architecture of the network will be discussed, followed by its specific application in this context.

The Deep Sets network is a set-based architecture, treating the NN input as a set of smaller inputs. In the case of b-tagging, the jet being classified as b, c, or light jet is treated as a set of tracks, with each track information serving as an input to the neural network. In the mass regression task, each jet information will be utilized as input.

The network architecture was built in this way because of the theorem 2 derived in [82] which states that a function $f(X)$ operating on a set X, where X is a set of vectors $\{x_1, ..., x_m\}$, is invariant to the permutation of instances in X if f can be decomposed as:

$$f(X) = \rho(\sum_{xe\in} \phi(x)) \tag{6.27}$$

The x_i for example in the b-tagging case are the tracks of the jet. The network job is to given X (the jet) composed of x_i (tracks) act like the function $f(X)$ which yields the probability of a jet being a light b or c jet.

And rewriting $f(X)$ as $\rho(\sum_{xe\in} \phi(x))$ enables the function f to be permutation invariant, and as a bonus act on a variable number of x_i and that's is where the variable number of inputs comes from.

As a neural network is a universal function approximator, we use two different feed-forward networks to model the ρ and ϕ functions. This process can be better visualised in Figure 6.9.

For the deep sets applications on the mass estimation, each jet information (p_x, p_y, p_z, E) is treated as the x_i arrays, and the ϕ network will act on each jet, return a $\phi(x_i)$, and the sum operation is performed over the outputs of each jet. After the sum over the jet

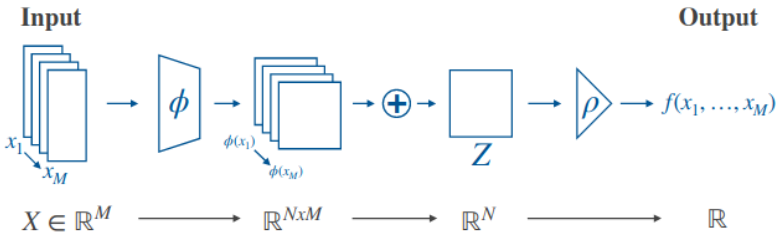


Figure 6.9: The deep sets architecture. [84]

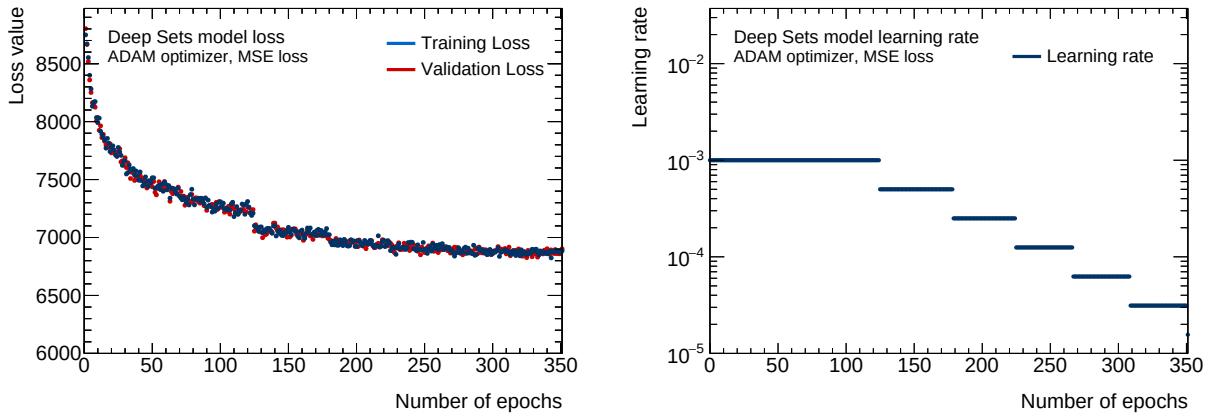
outputs, we will have a array of dimension d, which is the dimension of the $\phi(x_i)$ vectors. Supposing that the taus inputs have a dimension n we can now concatenate the two arrays, and we will have the jets and taus information. This information will then be again processed by the ρ network to produce the final output.

The information about taus cannot be used as input at the beginning of the network due to its different dimensions and inputs compared to jets. Therefore, it is added later after the processing of the jet information. Although adding the taus in this way does not make the network permutation invariant under the permutation of taus, the network is still permutation invariant under the permutation of jets.

In every application of the Deep Sets network, two networks must be built and trained. For simplicity, the network that processes the jet information will be referred to as the "jet net" (ϕ network) and the network that processes the tau information (as well as the jets information) will be referred to as the "tau net" (ρ network). The jet net has four layers, each with 16 nodes, and uses the leaky ReLU activation function. The tau net has four layers, each with 256 nodes.

The jet net does not have many nodes as the tau net because its objective is to choose from each jet the information that will be kept, and store it in the $\phi(x_i)$ output, so that the tau net will do the final processing and the mass estimation. Adding more nodes only made the network training slower and did not add performance.

The network training and validation losses and learning rates are shown in Figure 6.10.



(a) (b)

Figure 6.10: Loss (a) and learning rate (b) as a function of epochs in the Deep Sets model training.

6.4.3 Mixture density network

A Mixture Density Network (MDN) is a type of neural network that outputs a probability distribution or mixture model instead of a single prediction vector. The MDN operates similarly to a deep feed-forward network, but instead of producing a vector of predictions, it outputs the parameters of a PDF. For example, if the data follows a multivariate Gaussian distribution, the network would produce the means and covariance matrix that define the Gaussian. The PDF can then be constructed from the output of the MDN, and the loss function can be calculated as the negative log-likelihood. MDNs have the ability to represent arbitrary conditional probabilities that are conditioned on the input vector, similar to how conventional neural networks can represent any function [85]. Figure 6.11 illustrates a scheme of a Mixture Density Network.

One of the advantages of using a Mixture Density Network (MDN) is that it outputs the complete covariance matrix, which provides an estimation of the uncertainty of the predicted values. This estimation of uncertainty is not linked to experimental or statistical uncertainties, but rather reflects the network's confidence in its regression [86]

The conditional probability distribution of an n-dimensional quantity $\mathbf{y}$, given an input vector $\mathbf{x}$, can be expressed as:

$$p(\mathbf{y}|\mathbf{x}) = (2\pi)^{-n/2}|\Sigma|^{-1/2}e^{(\frac{-1}{2}(\mathbf{y}-\hat{y})\Sigma^{-1}(\mathbf{y}-\hat{y}))} \tag{6.28}$$

where Σ and $\hat{y}$ are the conditional covariance matrix and means, and they are functions of x. The Σ and $\hat{y}$ quantities should be the outputs of the neural networks, which we will use to build the PDF. Instead of making the neural network output all of the elements of the Σ matrix, we can use the Cholensky decomposition to write the Σ matrix, actually we can write the inverse of the matrix (Σ^{-1}). For example in the n=3 case:

$$\Sigma^{-1} = A^T A = \begin{pmatrix} \alpha_{11} & 0 & 0 \\ \alpha_{12} & \alpha_{22} & 0 \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ 0 & \alpha_{22} & \alpha_{23} \\ 0 & 0 & \alpha_{33} \end{pmatrix} \tag{6.29}$$

To obtain the Σ matrix one has to invert the above matrix. The determinant the matrix can now be simply calculated as:

$$|\Sigma|^{-1/2} = \alpha_{11}\alpha_{22}\alpha_{33}\alpha_{44} \tag{6.30}$$

As a result, if the goal is to output n quantities, the total number of outputs required to construct the Gaussian PDF is $n(n+3)/2$ [86]. With these outputs, we can calculate the negative log-likelihood of the Gaussian and use it as the loss function. The negative log-likelihood for a single event is given by:

$$\ell(\mathbf{y}, \hat{y}) = \frac{1}{2} log(|\Sigma|) + \frac{1}{2}(\mathbf{y} - \hat{y})^T \Sigma^{-1} (\mathbf{y} - \hat{y}) \tag{6.31}$$

With the necessary tools in place, we are now equipped to construct and train a Mixture Density Network. It is important to ensure that the activation functions utilized to obtain the diagonal coefficients of the A matrices are positive.

The evolution of the loss and the learning rate during the training of this model are illustrated in Figure 6.12.

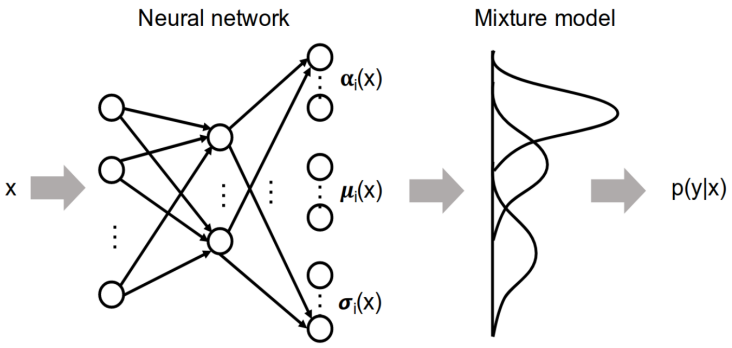
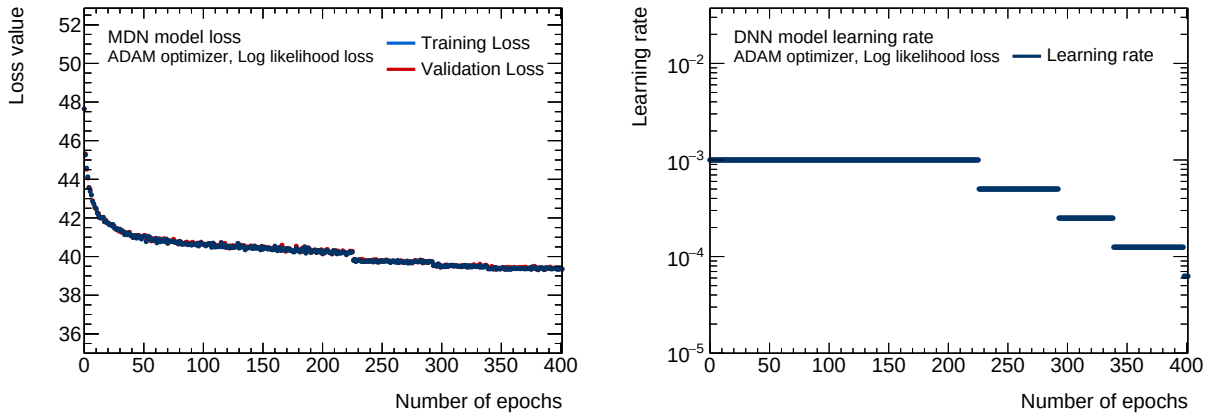


Figure 6.11: Mixture density network architecture [87]



(a) (b)

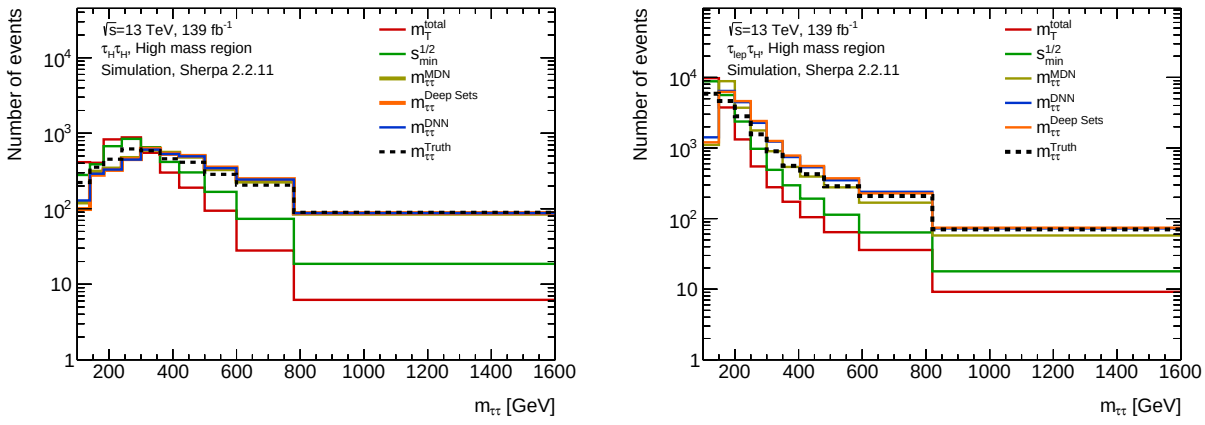
Figure 6.12: Loss (a) and learning rate (b) as a function of epochs in the Mixture density network model training.

CHAPTER 7

NN METHODS PERFORMANCE

The aim of this chapter is to assess the effectiveness of the machine learning models in mass estimation and to compare their results with those of established mass estimation methods.

After training on SHERPA 2.2.11 validation samples, the neural networks were applied to obtain a mass estimate for each event. The application of the neural network to the entire dataset produced the distributions shown in Figure 7.1. From the distributions the NN methods produced results that were closer to the generated mass distribution in high mass bins. However, they tended to slightly overestimate the truth mass distribution in those bins. The NN methods underestimated the truth mass in the lower bins of the distribution, suggesting that they tend to overestimate the mass in these bins and reconstruct them with higher masses. This behavior indicates that the NN methods have a high bias in the first bins, which overshoots the mass estimate.



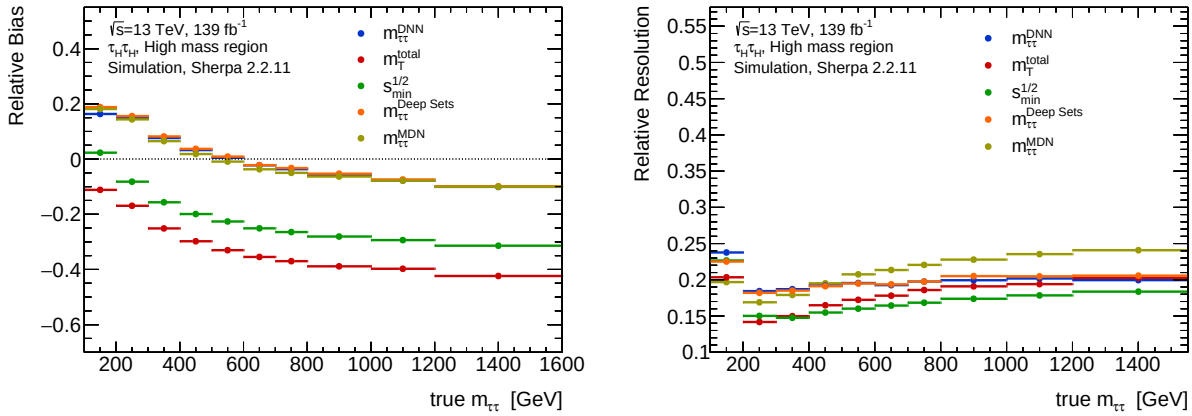
(a) (b)

Figure 7.1: Mass distributions in the high mass region of several mass estimation and NN methods in the (a) $\tau_{had}\tau_{had}$ channel and (b) $\tau_{lep}\tau_{H}$ channel.

Relative bias and resolution

By utilizing machine learning techniques on SHERPA samples, the performance of the samples in terms of bias and resolution at the high mass region can be analyzed. By employing the same selection and triggers discussed in Chapter 3, the relative bias and resolution curves can be calculated. The results are depicted in Figures 7.2 and 7.3 for the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels, respectively.

The neural network methods were observed to produce close to zero relative biases across almost the entire mass range, except for the initial bin, where it shows a high relative bias in both the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels. The relative resolutions of the neural network methods were found to be slightly inferior compared to other methods, yet remain competitive. Specifically, the MDN network demonstrated a lower resolution than other methods in most of the mass range. This can be attributed to the utilization of different loss functions. In particular, the MSE loss function calculates the squared difference between the target and predicted NN value, which magnifies the impact of values that are further from the target. As a result, the MSE loss function was observed to generate better resolution across most of the mass interval.

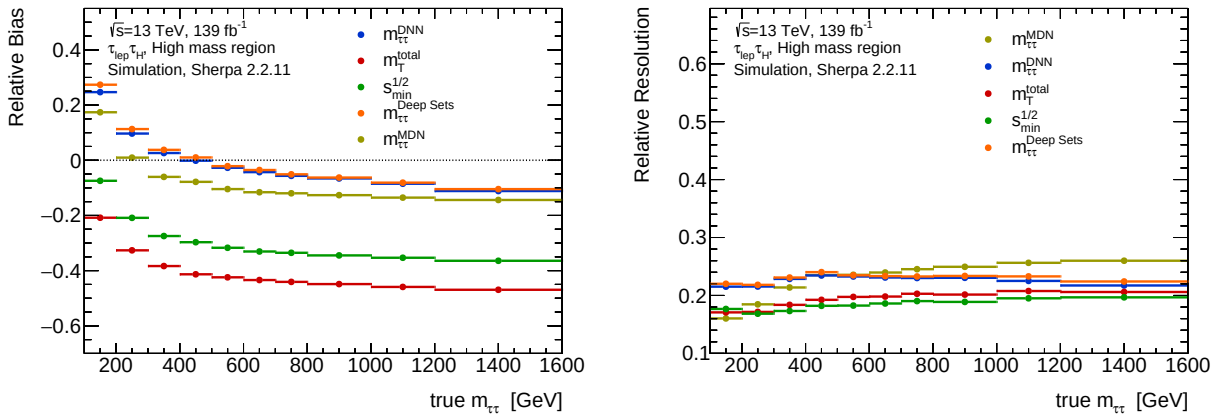


(a) (b)

Figure 7.2: Relative bias (a) and resolution (b) of the NN mass estimation methods in the $\tau_{had}\tau_{had}$ channel.

Moreover, the MDN network is obliged to generate 90 values, due to its covariance matrix, compared to the typical network which outputs only 12 values, which may result in a decrease in performance. The neural networks display similar trends in terms of bias in the $\tau_H\tau_H$ channel, however, the MDN network demonstrates slightly inferior results in the last bins of the high mass regions in the $\tau_{lep}\tau_{had}$ channel.

The DNN and DeepSets exhibit comparable behavior, with minor distinctions in their relative bias and resolution performances. This indicates that incorporating additional jet information into the network did not have a significant effect on the network's capability to estimate mass in the inclusive region. Further insights into the impact of additional

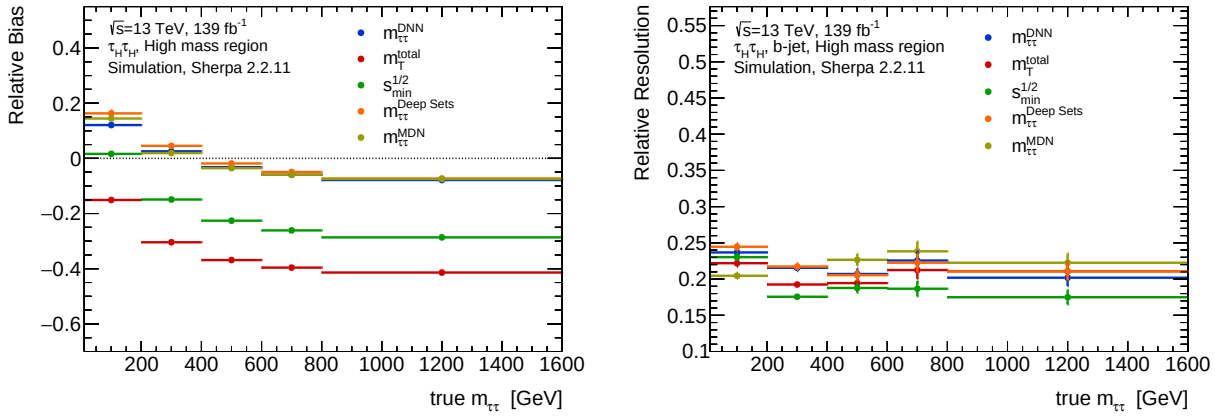


(a) (b)

Figure 7.3: Relative bias (a) and resolution (b) of the NN mass estimation methods in the $\tau_{lep}\tau_{had}$ channel.

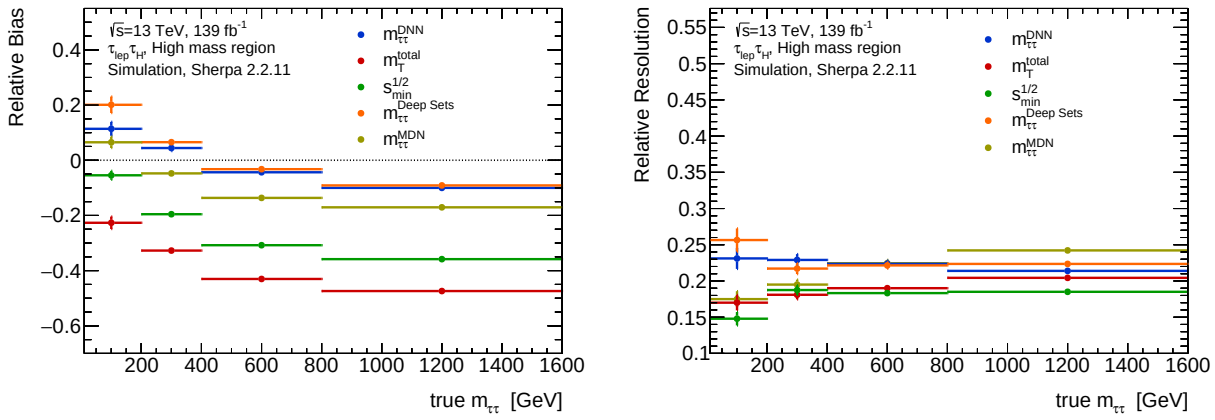
jet information on mass estimation may be obtained by examining the behavior of these methods in events where the tau pair is produced in conjunction with a b jet. The b-jet region is also significant for this measurement, as BSM models like leptoquarks have higher cross sections in the b-jet regions as reported in [19]. So it is fundamental that the mass estimation methods also have a good performance in this region.

The results of the application of neural networks to the b-jet region are presented in Figure 7.4 for the $\tau_{had}\tau_{had}$ channel and 7.5 shows the behavior in the $\tau_{lep}\tau_{had}$ channel. Due to the lower cross section of Z boson production with one b-jet compared to the inclusive cross section, the same binning used in the plots for the inclusive region was not feasible. The plot demonstrates that, once more, there is no improvement in performance through the use of DeepSets in the b-jet region. This suggests that the network is capable of learning to estimate mass by analyzing only the reconstructed tau and missing energy signature, and that the incorporation of jet information only worsened the networks resolution.



(a) (b)

Figure 7.4: Relative bias (a) and resolution (b) in the one b-jet region in the $\tau_{had}\tau_{had}$ channel.



(a) (b)

Figure 7.5: Relative bias (a) and resolution (b) in the one b-jet region in the $\tau_{lep}\tau_{had}$ channel.

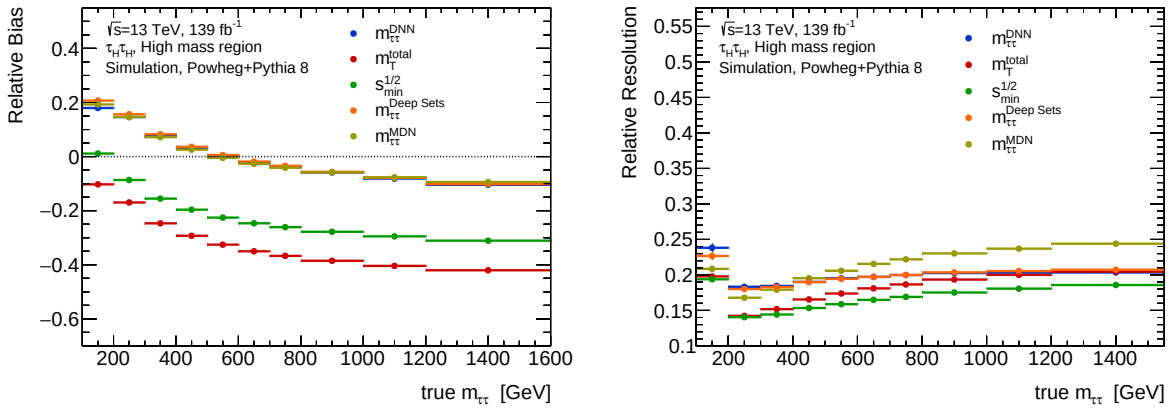
7.1 Robustness of the neural network

While neural networks were capable of delivering good results, it is crucial to exercise caution in interpreting their outputs. Neural networks may become overtrained on a specific MC sample and fail to produce comparable results when applied to data or other MC samples. To assess the robustness of the network and determine if it has become overtrained on SHERPA samples, the same inclusive relative bias and resolution results were obtained using a diferent generator. The results are presented in Figures 7.6 and 7.7.

The behavior of bias and resolution when applied to POWHEG samples is consistent with the results obtained from SHERPA 2.2.11 samples, implying that the neural network models were not overtrained and produced comparable performance when utilized with distinct and independent MC event generators.

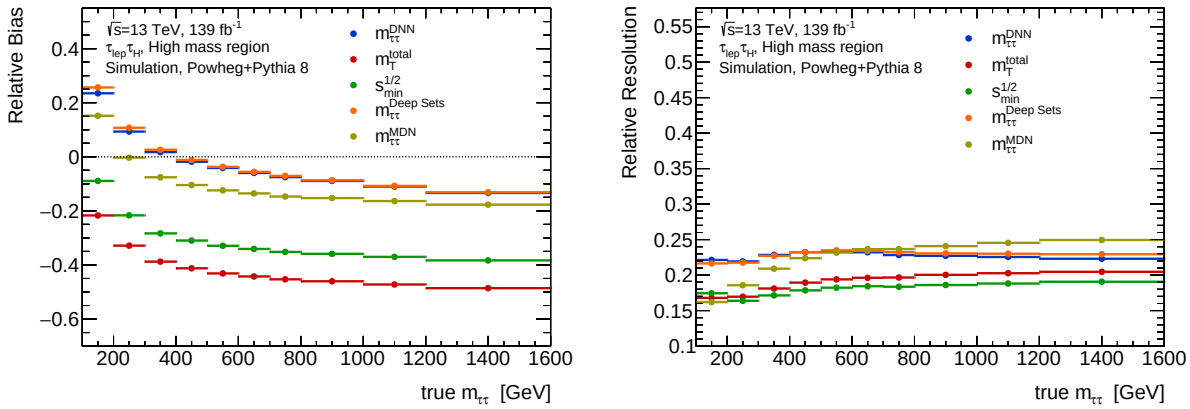
There may be slight variations in the mass estimation models, both those based on neural networks and those described in literature, however, these differences are likely due to the limited statistics in the SHERPA samples and the different modeling of the process in SHERPA and POWHEG generators. In the subsequent section, the networks will also be applied to data and background MC simulations to further evaluate their performance.

Since the MDN model showed lower performance compared to the other NN models, the MDN mass estimation will not be considered for future studies.



(a) (b)

Figure 7.6: Relative bias (a) and resolution (b) of the mass estimation methods in the POWHEG+PYTHIA 8 samples in the $\tau_{had}\tau_{had}$ chanel.



(a) (b)

Figure 7.7: Relative bias (a) and resolution (b) of the mass estimation methods in the POWHEG+PYTHIA 8 samples in the $\tau_{lep}\tau_{had}$ channel.

7.2 Response matrices

The response matrix is a crucial aspect in the measurement of unfolding and differential cross sections [88]. It is important that the response matrix be as diagonal as possible for the unfolding process to be effective.

The response matrix displays the probability that an event generated in a specific mass bin (i) will be reconstructed in another mass bin (j). Figures 7.8 and 7.9 present the response matrices for the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels, respectively. The X-axis represents the generated mass bins, while the Y-axis represents the reconstructed mass bins. For bin element in the migration matrix, the probability that an event generated in the given truth mass bin (i)will be reconstructed in a specific reco mass bin (j) is assigned. It should be noted that the sum of the bins in each column must equal one, as every generated event must be reconstructed with a specific reco mass. Although the columns in the plots may not sum to one, this is because the matrix only displays values greater than two.

The NN methods were able to produce response matrices that are significantly more diagonal in the high mass bins compared to those generated by the $s_{min}^{1/2}$ and m_T^{total} methods. This behavior is also observed in the $\tau_{lep}\tau_{had}$ channel. This is due to the bias of the $s_{min}^{1/2}$ and m_T^{total} methods, which tend to reconstruct events with lower masses than they were generated with.

The NN methods present worse results in the first bins of the mass distribution, where a significant migration is observed upward, these results are consistent with the bias curves. The NN method exhibits larger biases in the low mass bins, while the $s_{min}^{1/2}$ and m_T^{total} methods have closer-to-zero biases in these bins, resulting in less migration in the first mass bins but worse performance in the final bins of the distribution.

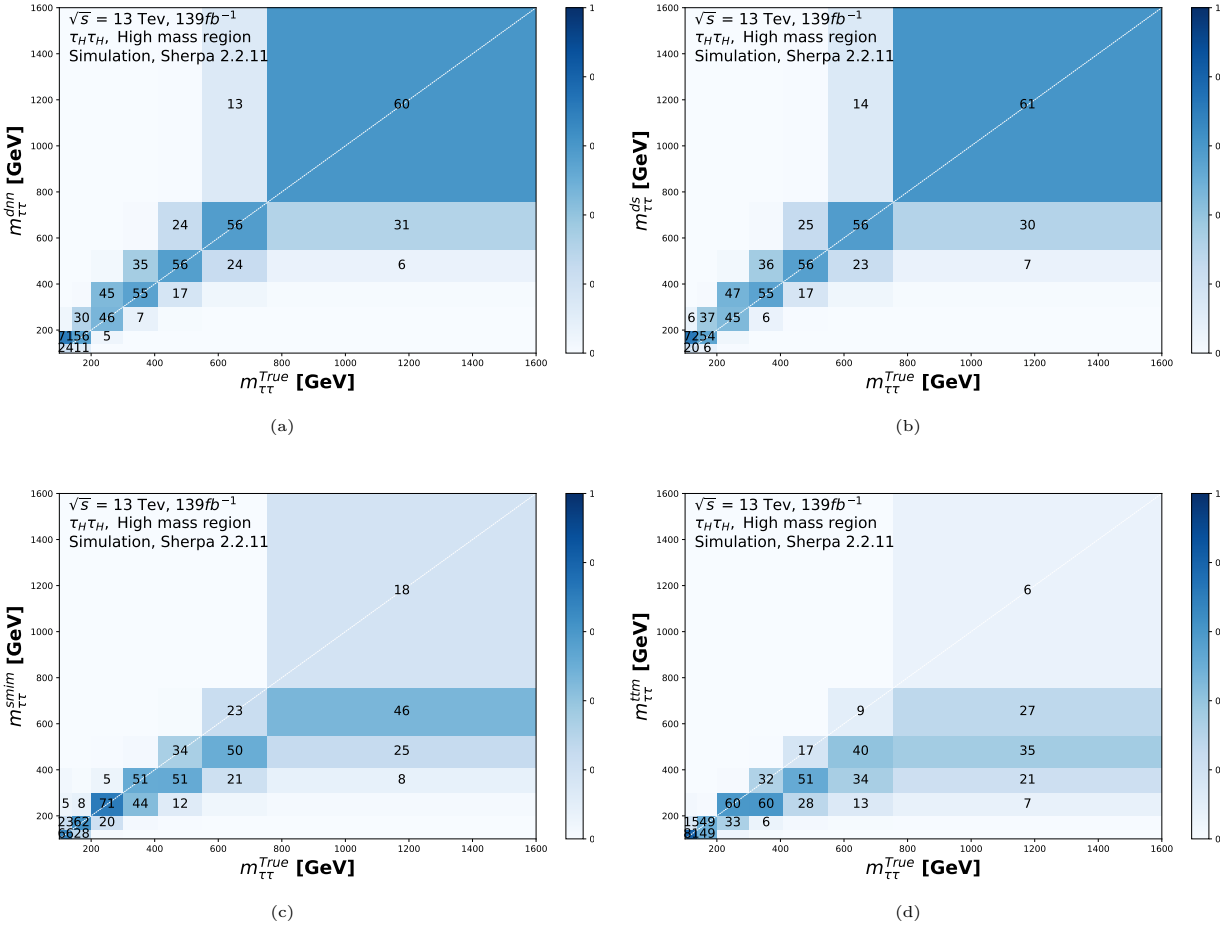


Figure 7.8: Response matrices in the HH channel for the following mass estimation methods (a) DNN (b) DeepSets (c) $s_{min}^{1/2}$ (d) m_T^{total} in the $\tau_{had}\tau_{had}$ channel

7.3 Correcting the bias of the mass estimation methods

Given that we possess the knowledge of the bias curves for a specific mass estimation method, it is feasible to adjust these bias distributions and mitigate the method's bias by fitting the bias distribution and correcting the bias. This study aims to carry out this procedure and correct the biases of the evaluated estimation methods. This adjustment will also have an impact on the methods' relative resolution and, consequently, on the response matrices, which will be recalculated. In order to correct the mass estimation for the bias, we must multiply the mass estimation by $(1/(1+b))$, where b is the bias in that given mass bin. The calculation that shows how the relative resolution responds to the bias correction is available in the Appendix B.

The goal of the fit is to make the mass migration matrix more diagonal for the unfolding process (not discussed in this work). However, it should be noted that correcting the bias

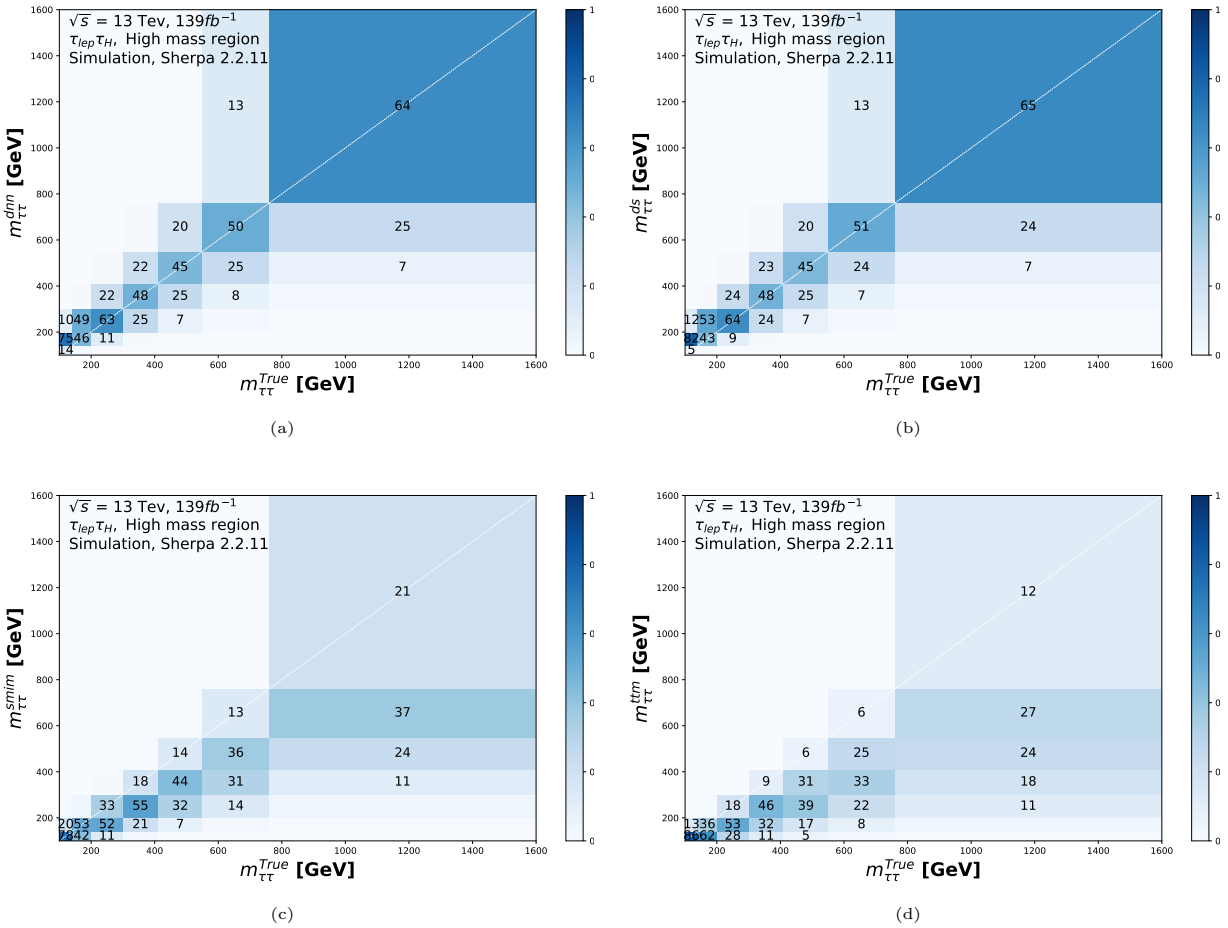


Figure 7.9: Response matrices in the le-had channel for the following mass estimation methods (a) DNN (b) DeepSets (c) $s_{min}^{1/2}$ (d) m_T^{total} in the $\tau_{lep}\tau_{had}$ channel

of a method will also result in a reduction of the method's resolution. In the following sections, the effects of this bias correction on the mass estimation methods will be studied. First, it is necessary to derive the bias curve of a method as a function of the mass estimated by the same method. The curves studied in previous chapters were plotted in truth mass bins, but since in data we do not have access to the generated mass, these curves must be derived again.

For example, in this study, the bias of the $s_{min}^{1/2}$ method will be studied in bins of $s_{min}^{1/2}$ mass, in order to perform a fit and correct the mass of each bin. The bias curves of each method and the fitted curve are available in Figure 7.10.

The bias correction and relative resolution for various mass estimation methods are presented in Figure 7.11 for both the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels. As can be seen, a larger correction to the bias leads to an increase in the relative resolution. This is evident in the $s_{min}^{1/2}$ and m_T^{total} mass estimation methods, which required substantial correction to the bias in the high mass region, leading to a significant decrease in resolution in that

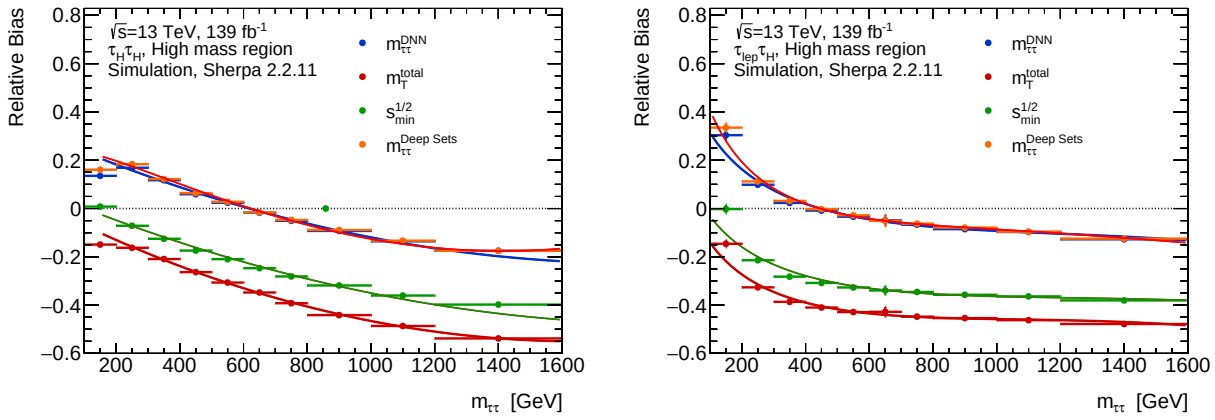


Figure 7.10: Bias curves of each estimation method and a curve fit for each in the (a) $\tau_{had}\tau_{had}$ and (b) $\tau_{lep}\tau_{had}$ channel.

region.

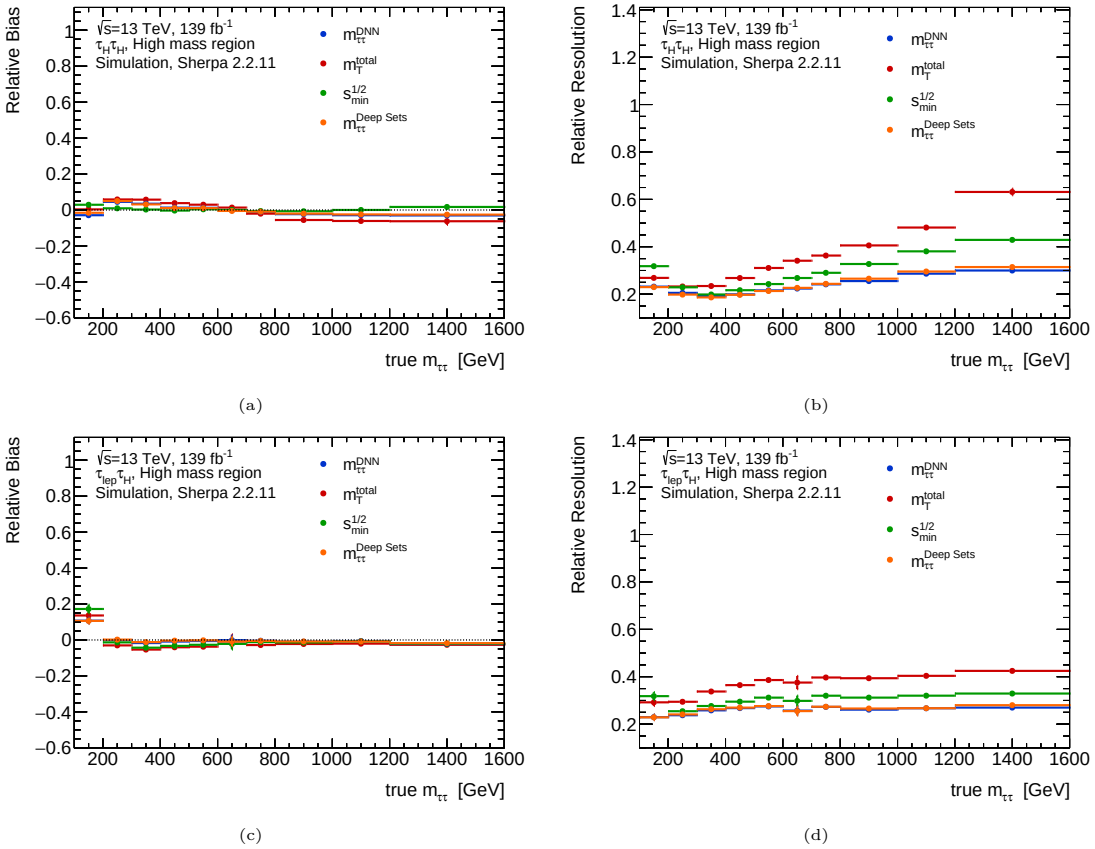
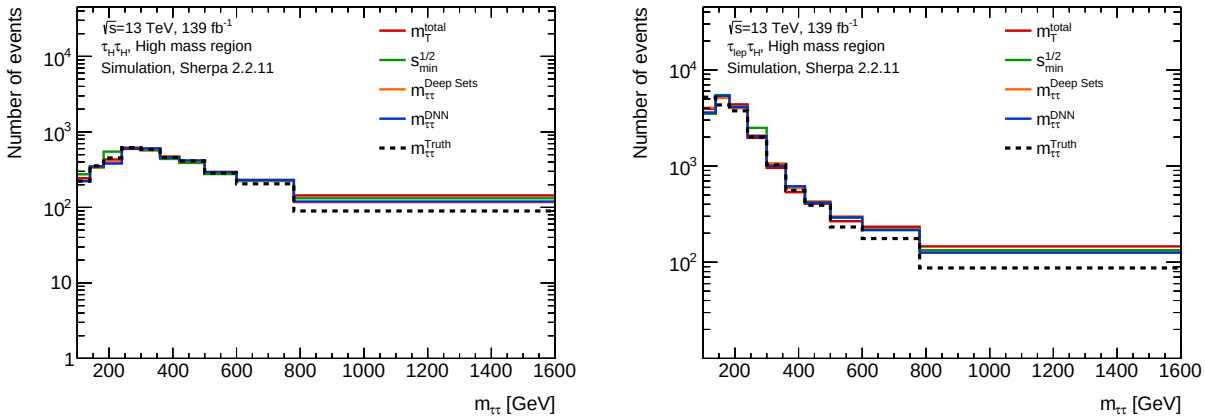


Figure 7.11: Relative bias and resolution curves of the mass estimation methods in the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channel after the bias correction.

Additionally, the fits in the $\tau_{lep}\tau_{had}$ channel underestimated the bias in the first bin. Despite the correction process, the bias in the $\tau_{lep}\tau_{had}$ channel was not fully eliminated in that bin, and a bias of approximately 10% remains. Examining the mass distributions after bias correction in Figure 7.12 reveals that there is an undershoot in the number

of events in the same bin of the $\tau_{lep}\tau_{had}$ distribution. However, there is a noticeable overall improvement in the distributions after the correction when compared to the $m_{\tau\tau}^{truth}$ distribution.

Now, we can recompute the response matrices for the bias-corrected distributions. The response matrices for the methods after bias correction are available in Figure 7.13 for the $\tau_{had}\tau_{had}$ channel, and Figure 7.14 shows the matrices for the LH channel.



(a) (b)

Figure 7.12: Detector and particle level mass distributions of the mass estimation methods after the bias correction.

The response matrices for the m_T^{total} and $\sqrt{s}_{min}^{1/2}$ methods show significant improvement in the diagonals for the high mass bins. The methods are now more comparable with the NN models in terms of diagonals at very high mass. With respect to the NN methods, the most significant improvement was observed in the first mass bins where a substantial bias was previously identified. However, in some bins, the deterioration in resolution resulted in an increase in bin migrations. After the bias correction, the methods are more competitive with each other.

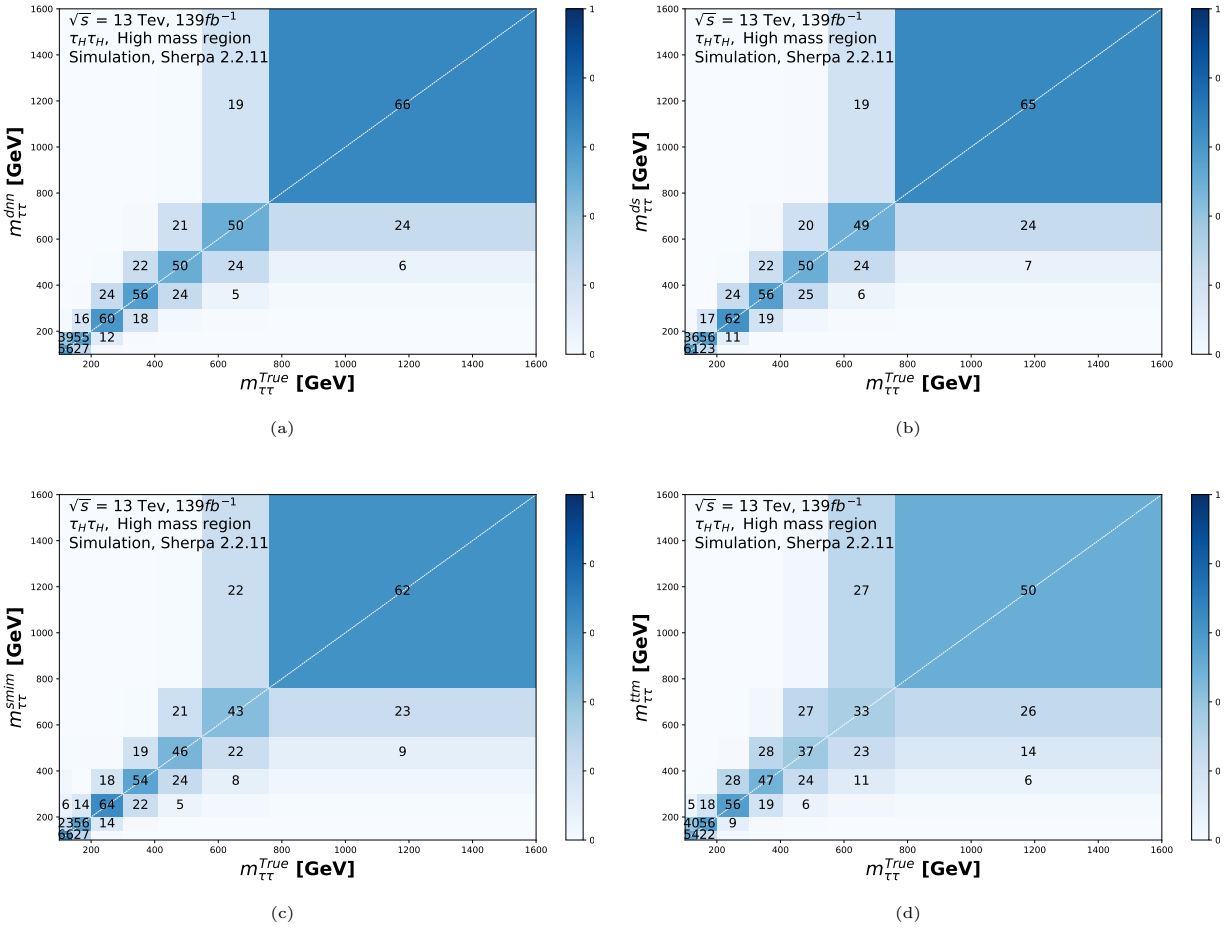
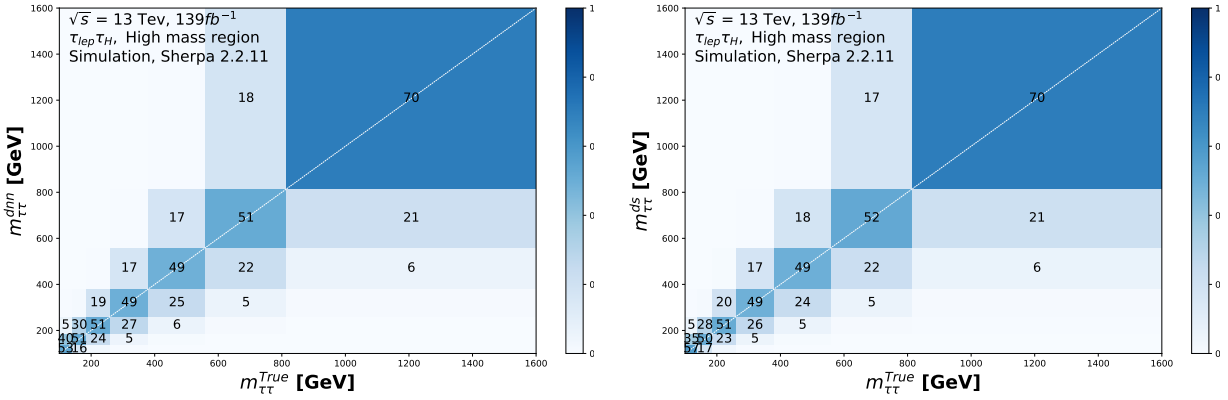
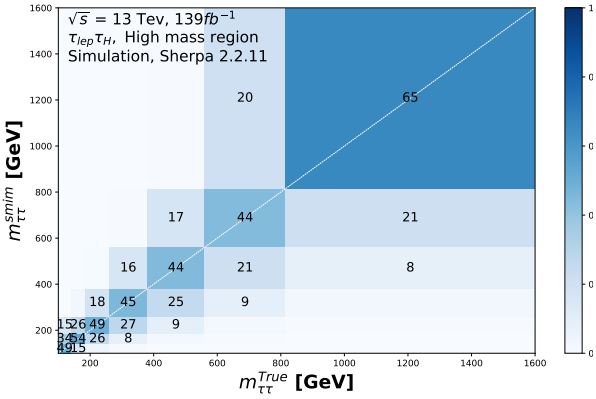


Figure 7.13: Bias corrected response matrices in the $\tau_{had}\tau_{had}$ channel for the following mass estimation methods (a) DNN (b) DeepSets (c) $s_{min}^{1/2}$ (d) m_T^{total} in the $\tau_{had}\tau_{had}$ channel

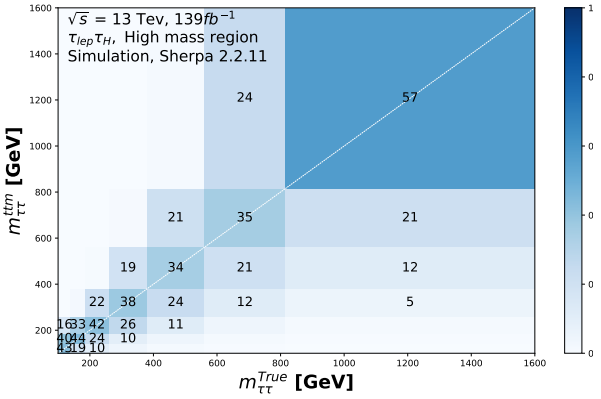


(a)

(b)



(c)



(d)

Figure 7.14: Bias corrected response matrices in the $\tau_{lep}\tau_{had}$ channel for the following mass estimation methods (a) DNN (b) DeepSets (c) $s_{min}^{1/2}$ (d) TTM in the $\tau_{lep}\tau_{had}$ channel

7.4 Background and Data reconstruction

Currently, the performance of the mass estimation methods has been thoroughly investigated in $Z \to \tau\tau$ signal samples. This investigation has included examination of the methods behavior at the resonance, in the high mass region, and with a different Monte Carlo generator. However, the reconstruction of backgrounds by these mass estimation methods has not yet been studied and analyzed. It is critical to understand the background reconstruction as it can have a significant impact on the signal-to-background ratio in the analysis.

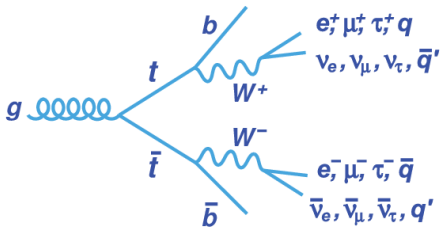
As we also search for potential BSM contributions to di-tau production in the high mass region, it is important to have a mass estimation method that accurately reconstructs the signal in this region without migrating background into it. The goal is to achieve the best possible signal-to-background ratio in the high mass region, thereby increasing the sensitivity of BSM searches at high masses.

In this section, we will apply various mass estimation methods to the full Run 2 MC and data samples to assess the agreement between MC and data, and to evaluate the performance of the methods in terms of signal-to-background ratio. The samples and selections used for data and MC were outlined in Chapter 3. It is noteworthy that the bias correction studied in the preceding chapter is not utilized in the present chapter.

In order to produce the distributions signal, background and data the analysis framework was used, where in addition to the mass estimation methods introduced in Chapter 5, the trained neural networks were also integrated into the analysis framework. This integration was performed using the ONNX runtime framework, which allows for the use of machine learning models, built and trined in python to be used in C++ [89]. For each event that meets the selection criteria outlined in section 3, the mass estimation methods are applied to the detector-level information of that event to obtain a mass estimation, and the estimation was use to fill the method histogram. This process is carried out for the signal, background, and data samples to generate the mass distributions and analyze the signal-to-background ratio and the consistency with the data.

In the studied $Z \to \tau\tau$ process, the two most significant backgrounds in the high mass region are the $W + jets$ and the $t\bar{t}$ backgrounds [20]. The $W + jets$ background occurs when a jet produced alongside the W boson is misidentified as a tau lepton, and it is particularly important in the inclusive $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels. The $t\bar{t}$ background can take two forms: it may result from the two truth taus in the $t\bar{t}$ decay chain, or it may arise from a fake tau, where a b jet from a semileptonic decay is misidentified as a tau. A leading-order Feynman diagram of the $t\bar{t}$ decay chain is shown in Figure 7.15. In regions where the di-tau pair is produced with b-jets, the $t\bar{t}$ background becomes predominant.

Conversely, in the high mass inclusive or 0-b jet region, the $W + jets$ background is the main source of contamination.



(a)

Figure 7.15: Leading-order Feynman diagram for $t\bar{t}$ decay. [90]

In addition to the $W+jets$ and $t\bar{t}$ backgrounds, there are other significant backgrounds such as the $Z+jets$ and single top backgrounds. The $Z+jets$ background is particularly important in the $\tau_{lep}\tau_H$ channel, while the single top background is important in the b-jet regions. However, these backgrounds are less dominant compared to the $W+jets$ and $t\bar{t}$ backgrounds.

Data

Now, we will examine how the methods reconstruct the data, and assess the accuracy of the estimation methods. It is important that the method has a similar reconstruction for both MC simulations and real data. Applying the method to real data provides a robust test for the ML techniques, as it ensures that the NN does not differentiate between MC and data, which can occur in some classification tasks.

As the region where the di-tau pair is produced with one b-jet is significant in searches for BSM physics [19], the distributions will be separated into 0-b jet and one b-jet categories. In the 0-b jet category, no b-jet is identified with the tau pair, while in the one b-jet category, at least one b-jet is identified alongside the tau pair.

To begin our analysis, we will examine how the deep DNN ML method reconstructs the MC simulations and real data in the high mass region. The results of using the DNN as a mass estimation technique are displayed in Figure 7.16, where the 0 and 1-b jet distributions in the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels are shown.

In the plots, only Monte Carlo background simulations are considered and a perfect closure between MC and real data is not expected, particularly in the $\tau_{had}\tau_{had}$ channel where fake backgrounds estimated by data-driven methods are of crucial importance. The QCD multi-jet background is not included in the plots because the MC simulations do not accurately model this background [51]. Thus, in the $\tau_{had}\tau_{had}$ channel, where this contribution is larger, especially in the first mass bins, there are significant discrepancies

between data and MC in the distribution displayed in Figure 7.16. Additionally, these are pre-fit plots, and the closure between MC and data may be further improved by fitting the data to the signal and control regions.

Despite this, good closure is observed between data and MC, in the 1-b jet regions and in the 0-b jet of the $\tau_{lep}\tau_{had}$ plots and in the 1-b jet regions, where we do have true taus coming from $t\bar{t}$ and single top events.

The plots include two pull plot panels. The first displays the data/MC ratio, which provides information on the closure of data to MC simulation. The second panel shows the signal-to-background ratio, which is calculated using only the MC simulations. This ratio is the division of the $Z \to \tau\tau$ signal by the sum of all background contributions in a given mass bin.

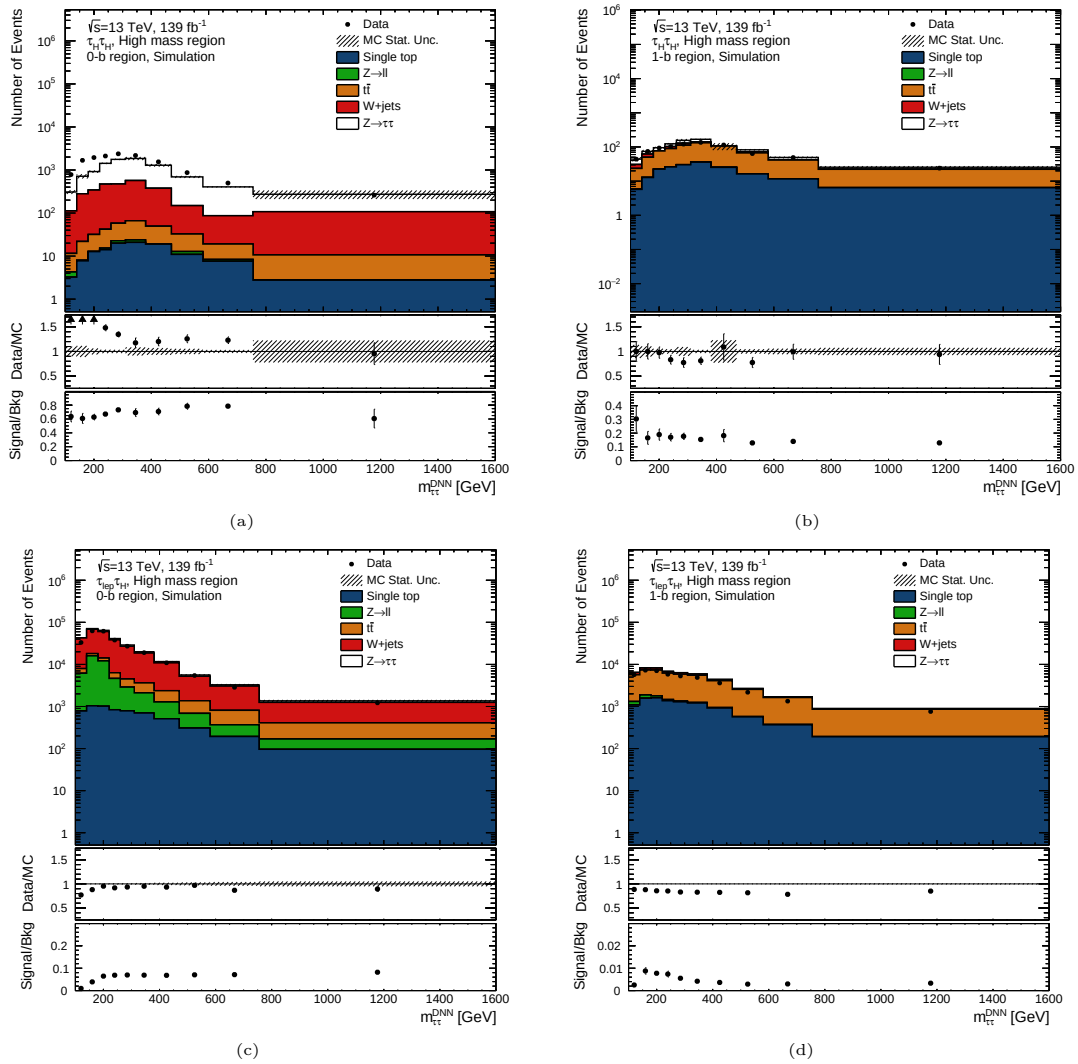
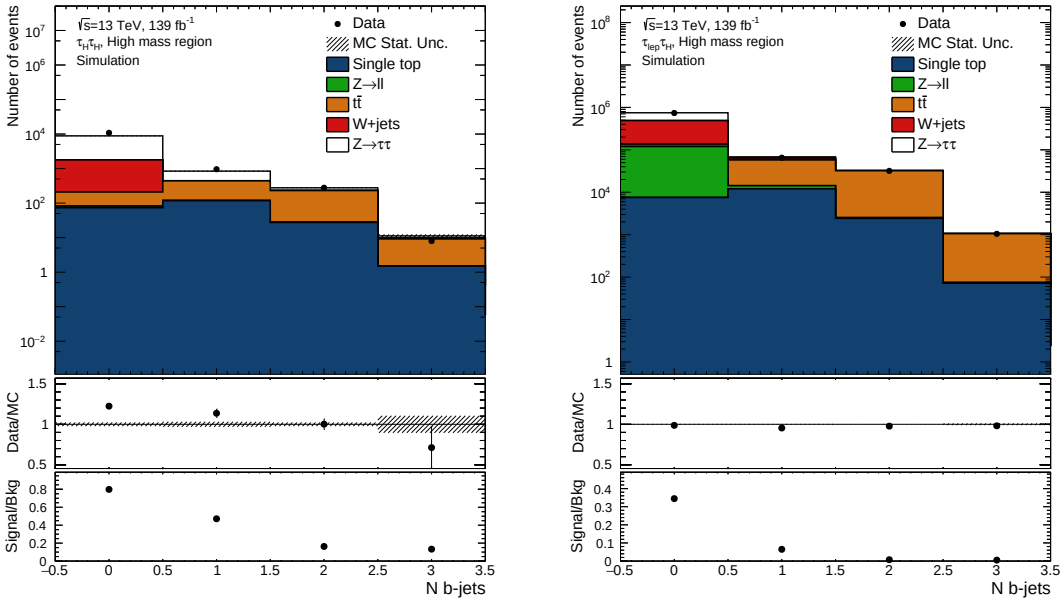


Figure 7.16: Data and MC signal and background distributions of the DNN mass estimation methods. The plots show the distributions in the $\tau_{had}\tau_{had}$ 0-b and 1-b regions for the $\tau_{lep}\tau_{had}$ 0-b jet and 1-b jet.

The signal-to-background ratios have the best results in the 0-b jet $\tau_{had}\tau_{had}$ region,

where in the 1-b region, the signal gets completely dominated by $t\bar{t}$. The $t\bar{t}$ background associated with one b-jet is significantly higher than Drell-Yan production, making it challenging to discriminate this background in the 1-b jet region. As depicted in Figure 7.17, the expected number of tau pair events is shown in terms of the number of n-b jets. In the b-jet regions, the $t\bar{t}$ background is substantially larger than the $Z \to \tau\tau$ process.



(a) (b)

Figure 7.17: Number of tau pair candidates produced alongside b-jets for the $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channel .

The m_T^{total} variable is an effective means of discrimination between Drell-Yan and $t\bar{t}$ backgrounds. Figure 7.18 displays the plots of data and MC for the m_T^{total} method. It can be observed that in the 1-b region of the $\tau_{had}\tau_{had}$ channel, the m_T^{total} method provides a signal-to-background ratio approximately five times higher than that of the DNN method. This is desirable for leptoquark searches, as the largest contributions from LQs are expected in the $\tau_{had}\tau_{had}$ 1-b jet region.

The m_T^{total} method is effective in exploiting differences in the missing Et and transverse mass distributions between the $t\bar{t}$ and $Z \to \tau\tau$ processes, contributing to its better performance. Additionally, the m_T^{total} method demonstrates improved performance in the high mass region of the $\tau_{lep}\tau_H$ channel. The m_T^{total} method also performs better at the high mass region of the $\tau_{lep}\tau_H$ channel.

However, in the 0-b jet region, the method has a tendency to reconstruct the W+jets background with higher masses due to the known high transverse masses of W bosons. As a result, the last mass bin exhibits a poor signal-to-background ratio, when compared to the DNN method.

The distributions of the DeepSets ML model can also be analyzed. These distributions are displayed in Figure 7.19. The results are similar to those of the DNN model, both

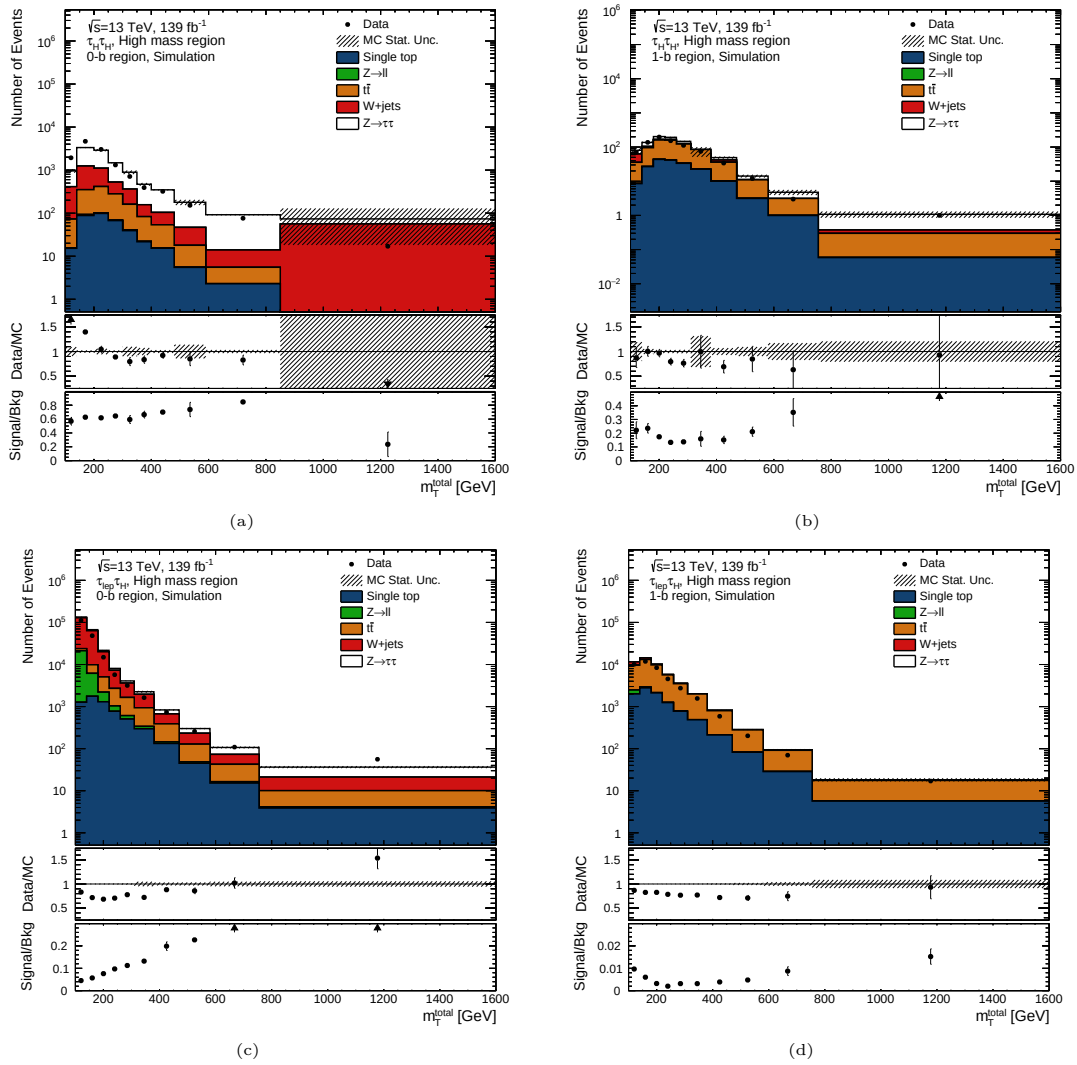


Figure 7.18: Data and MC signal and background distributions of the m_T^{total} mass estimation methods. The plots show the distributions in the $\tau_{had}\tau_{had}$ 0-b and 1-b regions for the $\tau_{lep}\tau_{had}$ 0-b jet and 1-b jet.

in terms of closure and signal-to-background ratio, as expected, as both methods have performed similarly up to this point.

Lastly, Figure 7.20 displays the distributions of the $s_{min}^{1/2}$ method. The signal-to-background ratio of the $s_{min}^{1/2}$ method is comparable to that of the NN methods, although it exhibits a slightly better ratio in the highest mass bins of the distributions.

We can explain why the m_T^{total} method is the best-performing method in the signal-to-background ratio in the b-jet region compared to the other studied methods because of the way it reconstructs the $t\bar{t}$ background. Since some part of the $t\bar{t}$ background contains truth events, we can analyze the mass bias of such events.

Figure 7.21 displays the methods' biases in the reconstruction of $t\bar{t}$ events. The neural network methods exhibit a high bias in the first mass bins, which is an undesired behavior as it leads to migration of low mass $t\bar{t}$ events into the later bins of the high mass region

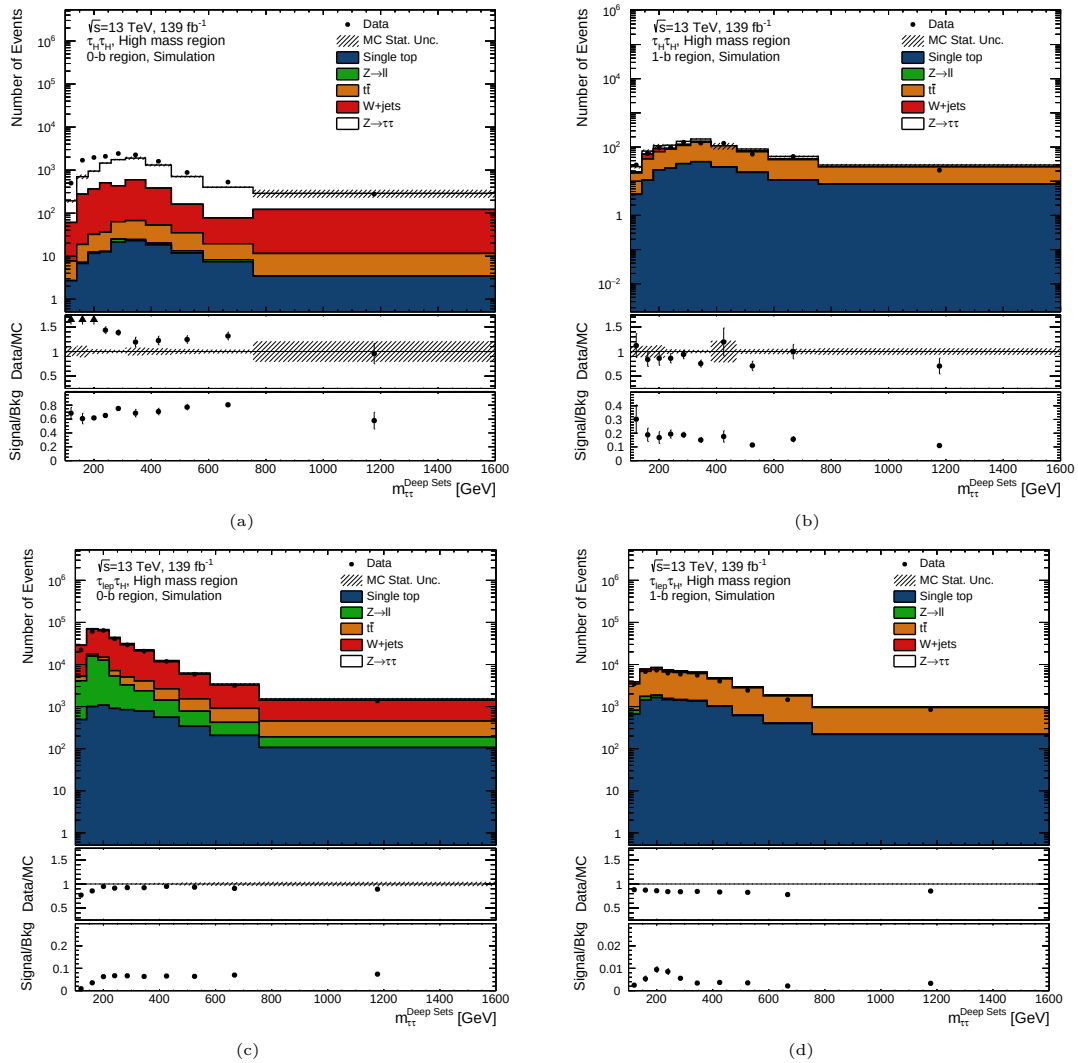


Figure 7.19: Data and MC signal and background distributions of the DeepSets mass estimation methods. The plots show the distributions in the $\tau_{had}\tau_{had}$ 0-b and 1-b regions for the $\tau_{lep}\tau_{had}$ 0-b jet and 1-b jet.

and contaminating the region of interest for many BSM searches.

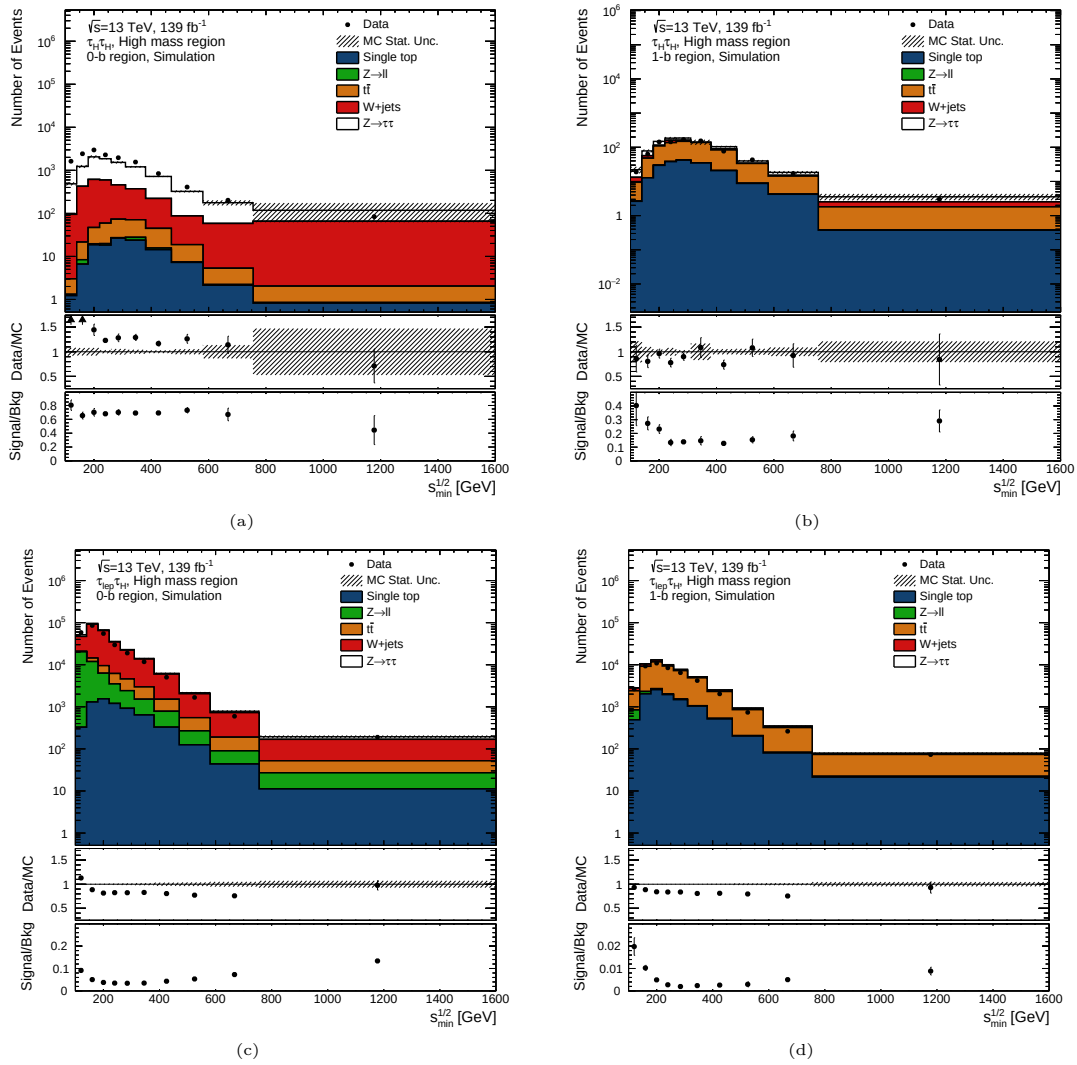


Figure 7.20: Data and MC signal and background distributions of the $s_{min}^{1/2}$ mass estimation methods. The plots show the distributions in the $\tau_{had}\tau_{had}$ 0-b and 1-b regions for the $\tau_{lep}\tau_{had}$ 0-b jet and 1-b jet.

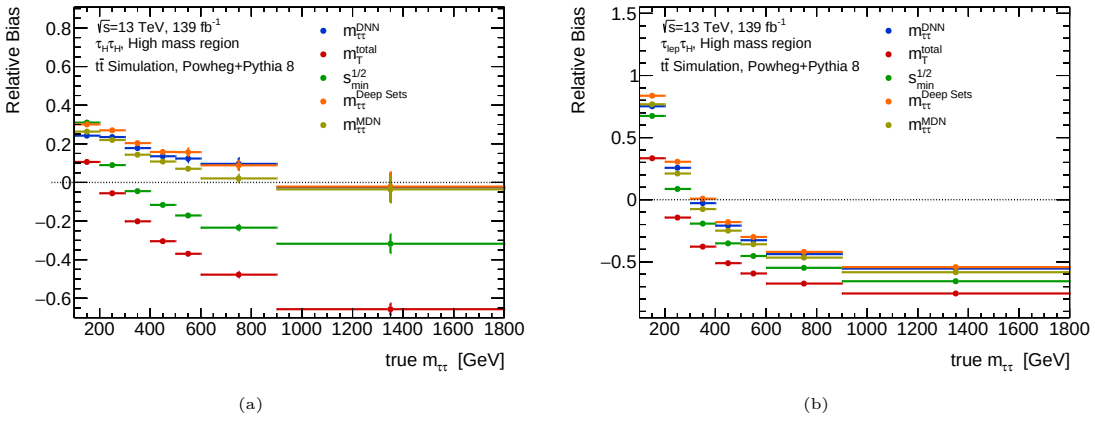


Figure 7.21: Relative bias and resolution of the different mass estimation methods in the $t\bar{t}$ background mass estimation.

CHAPTER 8

CONCLUSIONS

The high-mass region of the di-tau spectrum is an area that remains incompletely explored, despite its significant importance for various beyond the Standard Model (BSM) searches. Moreover, the lack of measurements in this region represents a limitation that necessitates attention. Mass estimation techniques are indispensable for both measurements and BSM searches. Specifically, for accurate measurements, a method that yields a diagonal response matrix is of particular importance. On the other hand, for BSM searches, a high signal-to-background ratio is critical.

The MMC method, which was commonly used for mass estimation on the Z peak, was found to be inadequate for the high-mass region due to its high failure rate. Other methods that were already available in the literature like the total transverse mass and the $\hat{s}_{min}^{1/2}$ were tested both in resonance as in the high mass region. In the resonance none of the available mass estimation methods were able to keep up with the MMC method performance. Also in the high mass region, the m_T^{total} and $\hat{s}_{min}^{1/2}$ pressente high negative biases, causing the migration of events to lower mass bins. Consequently, non-diagonal migration matrices were obtained.

By applying machine learning techniques to the mass estimation task, it was possible to achieve more diagonal migration matrices than with other methods. However, in the b-jet region, the signal-to-background ratio of the machine learning techniques was inferior to that of other methods, making it an unsuitable candidate for leptoquark searches at its current state. It will require further studies for improving its performance in the signal-to-background ratio in this region. The Total Transverse Mass method, which exploits the diferent momentum, missing transverse energy, and transverse mass signatures of the diferent process, was able to deliver the best signal-to-background ratio in the high-mass region of the b-jet region. However, it tends to reconstruct $W + jets$ events with higher

masses, which populates the high-mass region of the 0 b-jet region with W background. In the case of leptoquarks, the b-jet region is the primary area of interest. However, for models with the largest production cross section in the 0 b-jet region, the m_T^{total} method may not be a suitable mass estimation technique. In the 0 b-jet region, both the machine learning and $s_{min}^{1/2}$ methods exhibited similar signal-to-background results.

The bias correction technique improved the diagonalization of the response matrices, resulting in improved performance of the ML techniques at low mass, and enhanced performance of the m_T^{total} and $\hat{s}_{min}^{1/2}$ methods in the high mass region. However, the m_T^{total} method did not achieve a diagonal response matrix even after applying the correction. The $\hat{s}_{min}^{1/2}$ method showed the most significant improvement in performance, with only minor enhancements seen for the ML methods, primarily in the low mass region.

In summary, mass estimation in the high-mass region is a complex task, and no single method has been found to yield very good results across all regions of interest. The m_T^{total} method was able to achieve a good signal-to-background ratio in the 1-b jet region, making it a strong candidate for leptoquark searches. However, its response matrix was not diagonal, rendering it unsuitable for unfolding. On the other hand, machine learning methods were able to produce diagonal response matrices and competitive signal-to-background ratios compared to the $s_{min}^{1/2}$ method in both the 0 and 1-b jet regions. Therefore, they may be suitable candidates for unfolding. Additionally, given the excellent performance of the MMC method at the resonance, future studies should explore the re-factorization (derivation of new PDFs) of the MMC method, as it may be a promising candidate for mass estimation in the high-mass region.

All of the mass estimation methods discussed herein, including the machine learning models, have been implemented into the analysis framework and are readily available for use by the analysis team.

APPENDIX A

TRIGGER LIST USED IN THE HH AND LH CHANNEL

Table A.1: Trigger list for the di-tau semi-leptonic decay channels. The triggers in the same period of each data taking year are applied in a logical OR.

Year	Period	Trigger name (EH channel)	Trigger name (MH channel)
2015	all	HLT_e24_lhmedium_L1EM20VH	HLT_mu20_iloose_L1MU15
		HLT_e60_lhmedium	HLT_mu40
		HLT_e120_lhloose	
2016	A-D3	HLT_e26_lhtight_nod0_ivarloose	HLT_mu24_ivarmedium
		HLT_e60_lhmedium_nod0	HLT_mu50
		HLT_e140_lhloose_nod0	
	D4-end	HLT_e26_lhtight_nod0_ivarloose	HLT_mu26_ivarmedium
		HLT_e60_lhmedium_nod0	HLT_mu50
		HLT_e140_lhloose_nod0	
2017	All	HLT_e26_lhtight_nod0_ivarloose	HLT_mu26_ivarmedium
		HLT_e60_lhmedium_nod0	HLT_mu50
		HLT_e140_lhloose_nod0	
2018	All	HLT_e26_lhtight_nod0_ivarloose	HLT_mu26_ivarmedium
		HLT_e60_lhmedium_nod0	HLT_mu50
		HLT_e140_lhloose_nod0	

Table A.2: Trigger list for the di-tau fully-hadronic channel

Year	Period	Trigger name
2015	All	HLT_tau80_medium1_tracktwo_L1TAU60
		HLT_tau35_medium1_tracktwo_tau25_medium1_tracktwo_L1TAU20IM_2TAU12IM
2016	A	HLT_tau80_medium1_tracktwo_L1TAU60
		HLT_tau80_medium1_tracktwo_L1TAU60_tau50_medium1_tracktwo_L1TAU12
	B-D3	HLT_tau125_medium1_tracktwo
		HLT_tau80_medium1_tracktwo_L1TAU60_tau50_medium1_tracktwo_L1TAU12
	D4-end	HLT_tau160_medium1_tracktwo
		HLT_tau80_medium1_tracktwo_L1TAU60_tau50_medium1_tracktwo_L1TAU12
2017	B1-B7	HLT_tau160_medium1_tracktwo
		HLT_tau80_medium1_tracktwo_L1TAU60_tau50_medium1_tracktwo_L1TAU12
	B8-end	HLT_tau160_medium1_tracktwo_L1TAU100
		HLT_tau80_medium1_tracktwo_L1TAU60_tau60_medium1_tracktwo_L1TAU40
2018	B-J	HLT_tau160_medium1_tracktwoEF_L1TAU100
		HLT_tau80_medium1_tracktwoEF_L1TAU60_tau60_medium1_tracktwoEF_L1TAU40
	K-endxs	HLT_tau160_medium1_tracktwoEF_L1TAU100
		HLT_tau160_mediumRNN_tracktwoMVA_L1TAU100
		HLT_tau80_medium1_tracktwoEF_L1TAU60_tau60_medium1_tracktwoEF_L1TAU40
		HLT_tau80_mediumRNN_tracktwoMVA_L1TAU60_tau60_mediumRNN_tracktwoMVA_L1TAU40

APPENDIX B

BIAS CORRECTION

Once we do a bias correction, we want to make the relavite bias in a given mass bin 0. The bias can be defined as:

$$b = \frac{1}{n}\sum_{i=0}^{n}\frac{m_{\tau\tau}^{Est} - m_{\tau\tau}^{truth}}{m_{\tau\tau}^{truth}} \tag{B.1}$$

Which is the mean of the relative mass estimation per event. To make the bias equals zero, we will multiply each mass estimation by $1/(1+b)$ where b is the bias before the correction, of course. Then the above equation become:

$$b = \frac{1}{n}\sum_{i=0}^{n}\frac{m_{\tau\tau}^{Est}(1/(1+b)) - m_{\tau\tau}^{truth}}{m_{\tau\tau}^{truth}} \tag{B.2}$$

$$b = \frac{1}{1+b}\frac{1}{n}\sum_{i=0}^{n}\frac{m_{\tau\tau}^{Est} - m_{\tau\tau}^{truth}(1+b)}{m_{\tau\tau}^{truth}} \tag{B.3}$$

$$(1+b)b = \frac{1}{n}\sum_{i=0}^{n}\frac{m_{\tau\tau}^{Est} - m_{\tau\tau}^{truth}}{m_{\tau\tau}^{truth}} - \frac{1}{n}\sum b \tag{B.4}$$

$$(1+b)b = \frac{1}{n}\sum_{i=0}^{n}\frac{m_{\tau\tau}^{Est} - m_{\tau\tau}^{truth}}{m_{\tau\tau}^{truth}} - \frac{1}{n}\sum b \tag{B.5}$$

$$(1+b)*b = \frac{1}{n}\sum_{i=0}^{n}\frac{m_{\tau\tau}^{Est} - m_{\tau\tau}^{truth}}{m_{\tau\tau}^{truth}} - b \tag{B.6}$$

and hence, by defintition:

$$(1+b)*b = b - b = 0 \tag{B.7}$$

There are two solutions for the above equations, and they are $b = 0$ or $b = -1$. Since $b = -1$ correspond to the $m_{\tau\tau}^{Est}$ equals zero to all events, the $b = 0$ is the only physical solution. And hecen, if the mass estimations are multiplied by a given 1/(1+b) the bias in that bin will be corrected.

B.1 Effects of the bias correction in the resolution

Doing such a correction in the bias also affects the relative resolution. We can once more calculate the shift that will happens in the resolution once we do a bias correction. The relative resolution, which I will call RMS can be defined also:

$$RMS = \sqrt{\frac{1}{n}\sum (x_i - \mu)^2} \tag{B.8}$$

In our problem, the RMS can be defined as:

$$RMS = \sqrt{\frac{1}{n}\sum \left(\frac{m_{\tau\tau}^{Est} - m_{\tau\tau}^{truth}}{m_{\tau\tau}^{truth}} - b\right)^2} \tag{B.9}$$

This is the RMS before any bias correction. After the bias correction, we can approximate $b \approx 0$ and make the take the estimated mass as:

$$m_{\tau\tau}^{Est} \rightarrow \frac{m_{\tau\tau}^{Est}}{1 + b'} \tag{B.10}$$

where $b^{'}$ is the bias before the correction. Then, we $RMS^{'}$ which is the RMS after the bias correction can be written as:

$$RMS^{'} = \sqrt{\frac{1}{n}\sum \left(\frac{\frac{m_{\tau\tau}^{Est}}{1+b'} - m_{\tau\tau}^{truth}}{m_{\tau\tau}^{truth}}\right)^2} \tag{B.11}$$

Then:

$$RMS^{'} = \sqrt{\left(\frac{1}{1+b'}\right)^2 \frac{1}{n}\sum \left(\frac{m_{\tau\tau}^{Est} - m_{\tau\tau}^{truth}(1+b')}{m_{\tau\tau}^{truth}}\right)^2} \tag{B.12}$$

and the final expression is:

$$RMS^{'} = \frac{1}{1+b'}\sqrt{\frac{1}{n}\sum \left(\frac{m_{\tau\tau}^{Est} - m_{\tau\tau}^{truth}}{m_{\tau\tau}^{truth}} - b'\right)^2} \tag{B.13}$$

substituting the RMS in the right term, we have:

$$RMS^{'} = \frac{1}{1+b'} RMS \tag{B.14}$$

and hence we can link the before and after correction relative resolutions.

BIBLIOGRAPHY

[1] The ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, Sep 2012. http://dx.doi.org/10.1016/j.physletb.2012.08.020.

[2] The CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, Sep 2012. http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[3] M et all Tanabashi. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018. https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[4] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[5] S. Weinberg. *Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature*. Vintage books. Knopf Doubleday Publishing Group, 1994.

[6] Standard model of elementary particles. https://commons.wikimedia.org/w/index.php?curid=4286964. Accessed: 2023-02-13.

[7] Michel Davier, Andreas Höcker, and Zhiqing Zhang. The physics of hadronic tau decays. *Reviews of Modern Physics*, 78(4):1043–1109, Oct 2006. http://dx.doi.org/10.1103/RevModPhys.78.1043.

[8] A profile likelihood approach to measure the top quark mass in the lepton+jets channel at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2022. https://cds.cern.ch/record/2806509.

[9] Luigi Di Lella and Carlo Rubbia. The discovery of the w and z particles. *Adv. Ser. Dir. High Energy Phys.*, 23:137–163, 2015. https://cds.cern.ch/record/2103277.

[10] Andrei Angelescu and Giorgio Arcadi. Dark matter phenomenology of SM and enlarged higgs sectors extended with vector-like leptons. *The European Physical Journal C*, 77(7), jul 2017. https://doi.org/10.1140%2Fepjc%2Fs10052-017-5015-2.

[11] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and antimatter in the universe. *New Journal of Physics*, 14(9):095012, sep 2012. https://doi.org/10.1088%2F1367-2630%2F14%2F9%2F095012.

[12] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. Search for neutrinos from the sun. *Phys. Rev. Lett.*, 20:1205–1209, May 1968. https://link.aps.org/doi/10.1103/PhysRevLett.20.1205.

[13] Average of r(d) and r(d*+) for spring 2019. https://hflav-eos.web.cern.ch/hflav-eos/semi/spring19/html/RDsDsstar/RDRDs.html. Accessed: 2010-09-30.

[14] The HFLAV collaboration. New test of lepton universality using the first simultaneous measurement of r(d) and r(d*) observables at lhcb. https://indico.cern.ch/event/1187939/attachments/2530158/4355180/DTaunu_CERNSeminar.pdf, 2021.

[15] LHCb collaboration. Test of lepton universality in beauty-quark decays. *arXiv:2103.11769*, 2021.

[16] LHCb collaboration. Test of lepton universality in $b \to s\ell^+\ell^-$ decays, 2022.

[17] David London. Anomalies in b decays: A sign of new physics?, 2019. https://arxiv.org/abs/1911.06238.

[18] A.M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, and A. Search for singly and pair-produced leptoquarks coupling to third-generation fermions in proton-proton collisions at s=13 tev. *Physics Letters B*, 819:136446, 2021. https://www.sciencedirect.com/science/article/pii/S0370269321003865.

[19] Claudia Cornella, Darius A. Faroughy, Javier Fuentes-Martín, Gino Isidori, and Matthias Neubert. Reading the footprints of the b-meson flavor anomalies. *Journal of High Energy Physics*, 2021(8), aug 2021. https://doi.org/10.1007%2Fjhep08%282021%29050.

[20] Search for new phenomena in pp collisions in final states with tau leptons, b-jets, and missing transverse momentum with the ATLAS detector. Technical report, CERN, Geneva, Mar 2021. https://cds.cern.ch/record/2759282.

[21] The search for a third-generation leptoquark coupling to a τ lepton and a b quark through single, pair and nonresonant production at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2022. https://cds.cern.ch/record/2815309.

[22] Till Tantau. *The TikZ and PGF Packages*. http://sourceforge.net/projects/pgf/.

[23] Vincenzo Cirigliano, Adam Falkowski, Martín González-Alonso, and Antonio Rodríguez-Sánchez. Hadronic τ decays as new physics probes in the lhc era. *Phys. Rev. Lett.*, 122:221801, Jun 2019. https://link.aps.org/doi/10.1103/PhysRevLett.122.221801.

[24] The ATLAS collaboration. Measurement of the higgs boson coupling properties in the h $\rightarrow$ zz* $\rightarrow$ 4 decay channel at $\sqrt{s} = 13$ tev with the atlas detector. *Journal of High Energy Physics*, 2018(3), Mar 2018. http://dx.doi.org/10.1007/JHEP03(2018)095.

[25] The ATLAS collaboration. Measurements of higgs boson production cross-sections in the h $\rightarrow$ - decay channel in pp collisions at $$ \sqrt{s} $$ = 13 TeV with the ATLAS detector. *Journal of High Energy Physics*, 2022(8), aug 2022.

[26] The ATLAS collaboration. Measurement of the $z \rightarrow \tau\tau$ cross section with the ATLAS detector. *Physical Review D*, 84(11), dec 2011. https://doi.org/10.1103$%$2Fphysrevd.84.112006.

[27] Cesare Calabria. $Z \rightarrow \tau\tau$ production at CMS. *Nuovo Cimento C*, 034(6):224–225, 2011. https://cds.cern.ch/record/1363339.

[28] Measurement of Higgs boson production and decay to the $\tau\tau$ final state. Technical report, CERN, Geneva, 2019. https://cds.cern.ch/record/2668685,.

[29] The ATLAS collaboration. Measurement of the high-mass drell-yan differential cross-section in pp collisions at $\sqrt{s} = 7$ tev using the atlas detector. *Physics Letters B*, 725(4-5):223–242, oct 2013. https://doi.org/10.1016$%$2Fj.physletb.2013.07.049.

[30] CMS Collaboration. Searches for additional higgs bosons and for vector leptoquarks in final states in proton-proton collisions at $\sqrt{s} = 13$ tev. https://arxiv.org/abs/2208.02717, 2022.

[31] The ATLAS collaboration. Search for heavy higgs bosons decaying into two tau leptons with the ATLAS detector. *Physical Review Letters*. https://doi.org/10.1103$%$2Fphysrevlett.125.051801.

[32] A. Elagin, P. Murat, A. Pranko, and A. Safonov. 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(1):481–489, Oct 2011. http://dx.doi.org/10.1016/j.nima.2011.07.009.

[33] The ATLAS Collaboration. Expected performance of the atlas experiment - detector, trigger and physics, 2009. https://arxiv.org/abs/0901.0512.

[34] Partha Konar, Kyoungchul Kong, Konstantin T. Matchev, and Myeonghun Park. Reco level s√min and subsystem improved s√min. *Journal of High Energy Physics*, 2011(6), jun 2011. https://doi.org/10.1007%2Fjhep06%282011%29041.

[35] The ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation*, 3(08):S08003–S08003, aug 2008. https://doi.org/10.1088/1748-0221/3/08/s08003.

[36] Public atlas luminosity results for run-2 of the lhc. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. Accessed: 2023-02-13.

[37] The ATLAS colaboration. The ATLAS inner detector commissioning and calibration. *The European Physical Journal C*, 70(3):787–821, aug 2010. https://doi.org/10.1140$%$2Fepjc$%$2Fs10052-010-1366-7.

[38] Muon Momentum Resolution in First Pass Reconstruction of pp Collision Data Recorded by ATLAS in 2010. Technical report, CERN, Geneva, 2011. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2011-046.

[39] The ATLAS colaboration. Measurement of the photon identification efficiencies with the ATLAS detector using LHC run 2 data collected in 2015 and 2016. *The European Physical Journal C*, 79(3), mar 2019. https://doi.org/10.1140$%$2Fepjc$%$2Fs10052-019-6650-6.

[40] Francesca Cavallari. Performance of calorimeters at the lhc. *Journal of Physics: Conference Series*, 293(1):012001, apr 2011. https://dx.doi.org/10.1088/1742-6596/293/1/012001.

[41] How atlas detects particles. http://collider.physics.ox.ac.uk/detecting.html. Accessed: 2023-02-13.

[42] The ATLAS collaboration. Performance of the ATLAS trigger system in 2015. *The European Physical Journal C*, 77(5), may 2017. https://doi.org/10.1140%2Fepjc%2Fs10052-017-4852-3.

[43] ATLAS. Reconstruction, energy calibration, and identification of hadronically decaying tau leptons in the atlas experiment for run-2 of the lhc. Technical report, CERN, Geneva, 2015. https://cds.cern.ch/record/2064383.

[44] Salman et all Habib. Ascr/hep exascale requirements review report, 2016. https://arxiv.org/abs/1603.09303.

[45] Stefan Höche, Frank Krauss, Steffen Schumann, and Frank Siegert. QCD matrix elements and truncated showers. *Journal of High Energy Physics*, 2009(05):053–053, may 2009. https://doi.org/10.1088%2F1126-6708%2F2009%2F05%2F053.

[46] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. QCD matrix elements parton showers. the NLO case. *Journal of High Energy Physics*, 2013(4), apr 2013. https://doi.org/10.1007%2Fjhep04%282013%29027.

[47] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to PYTHIA 8.1. *Computer Physics Communications*, 178(11):852–867, jun 2008. https://doi.org/10.1016$%$2Fj.cpc.2008.01.036.

[48] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower monte carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010(6), jun 2010. https://doi.org/10.1007$%$2Fjhep06$%$282010$%$29043.

[49] Thomas Gehrmann, Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. NLO QCD matrix elements parton showers in ee- → hadrons. *Journal of High Energy Physics*, 2013(1), jan 2013. https://doi.org/10.1007%2Fjhep01%282013%29144.

[50] The ATLAS collaboration. Measurements of the production cross-section for a z boson in association with b-jets in proton-proton collisions at $$ \sqrt{s} $$ = 13 TeV with the ATLAS detector. *Journal of High Energy Physics*, 2020(7), jul 2020. https://doi.org/10.1007$%$2Fjhep07$%$282020%29044.

[51] The ATLAS collaboration. Measurements of the $$w$$ w production cross sections in association with jets with the ATLAS detector. *The European Physical Journal C*, 75(2), feb 2015. https://doi.org/10.1140$%$2Fepjc$%$2Fs10052-015-3262-7.

[52] S. Agostinelli, J. Allison, and K. Amako et all. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[53] Matthew Gignac, Guglielmo Frattari, Stephen Jiggins, Francesco Giuli, Richards Gonzalez, Christian Gutschow, Janina Anna Krzysiak, Marco Lisboa Leite, Federico Sforza, Lailin Xu, Yuzhan Zhao, Frank Siegert, Valerio Ippolito, Francesco Armando Di Bello, and Monika Katarzyna Juzek. Single Vector-Boson Monte Carlo setups for ATLAS Run 2 physics. Technical report, CERN, Geneva, Feb 2021. https://cds.cern.ch/record/2753199.

[54] The ATLAS collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC run 1. *The European Physical Journal C*, 77(7), jul 2017. https://doi.org/10.1140%2Fepjc%2Fs10052-017-5004-5.

[55] T Barillari, E Bergeaas Kuutmann, T Carli, J Erdmann, P Giovannini, K J Grahn, C Issever, A Jantsch, A Kiryunin, K Lohwasser, A Maslennikov, S Menke, H Oberlack, G Pospelov, E Rauter, P Schacht, F Spanó, P Speckmayer, P Stavina, and P Strízenec. Local Hadronic Calibration. Technical report, CERN, Geneva, Jun 2008. https://cds.cern.ch/record/1112035.

[56] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ik/isubit/i/subjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008. https://doi.org/10.1088/1126-6708/2008/04/063.

[57] The ATLAS collaboration. Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment. *The European Physical Journal C*, 76(5), may 2016. https://doi.org/10.1140$%$2Fepjc$%$2Fs10052-016-4110-0.

[58] Christian Limbach. Reconstruction and identification of tau leptons in atlas. *Nuclear and Particle Physics Proceedings*, 260:195–198, 2015. https://www.sciencedirect.com/science/article/pii/S2405601415000991.

[59] Sara Alderweireldt, Adam Bailey, Alessandra Betti, Quentin Buat, Yuchen Cai, Juan Carlos Jr Cardenas, Antonio De Maria, Sinead Farrington, Luca Fiorini, Christian Grefe, Guillermo Nicolas Hamity, Xiaozhong Huang, Michael Hubner, Michel Janus, Anna Kaczmarska, David Kirchmeier, Will Kostecka, Shiyi Liang, Yang Liu, Max Maerker, Bertrand Martin Dit Latour, Michaela Mlynarikova, Serhat Ordek, Gianantonio Pezzullo, Duc Bao Ta, Nadav Michael Tamir, Akanksha Vishwakarma, Estifa'A Zaid, Bowen Zhang, and Xuai Zhuang. Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data. Technical report, CERN, Geneva, May 2022. https://cds.cern.ch/record/2808869.

[60] The ATLAS collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(11):970, 2019. https://cds.cern.ch/record/2682119.

[61] ATLAS. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV. *Eur. Phys. J., C*, 81:578, 2021. https://cds.cern.ch/record/2746302.

[62] The ATLAS collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $$\sqrt{s} = 13$$ $$\text {TeV}$$. *The European Physical Journal C*, 79(8), aug 2019. https://doi.org/10.1140%2Fepjc%2Fs10052-019-7140-6.

[63] Douglas Schaefer, Emma Tolley, Teng Jian Khoo, Matthew Henry Klein, Mariyan Petrov, Marianna Testa, Fang-ying Tsai, Marco Valente, Christopher John Young, Matteo Scornajenghi, Masahiko Saito, Till Eifert, and Sarah Heim. $E_{\mathrm{T}}^{\mathrm{miss}}$ performance in the ATLAS detector using 2015-2016 LHC p-p collisions. Technical report, CERN, Geneva, 2018. https://cds.cern.ch/record/2625229.

[64] A Salzburger. The ATLAS Track Extrapolation Package. Technical report, CERN, Geneva, 2007. https://cds.cern.ch/record/1038100.

[65] David Kirchmeier. Performance of tau and muon leptons reconstruction and identification in the ATLAS experiment using pp collisions at sqrt(s)=13 TeV and their prospects for the HL-LHC. 2018. https://cds.cern.ch/record/2634851.

[66] James Catmore. Title of the slides. https://indico.cern.ch/event/472469/contributions/1982677/attachments/1220934/1785823/intro_slides.pdf, 2019. Accessed: 2020–11-06.

[67] David Cameron, Alessandra Forti, Alexei Klimentov, Andrés Pacheco Pages, and David South. Evolution of ATLAS analysis workflows and tools for the HL-LHC era. Technical report, CERN, Geneva, 2021. https://cds.cern.ch/record/2772805.

[68] The ATLAS collaboration. Measurement of the w→ cross section in pp collisions at s=7 tev with the atlas experiment. *Physics Letters B*, 706(4):276–294, 2012.

[69] The CMS Collaboration. CMS physics technical design report, volume II: Physics performance. *Journal of Physics G: Nuclear and Particle Physics*, 34(6):995–1579, apr 2007. https://doi.org/10.1088/0954-3899/34/6/s01.

[70] Graeme Peter Niedermayer. Investigations of Calorimeter Clustering in ATLAS using Machine Learning, 2018. https://cds.cern.ch/record/2299903.

[71] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[72] Kiprono Elijah Koech. The basics of neural networks (neural network series). https://towardsdatascience.com/the-basics-of-neural-networks-neural-network-series-part-1-4419e343b2b, 2022. Accessed: 2022–02-26.

[73] Martin Isaksson and Madhavun. Create a drawing of a feed-forward neural network. https://github.com/martisak/dotnets, 2022. Accessed: 2022–02-26.

[74] J. Stuart Hunter. The exponentially weighted moving average. *Journal of Quality Technology*, 18(4):203–210, 1986. https://doi.org/10.1080/00224065.1986.11979014.

[75] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. https://arxiv.org/abs/1412.6980.

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

[77] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. http://jmlr.org/papers/v15/srivastava14a.html.

[78] Jingzhao Zhang, Tianxing He, Suvrit Sra, and Ali Jadbabaie. Why gradient clipping accelerates training: A theoretical justification for adaptivity, 2019. https://arxiv.org/abs/1905.11881.

[79] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. https://www.sciencedirect.com/science/article/pii/0893608089900208.

[80] Sunitha Basodi, Chunyan Ji, Haiping Zhang, and Yi Pan. Gradient amplification: An efficient way to train deep neural networks. *Big Data Mining and Analytics*, 3(3):196–207, 2020. https://arxiv.org/abs/2006.10560.

[81] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, 2015. https://arxiv.org/abs/1502.01852.

[82] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander Smola. Deep sets, 2017. https://arxiv.org/abs/1703.06114.

[83] Deep sets based neural networks for impact parameter flavour tagging in atlas. Technical report, CERN, Geneva, 2020. https://cds.cern.ch/record/2718948.

[84] Edward Wagstaff, Fabian B. Fuchs, Martin Engelcke, Ingmar Posner, and Michael Osborne. On the limitations of representing functions on sets, 2019. https://arxiv.org/abs/1901.09006.

[85] Christopher M Bishop. Mixture density networks. 1994. https://www.sciencedirect.com/science/article/pii/S0370269321003865.

[86] Christopher M Bishop. Mixture density networks. 1994. https://publications.aston.ac.uk/id/eprint/373/.

[87] Julian Vossen, Baptiste Feron, and Antonello Monti. Probabilistic forecasting of household electrical load using artificial neural networks. In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6, 2018. https://ieeexplore.ieee.org/document/8440559.

[88] Georgios Choudalakis. Unfolding in atlas, 2011. https://arxiv.org/abs/1104.2962.

[89] ONNX Runtime developers. Onnx runtime. https://onnxruntime.ai/, 2021. Version: x.y.z.

[90] CDF and D0 Collaborations. Combination of the top-quark mass measurements from the tevatron collider. *Physical Review D*, 86(9), nov 2012. https://arxiv.org/abs/1207.1069.