Search for Direct Production of Tau Sleptons with One Leptonically and One Hadronically Decaying Tau Lepton in Run-2 Data Taken with the ATLAS Detector

Ludwig-Maximilians University Munich

Faculty of Physics

Master thesis

Author: Anna Bertolini

Supervisor: PD Dr. Alexander Mann

Munich, November 2, 2020

Suche nach direkter Produktion von Tau-Sleptonen mit einem leptonisch und einem hadronisch zerfallenden Tau-Lepton mit Daten des ATLAS-Detektors aus Run-2

Ludwig-Maximilians Universität München
Fakultät für Physik

Masterarbeit

Vorgelegt von Anna Bertolini
Betreuer: PD Dr. Alexander Mann

München, 2. November, 2020

Abstract

Supersymmetry is a theory which solves several problems of the Standard Model, such as the hierarchy problem, the unification of coupling constants and the nature of dark matter. The basic idea of this theory is to introduce a new symmetry between fermions and bosons, providing bosonic superpartners to fermions and vice-versa.
In some supersymmetry models the partner of the tau lepton, the stau, is one of the lightest supersymmetric particles and could therefore probably be produced at the LHC. An analysis of direct stau production including two hadronically decaying tau leptons as final state has already been published, setting exclusion limits to the stau mass.
The purpose of this thesis is to study whether including the channel with one leptonically and one hadronically decaying tau would help to reach sensitivity to a phase space region which was not covered in the previous analysis with the ATLAS detector.
As the separation of signal and background in the studied channel is very challenging, boosted decision trees have been used. Several approaches to a definition of a signal region have been attempted and are presented.

Contents

1 Introduction

The relatively new history of particle physics has shown a great collaboration between experimental an theoretical physics. In some cases, experiments showed the necessity to postulate a new particle, like the neutrino. In others, particles were postulated in theory and observed experimentally afterwards, like the Higgs boson.
Theoretical predictions led to a theoretical framework called the standard model (SM) to describe fundamental particles and forces. Except some relevant corrections, like the neutrino masses, experiments confirmed every prediction making of the SM a very good description of the physical world.
However, the SM is just a very good approximation at low energies as it shows shortcomings at higher ones. In fact, besides excluding the description of gravity, this framework leaves some questions unanswered.
Even if it does not include gravity, the introduction of a new symmetry can solve some problems of the standard model. This new theory called *supersymmetry* solves the hierarchy problem, the unification of coupling constants and the problem of the nature of dark matter by predicting a class of new particles to the SM that can be detected for example at the large hadron collider (LHC) at CERN. These particles are expected to have higher masses than the standard model particles, otherwise they would have already been detected at accelerators. The high energies that can be reached from the LHC as well as the high amount of recorded data are promising preconditions for the search of these particles.

This work presents a part of an analysis that searches for the direct production of a pair of these supersymmetric particle, the scalar taus, or staus. This production suffers from a very low cross section and comes with a very high number of background events. Gaining sensitivity for this search is hence a very difficult challenge.
To cope with this problem a previous analysis searching for directly produced staus only included one of the three possible final states. To improve the sensitivity that could be reached with that search, in this thesis it is analysed if the inclusion of a second final state might be helpful. The final state with one hadronically decaying tau and one leptonically decaying tau was previously excluded due to reasons explained in section 6.2.
The structure of this thesis begins with an overview of the theoretical description of the standard model and supersymmetry and it's followed by a description of the LHC and the ATLAS detector, as well as the reconstruction of objects. After a brief introduction to the concept of a statistical analysis the search for directly produced stau pairs is introduced. Finally the two methods used to separate signal from background and different approaches to the definition of a signal region are presented.

2 Theory

2.1 The Standard Model

The beginning of the history of particle physics can be dated with the first discovery of a subatomic particle. The discovery of the electron by Thomson in the 19^{th} century signed the beginning of a series of discoveries of fundamental particles and forces. In the 20^{th} century the improving knowledge of quantum mechanics made a theoretical description of the fundamental interactions possible. Models of subatomic particles could be described by quantum field theory, which unifies the theory of classical fields with special relativity.
The desire of defining a united field theory to describe every fundamental interaction led to the formulation of the Standard Model (SM), a theory that covers three fundamental forces. Many experimental discoveries, like the observation of the W and Z bosons in 1984 or the Higgs boson in 2012 confirmed that the SM successfully describes three of the four fundamental forces: the electromagnetic, weak and strong force, while gravity is excluded.

2.1.1 Particle content

The particles of the SM can be grouped in two categories depending on their spin. Particles with a half-integer spin are called *fermions* and follow the Fermi-Dirac statistic. Instead, particles with an integer spin are called *bosons* and follow the Bose-Einstein statistic. Fermions make up ordinary matter and they do not interact with each other. The interaction is mediated by force carriers, the bosons.

Fermions

Fermions are further divided into leptons and quarks. For each group the particles are also divided into three generations. In table 2.1 can be seen an overview of the fermions.
The particles electron, muon and tau are usually referred to as heavy leptons, as the neutrinos have a consistently lower mass. In the SM neutrinos are actually assumed to be massless. However, the observation of neutrino oscillations demonstrates that at least two neutrinos have to be massive. [1] [2]
Quarks are referred to as up-type (u,c,t) and down-type (d,s,b). Every fermion interacts with gravity, but besides that, fermions can be further distinguished by the force they interact with. While neutrinos can only interact weakly, heavy leptons interact also electromagnetically. Beside these two forces, quarks interact also strongly.
In addition to these 12 fermions, the SM also includes an antiparticle for each of them.

Bosons

Bosons are the force carriers of the SM. All the bosons present in the SM can be seen in table 2.2 . The photon is the carrier of the electromagnetic force as it couples only to electric charge. The carrier of the strong force, the gluon, interacts only with colour charge which is a property of quarks. There are three colour charges (red, blue and green) and their respective anti-colours. Unlike the photon which is electrically neutral, the gluon has a colour charge itself. It is composed of one unit of colour and one of anti-color. The W and the Z bosons carry the weak force which interacts with every fermion. Finally the Higgs boson gives masses to the fermions when coupling to them.

	1^{st} generation	2^{nd} generation	3^{rd} generation
leptons	electron (e)	muon (μ)	tau (τ)
	electron neutrino (ν_e)	muon neutrino (ν_μ)	tau neutrino (ν_τ)
quarks	up (u)	charm (c)	top (t)
	down (d)	strange (s)	bottom (b)

Table 2.1: List of fermions in the standard model

Name	Symbol
photon	γ
gluon	g
W boson	$W^{\pm}$
Z boson	Z^0
Higgs boson	H

Table 2.2: List of bosons of the standard model

2.1.2 Mathematical description

Mathematically the SM can be described as a quantum field theory with a

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \tag{2.1}$$

internal gauge symmetry which describes an interaction of fermionic and bosonic fields. It has to be noted that unlike classical mechanics, in quantum physics particles are not localised, but are described as quantum excitations of an underlying field and are described as functions of space and time $\phi(\mathbf{x}, t)$. Similar to classical mechanics, a Langrangian density $\mathcal{L}(\phi_i, \partial_\mu \phi_i)$ as well as the Euler-Lagrange equation:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \right) = \frac{\partial \mathcal{L}}{\partial \phi_i} \tag{2.2}$$

are needed to calculate the equations of motion.
If one looks for example at the Lagrangian for a free $\frac{1}{2}$-spin particle:

$$\mathcal{L} = i\overline{\psi}\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi \tag{2.3}$$

The resulting equation of motion is the Dirac-equation:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \tag{2.4}$$

Quantum field theories are based on the concept of local gauge invariance:

$$\psi' = e^{iq\chi(\mathbf{x},t)}\psi \tag{2.5}$$

This invariance can be guaranteed by a substitution of the derivation term which introduces a field A_μ which couples to ψ :

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu(x) \tag{2.6}$$

Furthermore, if one adds also an extra term which accounts for a free photon, the following Lagrangian can be obtained:

$$\mathcal{L} = \overline{\psi}(i\gamma^{\mu\partial_\mu} - m)\psi + q\overline{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.7}$$

This Lagrangian describes quantum electrodynamics (QED) which can be regarded as an independent quantum field theory which describes the electromagnetic interaction and is invariant under the symmetry group U(1).
The part of the SM which is invariant under the symmetry group $SU(3)_C$ is quantum chromodynamics (QCD), i.e the interaction of gluons with colour-charged particles. However, the other two symmetry groups of the SM cannot be easily attributed to the electromagnetic and weak force. Instead, in the SM these two forces are unified into the electroweak interaction described by the symmetry groups $SU(2)_L \times U(1)_Y$.
The subscripts of each symmetry group refers to the coupling of the force carriers. While C stands for the colour charge, in the electroweak interaction the L refers to left-handed particles, while Y is the hypercharge which is combined together with the electric charge Q by the Gell-Mann–Nishijima formula:

$$Q = I_3 + \frac{1}{2}Y \tag{2.8}$$

where I_3 is the third component of the weak isospin. The SU(2) gauge group has three generators and is associated to a triplet of gauge bosons $W^1_\mu, W^2_\mu, W^3_\mu$ while the U(1) group is associated to one gauge boson B_μ. Two of the W gauge bosons can be identified with the physical particles $W^\pm$ that carry the weak force. Instead, the two left gauge bosons combine together to the physical photon and Z boson as the following:

$$\begin{aligned} A_\mu &= \quad B_\mu cos\theta_w + W^3_\mu sin\theta_w \\ Z_\mu &= -B_\mu sin\theta_w + W_\mu cos\theta_w \end{aligned}$$

where the mixing angle θ_w is also called Weinberg angle.
As described until now the SM does not include the masses of particles. However, experimental measurements clearly show that most of the fundamental particles are massive. To account for this fact the Higgs mechanism is introduced. The Higgs field is described by a complex scalar doublet $\Phi = \begin{pmatrix} \phi^\dagger \\ \phi^0 \end{pmatrix}$ with the following components:

$$\phi(x) = \frac{1}{\sqrt{(2)}}(\phi_1(x) + i\phi_2(x)) \tag{2.9}$$

The potential of the Higgs field

$$V(\Phi) = -\mu^2|\Phi|^2 + \lambda^2|\Phi|^4 \tag{2.10}$$

induces a spontaneous symmetry breaking of three of the four generators. The three bosons of the weak interactions couple to those three generators thereby gaining their mass. The remaining degree of freedom of the Higgs field instead represents the massive Higgs bosons which couples to fermions giving them their mass. [3]

2.2 Supersymmetry

Experimental observations demonstrate the high precision of the SM. However, it is a theory which lacks to explain certain aspects of particle physics, besides not including gravity. To overcome this shortcoming, theorists proposed a fundamental symmetry which relates bosons to fermions, the so-called *Supersymmetry* (SUSY).

2.2.1 The Hierarchy Problem

The SM can be seen as a good approximation for the electroweak energy scale M_W ($\sim 10^2 GeV$). However, at the Planck scale M_P ($\sim 10^{18} GeV$) gravitational effects cannot be neglected anymore.
The importance of the discrepancy between $M_W$and M_P for the reliability of the SM becomes clear when looking at the Higgs mass. The Higgs potential (see equation 2.1.2) needs a non vanishing expectation value, which occurs for $< H >= \sqrt{-\frac{\mu^2}{2\lambda}}$ if λ is positive. The Higgs mass has been experimentally measured to be around 125 GeV. Its potential parameters are then calculated to be μ^2 = -(92.9 GeV)^2 and $\lambda = 0.126$. However, the parameter μ receives considerable quantum corrections from every particle that couples directly or indirectly to the Higgs field. If the Higgs field couples for example with a fermion, then the correction would be:

$$\Delta\mu^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + ... \tag{2.11}$$

where Λ_{UV} is a cutoff value that can be interpreted as the energy scale at which the theory starts to alter. If this value is of order of M_P the quantum corrections to μ^2 can be 30 orders of magnitude larger than the required value. This is referred to as the hierarchy problem.
Renormalisation would manage to cancel out these divergences, but this would result in a very unnatural process of fine-tuning. Instead, SUSY could provide a more elegant solution to this problem. In fact loops of the new particles postulated by SUSY would cancel out the contributions of the SM, as bosons and fermions contribute with opposite signs to loop corrections. [4]

2.2.2 Unification of Coupling Constants

The couplings of the SM forces have an energy dependent strength. This fact is referred to as the *running* of the coupling constant. At the electroweak scale these constants have similar

strengths:

$$\alpha^{-1} : \alpha_W^{-1} : \alpha_S^{-1} = 128 : 30 : 90 \tag{2.12}$$

In QED the coupling constant increases with energy, while in the strong force gluon loops make it decrease. With a smaller slope, also the coupling of the weak force decreases with energy. This behaviour therefore tends to bring the value of the coupling constants together and it is reasonable to imagine that at a high energy scale the couplings would converge at one point. However, this is not the case in the SM, as can be inferred in figure 2.1 on the left. The existence of new particles as provided by SUSY would contribute with extra fermionic and bosonic loops and imply a convergence of the three coupling constant (figure 2.1 on the right).

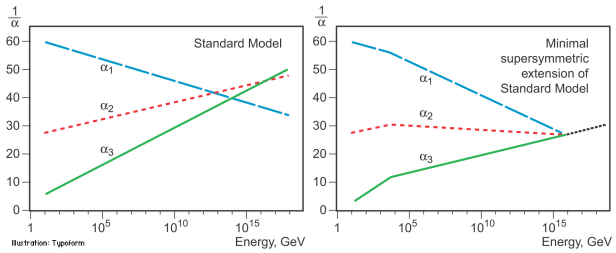


Figure 2.1: Illustration of the running of the coupling constant in the SM (left) and with SUSY (right)[5]

2.2.3 Dark Matter

Another aspect of particle physics the SM cannot provide an answer to is the nature of dark matter. Since the 1930s it is known that a significant part of the matter in the universe does not interact with photons. For this reason this unknown matter is referred to as *dark*. A compelling evidence of the existence of these unknown particles is given by the velocity distributions of stars which orbit the galactic centre. The equation of centripetal acceleration in a gravitational field

$$\frac{mv^2}{r} \approx \frac{Gm}{r^2} M(r) \tag{2.13}$$

tells us which mass distribution to expect for an observed velocity. Therefore, the inconsistent measurements of masses and velocities imply that the mass of a galaxy is made by a significant fraction of a non-luminous (dark) matter.
Further observations, like the small fluctuations in the Cosmic Microwave Background (CMB) add reliability to the existence of dark matter.
These observations therefore suggest the existence of previously unobserved particles that interact only weakly. In fact, the only SM particles which are known to interact only weakly are neutrinos. However, their relativistic nature as well as constraints on their mass make them accountable only for a small fraction of dark matter and cannot be the dark matter's only source.[6] The unknown particles which should form dark matter are referred to as weakly interacting massive particles (WIMPs).
An extension of the SM like SUSY would account for this type of particles, as many supersymmetric models imply that the lightest supersymmetric particle, the neutralino $\widetilde{\chi}_1^0$, is stable and only weakly interacting. [7]

2.2.4 Mathematical Description

As already mentioned, SUSY describes a symmetry between fermions and bosons. Bosonic states are transformed into fermionic states and vice versa. Such transformation is generated by a fermionic operator Q:

$$Q|\text{Boson}\rangle = |\text{Fermion}\rangle \tag{2.14}$$
$$Q|\text{Fermion}\rangle = |\text{Boson}\rangle \tag{2.15}$$

The generators Q and $Q^\dagger$ must satisfy the following commutation and anticommutation relations:

$$\{Q, Q^\dagger\} = P^\mu \tag{2.16}$$
$$\{Q, Q\} = \{Q, Q^\dagger\} = 0 \tag{2.17}$$
$$[P^\mu, Q] = [P^\mu, Q\dagger] = 0 \tag{2.18}$$

P^μ is the four-momentum generator of space-time translations. Irreducible representations of the supersymmetric algebra are called *supermultiplets*.
The SM particles and their supersymmetric partners are both contained in supermultiplets. The generators of supersymmetry Q and $Q^\dagger$ commute with the generators of gauge transformations. Particles of one multiplet must therefore have the same charge, isospin, and colour degrees of freedom. Particles residing in the same supermultiplet must also have equal eigenvalues, since $-P^2$ commutes with the operators Q and $Q^\dagger$. This implies that particles of the SM and their *superpartners* have to have the same mass.
However, if electromagnetically or strong interacting particles of the same masses as the SM particles would exist, they would already have been detected. Experimental observation therefore imply that SUSY must be a broken symmetry. In order to ensure the capability of SUSY to provide a solution to the hierarchy problem, this symmetry breaking has to be *soft*. This results into the Lagrangian

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}} \tag{2.19}$$

where $\mathcal{L}_{\text{SUSY}}$ contains gauge and Yukawa interactions, preserving the invariance of supersymmetry, while $\mathcal{L}_{soft}$ breaks the symmetry. [4]

2.2.5 The Minimal Supersymmetric Standard Model

The minimal supersymmetric standard model (MSSM) is the SUSY model which contains the lowest number of new particles. An overview of the included particles can be found in table 2.3.
The superpartners of SM fermion, are scalars, since they have a spin value of 0. These particles have a name which is constructed by an *s* and the usual SM particle name. For instance, the superpartner of a tau lepton is called a *stau*. Instead, fermionic supersymmetric particles are distinguished by the suffix *-ino*. The symbols of the SUSY particles are characterised by a tilde.
The simplest multiplet contains one chiral spin $\frac{1}{2}$ fermion and two scalars. This assures an equality of degrees of freedom $n_f = n_B$ since the fermion has two helicity degree of freedom

and the scalar fields have one degree of freedom each. This multiplet is composed of left- and right-handed quarks and leptons. Due to the helicity of the SM particles also its superpartners are usually written with an index of handedness, which refers to the handedness of the respective SM particles.
A further supermultiplet contains a spin 1 gauge boson whose superpartner is a fermion. This multiplet represents the SM gauge bosons ($W^{\pm}$, W^3, B, see section 2.1.2) as well as their superpartners called *gauginos* (gluinos, winos and binos). Like in the SM, where W and B mix to the Z boson and the photon, gauginos mix to zino and photino.
The Higgs boson resides in the chiral supermultiplet. However, unlike in the SM the MSSM needs two Higgs bosons with different hypercharge. These boson are called H_u and H_d as each one couples to up-type or down-type quarks only. The Higgs boson as described in the SM has to be seen as a linear combination of H_u and H_d. The superpartners of the Higgs bosons are called *Higgsinos*.
Finally, the electroweak gauginos and the Higgsinos can mix resulting in *charginos* ($\widetilde{\chi}^{\pm}$) and neutralinos ($\widetilde{\chi}^0$). [8]

	fermion spin $\frac{1}{2}$	scalar spin 0
quark, squark	(u_L, d_L)	$(\widetilde{u}_L, \widetilde{d_L})$
	$u_R^{\dagger}$	$\widetilde{u}_L^*$
	$d_R^{\dagger}$	$\widetilde{d}_L^*$
lepton, slepton	(ν, e_L)	$(\widetilde{\nu}, \widetilde{e}_L)$
	$e_R^{\dagger}$	$\widetilde{e}_R^*$
higgsino, higgs	$(\widetilde{H}_d^0, \widetilde{H_d^-})$	(H_d^0, H_d^-)
	$(\widetilde{H}_d^+, \widetilde{H}_d^0)$	(H_d^+, H_d^0)
	boson spin 1	fermion spin $\frac{1}{2}$
gluon, gluino	g	$\widetilde{g}$
W boson, winos	$W^{\pm}W^0$	$\widetilde{W}^{\pm}\widetilde{W}^0$
B bososn, bino	B^0	$\widetilde{B}^0$

Table 2.3: Chiral and gauge supermultiplets of the MSSM [8] [4]. The L/R indices of supersymmetric scalar particles refers to the handedness of the SM particle.

2.2.6 R-Parity

An extension of the SM with SUSY would introduce terms that violate the barion and lepton number. Such processes have however never been observed. In particular, a violation of barion and lepton number would necessarily lead to a proton decay. In order not to question experimental measurements like the proton lifetime, a new symmetry is introduced, characterised by a new quantum number called R-parity

$$P_R = (-1)^{3(B-L)+2S} \tag{2.20}$$

where L is the lepton number, B the baryon number and S the spin of the particle. This new quantum number is +1 for SM particles and -1 for SUSY particles. This quantum number has to be conserved, therefore at each interaction vertex the number of SUSY particles must be even. This means that supersymmetric particles are always produced in pairs. A further consequence is that the lightest supersymmetric particle, cannot decay in a further lighter SUSY particle and therefore has to be stable. [8]

3 The ATLAS experiment at the LHC

3.1 The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) is a proton-proton collider built 170 m below the French-Swiss border near Geneva. It was build in a circular tunnel of 27 km that was previously constructed for the Large Electron-Positron collider (LEP) which was shut down in 2000.
The LHC started activity in 2008 and the first collision in 2010 started the first phase of data taking, called Run-1 which operated at the centre-of-mass energy $\sqrt{s} = 7$ TeV until 2011 and $\sqrt{s} = 8$ TeV in 2012. After nearly two years of shutdown in 2015, the LHC started operation again providing collisions at $\sqrt{s} = 13$ TeV. The so-called Run-2 lasted until 2018. [9]
Unlike particle-antiparticle colliders, proton-proton colliders need two rings with particles rotating in opposite directions. These two rings have four interaction points at which underground caverns were built to contain one detector each. The four experiments built at these interaction points are ATLAS, CMS, ALICE and LHCb[10].
Near the LHC is positioned a complex of smaller accelerator needed to pre-accelerate protons. These already accelerated protons are then injected in the LHC where they are finally accelerated to reach an energy of 6.5 GeV, making up a collision centre-of-mass energy of 13 TeV. The complete complex of accelerators can be seen in figure 3.1.
The two particle beams go through the ring directed by superconducting magnets with a field of 8 T. Protons are not accelerated singularly, but in bunches. Each proton beam has around 2500 bunches with a spacing of 25 ns and each bunch contains 1.1×10^{11} protons [11].
The high number of protons in a bunch causes collisions in addition to the the primary collision. This phenomenon is called *pile up*.

3.2 The ATLAS detector

The ATLAS detector (A Toroidal LHC Apparatus) is a multipurpose detector as it was not built for one specific measurement. It has a cylindrical form with a length of 46 m and a diameter of 25 m. A schematic view of the ATLAS detector can be seen in figure 3.2.
It is built concentrically around the collision point and built of layers of 6 subdetectors.

3.2.1 Coordinate System

The position of objects in the detector are described with coordinates which are defined with the origin at the collision point. The x-y plane is defined by the plane transversal to the beam, where the positive x-axis points to the centre of the LHC and the y-axis points upwards. The azimuthal angle ϕ is measured around the beamline in this plane. The z-axis is in direction

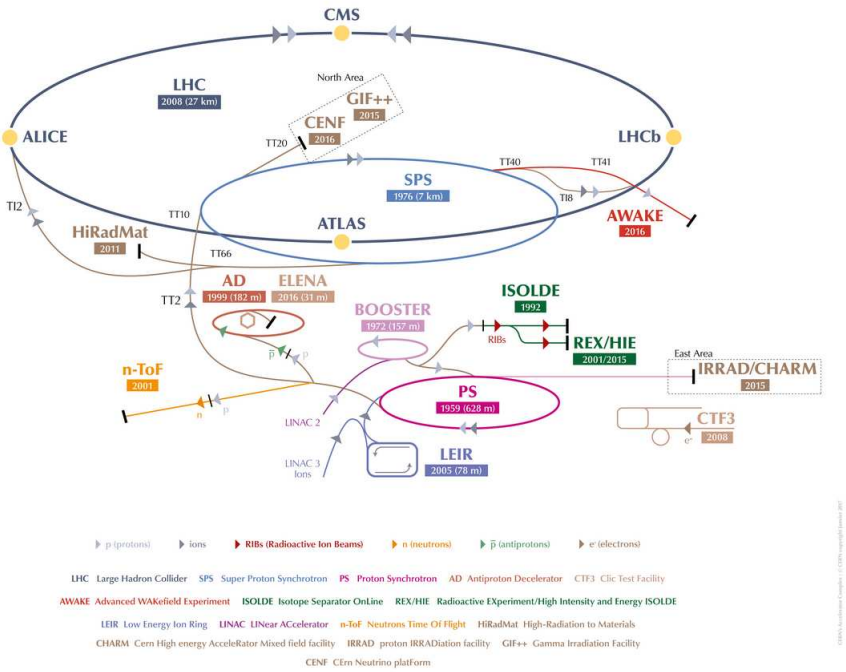


Figure 3.1: Overview of the accelerator complex at CERN [12]

of the beam and the polar angle θ gives the angle from the beam axis. The polar angle is used to calculate the pseudorapidity η[14]:

$$\eta = -\ln \tan(\frac{\theta}{2}) \tag{3.1}$$

This is used to define the distance R used to define object distances:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.2}$$

3.2.2 Subdetectors

The Inner Detector (ID)

The innermost part of the ATLAS detector is called inner detector (ID). It is subdivided in three parts.
Starting from the beam, the first layer of the ID is a silicon pixel detector. With this detector it is possible to make spatial measurements. The high spatial resolution allows for identification of particles with a very short lifetime like tau leptons and top quarks [15].
The following part is a semiconductor tracker (SCT) which consists of four layers of silicon microstrip detectors. These detectors provide measurements in the R-ϕ and z coordinates.
Finally, a transition radiation tracker (TRT) closes the ID. This detector consists of drift tubes (straws) which act as cathodes and are kept in a negative voltage. Layers of radiators are laced between the straws which are filled with a xenon based gas mixture. The energy

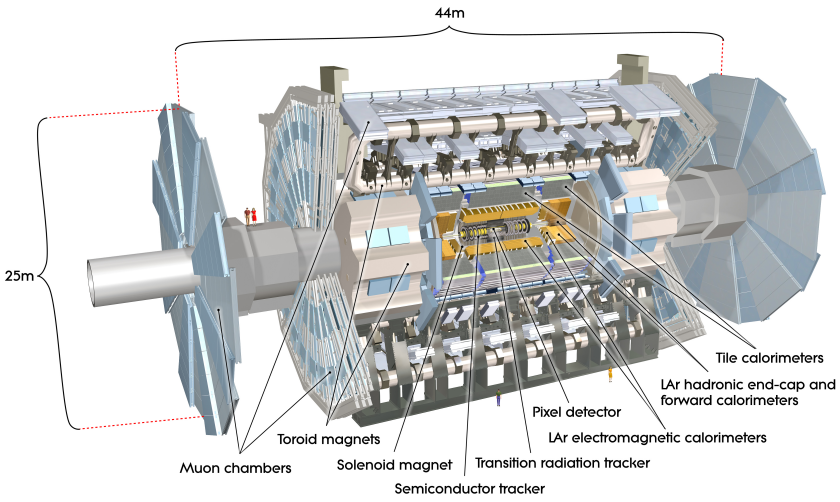


Figure 3.2: Schematic layout of the ATLAS detector [13]

deposits caused by transition radiation are used to identify particles, in particular for separating minimum ionizing particles from electrons.
Since the ID lies in a superconducting magnet which generates a magnetic field of 2 T in the direction parallel to the beam, the observed particle tracks are bent. It is therefore possible to calculate the momentum of a particle from its track curvature.[16]

Calorimeter

The middle part of the inner detector, the calorimeter, measures the energy of some particles particular by stopping them. It is composed of two parts: the electromagnetic calorimeter (ECal) and the hadronic calorimeter (HCal).
The inner part, the Ecal, is made of lead and liquid argon (LAr).[14] In this material photons, electrons and positrons are stopped after undergoing bremsstrahlung and pair creation processes. The number of the particles created by this shower process can be used to calculate the energy of the particle that came from the ID. Muons deposit a small amount of energy in this material but are not stopped by it as they are minimally ionising particles.
Also hadrons deposit a small amount of energy in the ECal but most of their energy is lost in the following HCal where they are completely stopped. For these particles instead of bremsstrahlung and pair production, hadronic interactions take place.

Muon Chamber

The most outer part of the ATLAS detector is the muon chamber (MS). Muons can reach this last detector part as they lost only a minimal part of their energy in the calorimeter. Also here magnets are positioned in order to bend the particle tracks and allow for momentum measurements.

Monitored Drift Tubes (MDTs) provide precision spatial measurements over most of the pseudorapidity plane, while cathode strip chambers (CSC) cover only large pseudorapidities. cosmic radiation can be vetoed.
Moreover a trigger system is present with resistive plate chambers (RPC) in the barrel and thin gap chambers (TGC) in the end-cap region.[14]

3.2.3 Trigger System

Due to the extremely high number of events that happen at the ATLAS interaction point it is not feasible to keep every single one. As not every event contains observations that might be relevant for a physical analysis a system of triggers is used to select relevant events.
The Run-2 trigger system consists of a hardware-based first level trigger (Level-1) and a high level trigger (HLT) based on software.
With the input of the calorimeter and the muon chamber the hardware-based trigger defines regions-of-interest (RoIs) of the detector. This helps to reduce the event rate from 30 MHz to 100 kHz with a decision time of 2.5 μs. Candidates in the RoIs are subdivided in different types: muons (MU), electromagnetic clusters (EM), jets (JET), tau (TAU) as well as sums of missing transverse energy (XE) and total energy (TE).
The selection of the Level-1 trigger is then further refined by the HLF which reduces the event rate to 1 kHz with a processing time of 200 ms. At this stage objects are subdivided in more precise categories which are relevant for physics analyses: electrons, muons, jets, b-jets, missing transverse energy, taus and b-hadrons. [17][18]

4 Object Reconstruction

The particles of the SM show different characteristics in the ATLAS detector and need different reconstruction methods. After their reconstruction additional requirements are applied on particle objects. An overview of the reconstruction method used for specific particle as well as specific object requirements applied for this analysis are explained in the following.

4.1 Jets

With exception of the top quark which decays into a bottom quark due to its high mass, other quarks cannot be identified because they undergo hadronisation. Therefore, they can be recognised in the detector as collimated energy deposits in the calorimeters and are called *jets*.
The reconstruction of jets is done in three steps: definition of the input four-vector, definition of the jet algorithm and parameters and grooming of jets (for large-R jets).
Firstly jets are reconstructed from topo-clusters which are topologically grouped energy clusters in calorimeter cells. These cells must have a high suppression of noise coming from pile-up events. Afterwards, different algorithms can be chosen. In this analysis a so-called anti-k_t algorithm is used. This algorithm can have two different distance parameters R, dividing jets into small-R and large-R jets. Small-R jets usually reconstruct quarks and gluons and have a value of R=0.4. Large-R instead have R=1.0 and represent hadronic decays of massive particles.Finally, since the large value R=1.0 causes a significant amount of pile-up noise, a strategy called *grooming* uses a set of algorithms to reject this background.
In this analysis baseline jets are selected which have $p_T > 20$ GeV and $|\eta| < 2.8$. An anti-k_T algorithm with R=0.4 is used.[19]

4.2 Electrons

To be reconstructed as electrons, candidates have to pass three criteria which characterise the signature of an electron. There must be a localised cluster of energy in the ECal, charged particle tracks have to be observed in the ID and finally both have to match in the $\eta \times \phi$ space
Electromagnetic energy cluster candidates are selected by an algorithm which goes through elements, or *seeds* of the $\eta \times \phi$. The scan starts at a seed with a transverse energy of at least 2.5 GeV and than moves schematically through the space. When two cluster are found to be candidates in elements of close proximity, the one with the highest E_T is selected. The efficiency of this algorithm increases with increasing transverse energy of the clusters.

Tracks are reconstructed by assembling clusters of *hits* in the ID tracking layers. The reconstruction is done by an algorithm which models the particles interaction with the detector material with a pion hypothesis. In the case where it can't be fitted with this hypothesis an electron hypothesis is used.
The electron reconstruction is then concluded by matching candidate tracks to calorimeter seed clusters candidates. If more than one track can be matched to a cluster a selection is done by an algorithm which considers the distance in the $\eta \times \phi$ space of the track and the cluster, as well as the number of hits in the silicon detectors.
Finally the electron identification is done with a likelihood method, where electrons originating from the primary collision (prompt electrons) are seen as signal. Background instead, are particles that stem from jet combinations with a similar signature, as well as electrons coming from photon conversions or heavy flavour decays. To discriminate between signal and background the discriminant d_L is used. This discriminant is subdivided in four values which define four operating points: *VeryLoose*, *Loose*, *Medium* and *Tight*, where objects of the tighter category are also included in the lower ones.
The choice of which working point to use can be made depending on the needs of the physical analysis. In addition to that other requirements can be made to increase the reconstruction efficiency. For this analysis signal electrons were required to have $|\eta| < 2.47$, $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ $\frac{d_0}{\sigma(d_0)} < 5$ and $\Delta z_0 sin(\theta) < 0.5$ mm.[20]

4.3 Muons

Since muons are charged particles, they leave a track in the ID which can be reconstructed similarly to electrons. However, in contrast to electrons, they are minimally ionising particles and can therefore not be characterised by high energy deposits in the calorimeter. They must hence be reconstructed from the information of the muon spectrometer.
Its reconstruction is mainly done by recognising segments formed by hit patterns in each muon chamber. These segments are then reconstructed as tracks with a straight line fit. In particular segments that can align to the tracks in the bending plane are selected. The coordinate orthogonal to the bending plane is measured in the RPC and TGC. An algorithm searches separately in the $\eta \times \phi$ space for segments in the CSC detectors.
When doing a combine reconstruction of the information in the MS and in the ID objects are subdivided in four muon types. When objects are reconstructed separately in the ID and in the MS and eventually fitted together, they are called *combined muons* (CB). Moreover, *segment tagged* (ST) muons are reconstructed from tracks in the ID that can be extrapolated to at least one segment in the MDT or CSC. If a track in the ID can be matched to a low energy deposit in the calorimeter they can be categorised as *calorimeter tagged* (CT) muons. Finally, extrapolated muons (ME) are muons which are reconstructed giving higher priority to the tracks in the MS chambers than to the ones in the ID. Before finalising the muon collection an overlap removal is applied in order to avoid double counting.
Similar to electrons also muon can be divided into four inclusive categories: *Loose*, *Medium*, *Tight* and *Hight-*p_{T}.
For this analysis signal muons are selected. They require $p_T > 25$ GeV, $|\eta| < 2.7$, $\frac{d_0}{\sigma(d_0)} < 3$ and $\Delta z_0 sin(\theta) < 0.5$.[21]

4.4 Tau Leptons

Being the heaviest leptons ($m(\tau) = 1.777\,\text{GeV}$ [22]), tau leptons have to be treated in a particular way. Due to their very short lifetime (2.9×10^{-13}s[22]) they decay before they can be reconstructed. They are therefore characterised by their daughter particles. However, leptonically decaying tau leptons are reconstructed as electrons or muons. In the final tau lepton collection only hadronic final states are included. These hadronic decays can be characterised by the number of charged hadrons, thereby subdividing tau leptons into 1-prong (for example $\tau \to \pi^{\pm}\nu_{\tau}$) or 3-prong taus ($\tau \to 3\pi^{\pm}\nu_{\tau}$), where the majority of these decays include also one or more neutral pions.
Candidates of the visible part of the hadronically decaying tau are reconstructed from jets using the anti-k_t algorithm with the value R=0.4. Moreover, candidates are required to have $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$. The vertex of a tau lepton is defined as the highest p_{T}-weighted candidate within an angular distance of R=0.2 from the seed jet. All the tracks inside a distance of R=0.4 are divided into *core* and *isolation* tracks, where core tracks refer to the number of prongs. This reconstruction makes it difficult to distinguish tau leptons from other jet signatures. To reject such unwanted signatures a recurrent neural network (RNN) classifier is used.
The RNN score is used to define four working points: *VeryLoose*, *Loose*, *Medium* and *Tight*. For this analysis baseline taus were used with the requirements: $p_T > 20$ GeV, $|\eta| \in [0, 1.37] or [1.52, 2.5]$ JetRNNSigTransMin > 0.05 and both 1 and 3 prong taus are accepted.[23]

4.5 Missing Transverse Energy ($E_{\text{T}}^{\text{miss}}$)

Since some particles cannot be detected their energy has to be reconstructed using the energy and momentum conservation. However, in proton-proton collision it is not possible to know the initial longitudinal momentum fraction that a parton is carrying. The energy conservation is therefore used in the transversal plane of the detector. The invisible energy of an event is therefore called missing transverse energy ($E_{\text{T}}^{\text{miss}}$).
Its reconstruction is divided in two contributions. One coming from the *hard event*, therefore including fully reconstructed jets, photons and leptons, and the second coming from the *soft event*, which includes other charged particle tracks.
The missing transverse momentum components are given by negative momentum contributions:

$$E_{x/y}^{m}iss = -\sum_{i\in\{hard\}} p_{x/y,i} - \sum_{j\in\{soft\}} p_{x/y,j} \tag{4.1}$$

however the amount of missing transverse energy is given by the magnitude of the vector $\mathbf{E}_{\mathbf{T}}^{\mathbf{miss}}$ which is positive.[24]

5 Statistical Data Analysis

In SUSY searches, experiments aim to discover a signal process that has been predicted in theory but has not already been observed. Such experiments are based on statistical data analysis and observations are made by counting events (*yields*). For this purpose, it is necessary to conduct an accurate analysis of simulated data before looking at the actual data recorded by the ATLAS detector. The simulated data also called Monte Carlo (MC) data, because of its simulation method, is subdivided in signal and background simulations. MC data allows for an estimation of expected data in the presence of the sought signal. A more detailed explanation of observed and simulated data will follow in chapter 7.

5.1 Statistical Significance

For the counting experiment, one defines a hypothesis H_0 which describes the case where only already known processes are present, so-called background processes. For SUSY searches background is represented by every SM process. The case of the presence of one or more signal processes is given by a hypothesis H_1. To evaluate the outcome of an experiment the agreement of a hypothesis H with the observed data is quantified by a so-called *p-value*. This is a measure for the probability that the background can produce a fluctuation equal to the excess present in the observed data. A hypothesis is then excluded when the p-value is below a certain threshold.
This value is usually converted into an equivalent measure, called significance (Z). Z represents the standard deviations of a gaussian distributed variable above which the probability is equal to the p-value. This relation of Z and the p-value can be seen in figure 5.1.
Mathematically, this is described as

$$Z = \Phi^{-1}(1 - p) \tag{5.1}$$

where Φ^{-1} is the inverse of the cumulative gaussian distribution.
To confirm a discovery in particle physics it is common to reject a background-only hypothesis with a p-value of $p = 2.87 \times 10^{-7}$, corresponding to Z=5. Instead, a signal hypothesis is rejected with a threshold value of p=0.05 which corresponds to Z=1.64. For exclusions, a Z value is usually translated into confidence levels (CL_s). A value of p=0.05 would therefore refer to an exclusion with 95% CL.
The significance value is calculated with the number of expected signal and background events given by MC data. The value can be calculated using different functions. Unless stated otherwise, the significance used in this work is the binomial significance given in *RooStats* [25], which includes the background statistical uncertainty and a flat systematic uncertainty of 30%. [26]

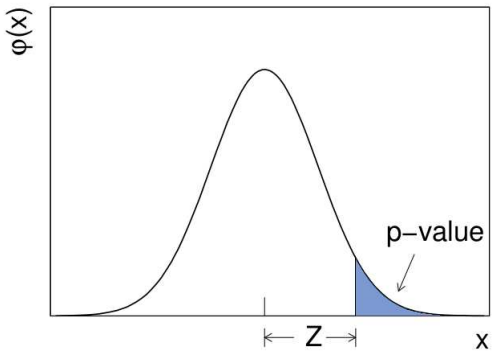


Figure 5.1: Relation between significance Z and the p-value

5.2 Region Definitions

The first step of an analysis is to check if the background estimation is realistic. This is done by looking at the distribution of some variables and check whether the background roughly corresponds to data. To avoid a bias in the analysis it is important to look only at *blinded* data, checking only for events with low signal expectation compared to background.
Once the models are considered to be realistic, they are binned into phase space regions defined by a set of variable requirements. These regions are subdivided in three types: signal, validation and control regions. Each type can then also be composed of one or more bins. Signal regions (SRs) are kinematical regions which are rich of signal events and are defined by a cut selection that maximises the significance. Control regions (CRs) instead, are dominated by background. Usually different backgrounds are divided into different CRs. These regions are used to estimate the background that contaminates SRs by normalising the MC estimation of the SM to data. This estimation is validated with the help of validation regions (VRs), which are usually placed in phase spaces between the SRs and the CRs.
In SUSY searches the particle masses are usually free parameters and the signal cross sections are mass dependent. For an analysis different signal samples with a fixed value for the SUSY particle masses are therefore produced. A signal process with fixed masses is referred to as a signal point. Due to the different masses each signal point has a different kinematic distribution and thereby different signal-background overlaps in several variables. A cut selection can therefore be good for some mass points but not for others. It is therefore best to define different SRs for example distinguish between low and high mass of a SUSY particle.

Once the SRs are defined one can calculate with the signal and background yields which significance could be reached for different points in the phase space spanned by the two particle masses. After extrapolation one can fit the region which includes regions with a significance value of at least 1.64. If no excess over SM expectation can be find, the limits of this region are used for setting exclusion limits for the SUSY particle masses. An example of exclusion limits can be found in the following chapter in figure 6.2.

6 Direct Stau Searches

In order to make SUSY searches more feasible, it is common to study single simplified models which have a reduced number of parameters. In a simplified model usually only a single process with a specific decay chain is considered, assuming the branching ratio to be of 100%. The masses of previously undetected particles are free parameters of the model. The simplified model presented in this work, only considers direct stau production. The superpartners of the two tau lepton chiral states $\widetilde{\tau}_L$ and $\widetilde{\tau}_R$ are considered to be mass degenerate but with different cross sections. A schematic representation of this production and its decay mode can be seen in figure 6.1.

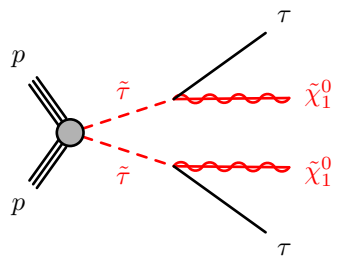


Figure 6.1: Schematic representation of direct stau production and its decay mode

One reason the direct stau production is an interesting model for the LHC is that in some SUSY models the stau is considered to be one of the lightest particles. This fact would make of the stau one of the first SUSY particles that might be produced at the LHC.
After its production, the stau particle decays into a tau lepton and a neutralino, with 100% branching ratio (BR). Almost instantly, the tau lepton then decays further into a neutrino and hadrons or leptons via a W boson. Because of this hadronic or leptonic decay mode, one distinguishes three channels: if both tau-leptons decay hadronically, one talks about the hadronic-hadronic channel (HadHad), while if both decay leptonically it is referred to as leptonic-leptonic channel (LepLep). This study analyses the case where one tau-lepton decays hadronically and the other leptonically, called the leptonic-hadronic channel (LepHad-channel).
The BR are $\frac{4}{9}$ both for HadHad and LepHad and $\frac{1}{9}$ for LepLep.

6.1 Previous Studies with the HadHad channel

An analysis searching for directly produced staus with the ATLAS detector has already been published, including only the HadHad channel [27]. The results presented no excess of data

of SM expectations. Therefore, exclusion limits on the stau mass were set. The results can be seen in figure 6.2. It can be seen that the HadHad channel was not sufficient to cover the complete kinematically available phase space. In particular it left a vast region without sensitivity in a compressed region. With a compressed region it is meant that the mass values of the stau and the neutralino are similar. This region is signed by a red ellipse in figure 6.2. In this work an attempt to enlarge the phase space region with sensitivity obtained in the published paper is presented. For this objective it is studied whether an analysis with the LepHad channel and eventual combination with the HadHad channel would lead to an improvement in sensitivity.

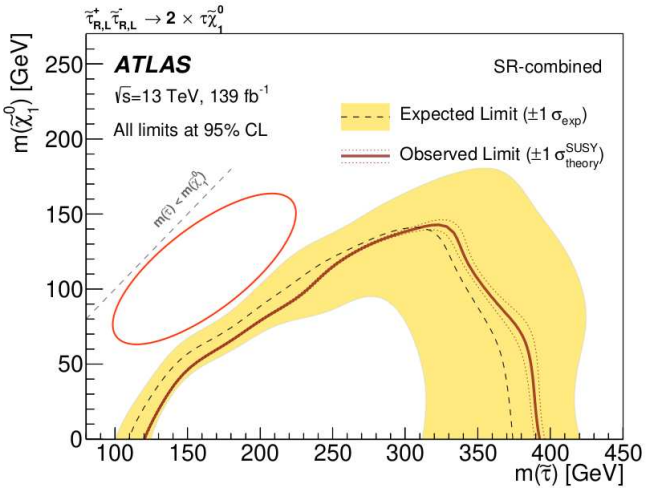


Figure 6.2: Results of the published analysis considering only the HadHad channel [27]

6.2 The LepHad Channel

Despite of its high branching ratio the LepHad channel was not included in a first analysis of the run 2 data, because it presents some difficulties in the separation of signal and background. Firstly, more events should be expected in the $W+$jets background, which is the biggest contribution in direct stau production. As it will be explained in chapter 7, $W+$jets is a background given by a leptonically decaying W-boson and extra jets which are misidentified as hadronically decaying tau leptons. It represents a background for the HadHad channel when the lepton originating from the boson is a hadronically decaying tau. For the case that the W decays into a light lepton (e/μ) or into a leptonically decaying tau it is a source of background for the LepHad channel.
Considering that a W-boson decays into a light lepton with a probability of $\frac{2}{3}$ and to a tau

with a probability of $\frac{1}{3}$, one can calculate the branching ratios of $W+$jets as following:

$$\begin{aligned}
\mathrm{BR(HadHad)} &= \mathrm{BR(W_{lep})} \times \mathrm{BR(W} \to \tau) \times \mathrm{BR}(\tau \to \mathrm{had}) \\
&= \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{27} \\
\mathrm{BR(LepHad)} &= \mathrm{BR(W_{lep})} \times \mathrm{BR(W} \to e/\mu) + \mathrm{BR(W_{lep})} \times \mathrm{BR(W} \to \tau) \times \mathrm{BR}(\tau \to \mathrm{lep}) \\
&= \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{7}{27}
\end{aligned}$$

Therefore, one would expect a much higher $W+$jets contribution with the LepHad channel, compared to the HadHad channel.
In addition to this drawback, it has to be considered that the leptonically decaying tau lepton always adds a neutrino to the final state. This neutrino adds more invisible energy to the event reconstruction. The smearing of the variable distribution that results from this effect makes it even harder to discriminate signal from background.
However, this makes of LepHad a very challenging channel, but still with a high branching ratio. In fact, this channel was not excluded in a search for scalar taus analysed with data taken by the CMS detector located at CERN [28] .
It has to be noted that only hadronically decaying taus are reconstructed by the ATLAS detector as tau leptons. Leptonically decaying tau leptons are always reconstructed as electrons or muons. The two channel would therefore have different trigger sets. The idea of combining the LepHad and the HadHad channel is therefore motivated by the low transverse momentum (p_{T}) thresholds of the lepton triggers. The compressed region where it is attempted to improve the sensitivity is highly populated by events with a low transverse momentum. The high p_{T} thresholds of the τ triggers therefore inevitably cause a lower number of signal events in that region.

The signature of the LepHad channel is made of one tau, one light lepton and missing transverse energy. Its background can be divided in reducible and irreducible background. If a SM signature is completely identical to the one of the signal, its is called *irreducible* background. For the case of the LepHad channel $Z+$jets, Multiboson, Higgs and Top have many irreducible background signatures. If a background shows the same signature, due to one or more misidentified particles it is called *reducible*, because it can be reduced with an improved reconstruction efficiency. $W+$jets and QCD processes have signatures which represent reducible background for the LepHad channel.
The signal and background model of this study is discussed more in detail in chapter 7.

6.2.1 Triggers and Preselection

To study the LepHad channel only events containing exactly one medium tau and one signal lepton are selected. In addition to that, event also have to pass a trigger selection. As already mentioned, the LepHad channel is expected to have more yields in the compressed region of the $\widetilde{\tau}$ and $\widetilde{\chi}_1^0$ mass phase space because of its lower trigger thresholds compared to the tau triggers.
Two types of triggers were used for HadHad studies: the asymmetric ditau trigger and the

	2015	2016-2018
electrons	HLT_e24_lhmedium_L1EM20VH OR HLT_e60_lhmedium OR HLT_e120_lhloose	HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR TrigHLT_e140_lhloose_nod0
muons	HLT_mu20_iloose_L1MU15 OR HLT_mu50	HLT_mu26_ivarmedium OR HLT_mu50

Table 6.1: List of trigger sets for electron and muons for two different data taking periods

	Plateau cut 2015	Plateau cut 2016-2018
electrons	25 GeV 61 GeV 121 GeV	27 GeV 61 GeV 141 GeV
muons	21 GeV 52.5 Gev GeV	25.2 GeV 52.5 GeV

Table 6.2: List of plateau cuts for electron and muons for two different data taking periods

ditau plus $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger. The exact triggers varied depending on the year of data taking, but two examples are:

HLT_**tau80**_medium1_tracktwo_L1TAU60_**tau50**_medium1_tracktwo_L1TAU12
HLT_**tau35**_medium1_tracktwo_**tau25**_medium1_tracktwo_**xe50**

With plateau cuts of $P_T(\tau 1) > 95\,\mathrm{GeV}$ and $P_T(\tau 2) > 65\,\mathrm{GeV}$ for the asymmetric trigger and $P_T(\tau 1) > 50\,\mathrm{GeV}$, $P_T(\tau 2) > 40\,\mathrm{GeV}$ and $E_{\mathrm{T}}^{\mathrm{miss}} > 130\,\mathrm{GeV}$. It can be noticed that low p_{T} thresholds can only be reached with an additional high $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement.
For the LepHad channel instead, it is possible to choose just lepton triggers which have lower thresholds. Different single lepton triggers have been combined with a logical "OR", meaning that at least one of them has to fire to select an event. The lowest offline cuts for these triggers are between 21-27 GeV, much lower than for tau triggers. A list of the used trigger selection for electrons and muons can be found in table 6.1 and the relative plateau cuts in table 6.2.

6.2.2 Variable Definition

From the objects that are reconstructed from the detector, like explained in chapter 4, different variables can be written. Mass, momentum, angles and other measurements are written separately for each object type. In addition, these variables can be used to create more complicated ones that might be very helpful in separating signal from background. The variables that are important for the LepHad channel are presented in the following.

Number of particles

A very important variable is the number of particles. To analyse a precise channel it is necessary to define it by selecting the particles in its final state. For example in the LepHad channel the number of taus N(τ) and the number of leptons N(ℓ) is relevant. For the background discrimination the number of jets N(jets) can be of interest. In particular the number of b-tagged jets N(b-tags) can be very helpful to remove background events coming from events including top quarks.

Transverse Momentum

A variable that is interesting for signal and background separation is the transverse momentum. The transverse momentum of the tau $P_T(\tau)$as well as the one of the lepton $P_T(\ell)$will show to be very relevant for the analysis. In some cases the lepton transverse momentum will be analysed for electrons and muons separately and will be referred to as $P_T(e)$and $P_T(\mu)$. It has to be noted that these are the momentum measurements of what in chapter 4 has been explained to be referred to as tau and leptons. Therefore, these are measurements of only the visible part of a hadronically or leptonically decaying tau leptons.

Transverse Mass

As neutrinos cannot be detected it is impossible to measure the mass of particles that decay to these not observable leptons. However, it is possible to estimate the mass of a particle decaying in one visible and one invisible particle with the information of the missing transverse energy using the equation

$$m_T = \sqrt{2 \cdot E_T^{miss} \cdot p_T(1 - cos(\Delta\phi))} \tag{6.1}$$

where p_T is the momentum of the visible particle and $\Delta\phi$ is the the difference of the angle of the visible particle and the missing transverse energy. This equation is a good estimation only if the invisible particle is massless.
For the LepHad channel the transverse mass of tau leptons and light leptons will be regarded. They will be often plotted singularly as $M_T(\tau)$ and $M_T(\ell)$but also its sum will show to be interesting for the LepHad channel. The variable $M_T(\tau)$ + $M_T(\ell)$will be therefore often referred to as sumM_T.

Invariant Mass

If a particle decays into detectable particles its mass can be calculated with the Lorentz invariant mass given by the relativistic energy momentum relation in the ultrarelativistic limit E $\gg$ m. For the LepHad channel the invariant mass reconstructed with the tau lepton and a light lepton $m_{inv}(\tau, \ell)$is of relevance. However, since neutralinos are involved in this decay $m_{inv}(\tau, \ell)$is not equal to the stau mass. Nonetheless, this variable is very important to remove backgrounds, especially events from $Z +$ jets.

Effective Mass

The effective mass is a very important variable for SUSY searches. It is defined as the sum of the momenta of every visible particles in the event and the missing transverse energy:

$$m_{eff} = \sum_i p_T^i + E_T^{miss} \tag{6.2}$$

Since the neutralino is recorded as $E_{\mathrm{T}}^{\mathrm{miss}}$ signal processes are expected to peak at higher values of m_{eff} than SM backgrounds.

Stransverse Mass

The stau mass cannot be calculated with the transverse mass as it decays into a massive neutralino. Neither the effective mass is a good variable for it, as it counts every particle in the event. Therefore, another variable is used to give a lower bound of $\mathrm{m}(\widetilde{\tau})$. It is called stransverse mass or m_{T2}. It is constructed by assuming that in a SUSY pair production where each slepton decays as

$$\widetilde{l} \rightarrow l\widetilde{\chi} \tag{6.3}$$

the following lower limit can be given to the slepton mass:

$$m_{\widetilde{l}}^2 \geq m_L^2 + M_{\widetilde{\chi}}^2 + 2(E_{Tl}E_{T\widetilde{\chi}} - \mathbf{p}_{Tl} \cdot \mathbf{p}_{T\widetilde{\chi}}) \tag{6.4}$$

If two separate missing transverse momenta, representing the two neutralinos were measurable, then one might calculate the lower limit as

$$m_{\widetilde{l}}^2 \geq max\{m_T^2(\mathbf{p}_{Tl^-}, \mathbf{p}_{T\widetilde{\chi_a}}), m_T^2(\mathbf{p}_{Tl^+}, \mathbf{p}_{T\widetilde{\chi_b}})\} \tag{6.5}$$

However, the missing transverse momentum can be measured only as the sum of the two neutralinos momenta and not for a single one. m_{T2} is therefore calculated by minimising over all possible 2-momenta:

$$m_{\widetilde{l}}^2 \geq m_{T2}^2 \equiv \min_{\mathbf{q_T}}[max\{m_T^2(\mathbf{p}_{Tl^-}, \mathbf{q}_T), m_T^2(\mathbf{p}_{Tl^+}, \mathbf{p_T^{miss}} - \mathbf{q_T})]\} \tag{6.6}$$

For the LepHad channel this variable is calculated from the tau and the lepton m_{T2} (τ, ℓ) and will be referred to just as m_{T2}. [29][27]

Azimuthal Angle

The azimuthal angle is given for every object (for example electronϕ, tauϕ etc.). However, a greater signal-background separation is given from the difference of the angles of two objects. The variables used for the LepHad channel are the difference between the tau and the lepton $\Delta\phi(\tau, \ell)$, the difference between the tau and the missing transverse energy $\Delta\phi(\tau, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}})$and the difference between the lepton and the missing transverse energy $\Delta\phi(\tau, \ell)$.

Pseudorapidity

The pseudorapidity η is also given for every object, but the difference between two objects is more relevant for background discrimination in the LepHad channel. The variable relevant for this work is the difference of pseudorapidity of the tau and the lepton $\Delta\eta(\tau, \ell)$.

Distance between particles

The pseudorapidity and the azimuthal angle differences can be used to measure the difference of two particles in the $\eta \times \phi$ space

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{6.7}$$

This variable can be used to calculate the difference between the tau and the lepton $\Delta R(\tau, \ell)$.

TauRNNJetScore

This variable is the score of the RNN algorithm used to identify taus as described in section 4.4. This variable is helpful to eliminate fake taus. This characteristic is very helpful for the LepHad channel as its greatest background contribution is given by $W+$ jets due to fake tau leptons.

Impact Parameters

To measure the distance of a vertex from the collision point impact parameter are used. The longitudinal impact parameter is give by $z_0 \sin\theta$. A more relevant one is the transverse impact parameter d_0, also expressed as d0-significance when dividing it by its uncertainty (d0sig $= \frac{d_0}{\sigma_{d_0}}$).
The transverse impact parameter significance is often used to veto particles coming from a second vertex. In the following analysis signal electrons are accepted with d0sig< 5 and electrons with d0sig < 3. However, as will be explained in section 9.4, this cut might be disadvantageous for the LepHad channel as the selected lepton is a second-vertex particle coming from the decay of a tau lepton.

7 Data and Monte Carlo Simulation

7.1 Data

The Data used for this analysis has been recorded between 2015 and 2018 and the centre-of-mass energy of Run-2 was 13 TeV. A fraction of the taken event were rejected, as their quality was not good enough to be analysed. The process that selects only suitable events is called *event cleaning*. The data that passes the quality requirements corresponds to an integrated luminosity of $\mathcal{L}$=139fb^{-1}. With the total cross section σ and the luminosity the total number of events taken by the ATLAS detector can be calculated as:

$$N = \sigma\mathcal{L} = \sigma \int \mathcal{L}dt \tag{7.1}$$

7.2 Monte Carlo Event Simulation

MC event generators are widely used in high energy physics. They play an essential role in quantifying the significance of experimental outcomes and interpret the results.
The collision of two protons like it happens at LHC requires a very precise simulation due to the numerous QCD interaction. Therefore, it is subdivided in 5 simulation steps:

- hard process
- parton shower
- hadronisation
- underlying event
- unstable particle decays

Since most of the events leave along the beam pipe direction, only relevant high-momentum processes are simulated. The process with the highest momentum transfer in a collision is called *hard scatter*. The cross section of this type of events is obtained from the corresponding matrix element. Moreover, the colour charged particles involved in the hard scatter radiate gluons. Afterwards, gluons themselves emit other gluons, as they carry a colour charge. This repetition of gluon emission results in a particle shower simulated in the *parton shower* part of MC generation. The principle of confinement which bounds partons into colour-neutral states is taken into account by the *hadronisation* step of the simulation. The initial quarks that did not take part in the hard scatter and take part into secondary interactions are simulated in an additional step, the *underlying event*. Finally, the further decay of many unstable hadrons is taken into account in a last generation step.[30]
In order to make the simulated data comparable with the recorded data a full detector

Process	Generator	Parton Shower
$W \to \ell\nu$ + jets	SHERPA 2.2.1	SHERPA
$Z \to \ell\ell$ + jets	SHERPA 2.2.1	SHERPA
WW/WZ/ZZ	SHERPA 2.2.1/2.2.2	SHERPA
$t\bar{t}$	Powheg-Box v2	Pythia 8.186
Single top	Powheg-Box v2	Pythia 8.186
tt+X	MadGraph5_aMC@NLO 2.2.2	Pythia 8.186
Higgs	Powheg-Box v2	Pythia 8.186
Signal	MadGraph5_aMC@NLO 2.6.2	Pythia8.186

Table 7.1: List of monte carlo generators for different processes

simulation has to be done. With the help of GEANT4 it is possible to safe the information of how the detector sees the data and very importantly, which particles were not correctly reconstructed by it. Simulated data therefore divides information in *truth* (particles that were actually generated) and *reconstructed* (how was the event interpreted by the detector) events. By keeping both information it is not only possible to know which particles were misidentified, but also which one were not detected at all.
To facilitate a direct comparison, simulated events and data are reconstructed with the same algorithms. Subsequently, the number of MC events is normalised to the cross section of the interaction process and than scaled with the luminosity of the considered real data.
Different backgrounds are simulated with different monte carlo generators. An overview of the used generators for each sample can be found in table 7.1. The different backgrounds are discussed more in detail in the next section.

7.3 Samples

7.3.1 Signal Samples

The direct production of a stau pair with $\widetilde{\tau} \to \widetilde{\chi}_1^0\, \tau$ as the only allowed decay mode is called *signal*. The mass of the $\widetilde{\tau}$ and the $\widetilde{\chi}_1^0$ are kept as free parameters of the model. Therefore, several signal sample with different fixed mass values have been produced. A sample with a fix value for these two masses is referred to as *signal point* or *mass point*. Since for the purpose of this thesis only low stau masses are of relevance only signal points until a stau mass of 280 GeV have been regarded. An overview of the available mass points can be seen displayed in a signal grid in figure 7.1

7.3.2 Background Samples

As already mentioned, background events are those events that have the same detector signature as the wanted signal. These signatures can be equal either because of a SM process with the exact same final state (irreducible background) or because of the misidentification of a particle (reducible background). The backgrounds which are relevant for this analysis are grouped into 6 backgrounds: $W+$ jets, $Z+$ jets, Multiboson, Top, Higgs and QCD. In the

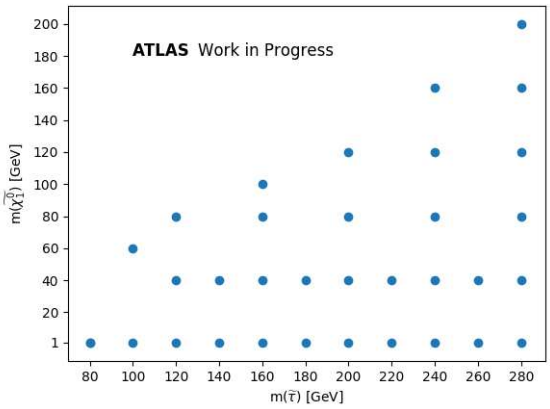


Figure 7.1: Available signal points for a low stau mass

following each background process is presented.

W +jets

This background type includes every process where the production of a leptonically decaying W boson is associated with two jets. For the LepHad channel this process represents a background source when the two jets are misidentified as tau. This can easily happen because of the similarity of the two signatures (see chapter 4). Jets that were incorrectly recognised as tau leptons are referred to as *fake taus*. To reduce the number of fake taus it is better to use taus with a low fake efficiency like for example medium or tight taus (see section 4.4). However, the $W+$jets background is originally produced with a high amount of fake taus. Selecting a low fake efficiency would inevitably lead to a strong reduction of the available raw data and make negative event weights evident.

Z +jets

The $Z+$jets background contains processes with a Z boson decaying into two leptons, together with jets. This process is a source of irreducible background when two tau leptons are produced and one successively decays hadronically and the other leptonically. It could also be a source of reducible background if the jets are misidentified as a tau lepton and the Z boson decays into two light leptons (e/μ), one of which is not reconstructed. A high number of events coming from a Z boson decay can be removed with a cut on the invariant mass ($m_{inv}(\tau, \ell)$, see section 6.2.2), accepting only events above 80-90 GeV. In fact, the mass distribution of a decaying Z bosons should peak at m_Z=91 GeV. However, when two tau leptons are produced, they both decay into hadron or leptons and a neutrino. Due to the impossibility to reconstruct the neutrinos, the invariant mass distribution of a Z boson process should have a peak under the actual Z boson mass value.

Multiboson

In the background Multiboson are grouped processes with more than one electroweak bosons. They can be subdivided in diboson processes, with leptonic or semileptonic decays, and triboson processes. Some of these final states consist of two taus and two neutrinos. If taus decay into the LepHad channel this process is a source of irreducible background. Other decay modes can also be a source of reducible background.

Top Quarks

In the Top backgrounds, many different processes containing at least one top-quark are included. In particular $t\bar{t}$ production is a significant source of background, as it almost instantly decays into a bottom quark and a W boson. The decay of two W bosons can show the same signature of the LepHad channel and it's therefore a source of irreducible background. However, this background can be highly reduced by applying a b-veto.

Higgs

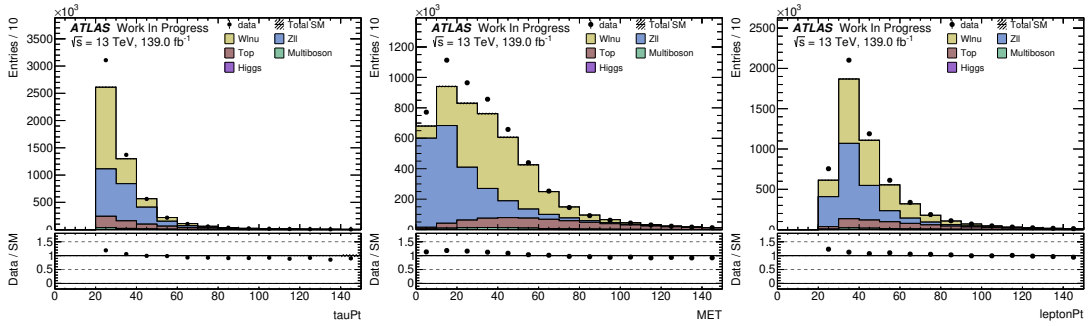
As Higgs processes often include two taus, they are a source of irreducible background. However, since their cross section is very low in comparison to other backgrounds, these processes do not represent a problematic background.

QCD and the Simplified ABCD Method

The contribution of multijet processes (QCD) to the LepHad channel backgrounds is not assured, before a previous blinded comparison of data an SM model expectations. The presence in this channel of one lepton should indeed highly reduce this background. However, as can be seen in figure 7.2 after the LepHad preselection (see section 6.2.1) a comparison of blinded data with the MC processes listed until now revealed a mismodeling which indicates a background understimation , probably due to an exclusion of the QCD background.
As MC simulations of multijet processes are highly inaccurate, QCD estimation are often provided by data-driven methods. In this work a method called simplified ABCD-method is used and presented in the following:

The idea behind the ABCD-method is to produce a QCD sample, by reweighting the MC samples of the other backgrounds, by a comparison with data. A phase space is spanned between to variables of choice which have to be uncorrelated and divide it into six regions. This process should not be confused with the subdivision in regions (explained in section 5.2) when regarding the complete analysis of the LepHad channel. The subdivision used for the QCD estimation in this thesis can be found in 7.3.

One region, rich of signal and poor of background, will be referred to as signal region D (SR-D). In the defined control regions (CR A,B,C), nearly no signal is expected. These regions has to be orthogonal in the phase space and in between them should be defined two validation



(a) **(b)** **(c)**

Figure 7.2: Comparison of (blinded) data and MC expectations without QCD estimations after passing a trigger selection and a preselection of one medium tau and one signal lepton. The distribution of the tau lepton transverse momentum (a) as well as missing transverse energy (b) and the lepton transverse momentum (c) all show a background underestimation.

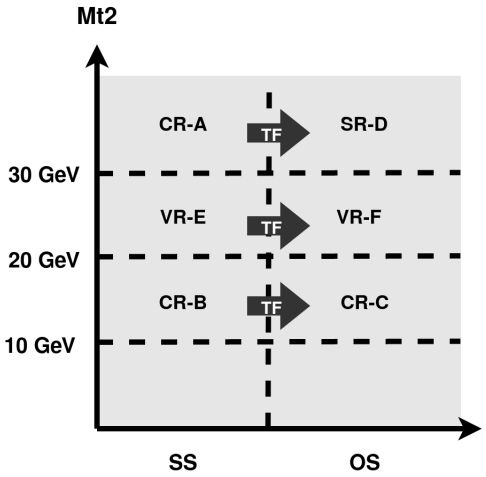
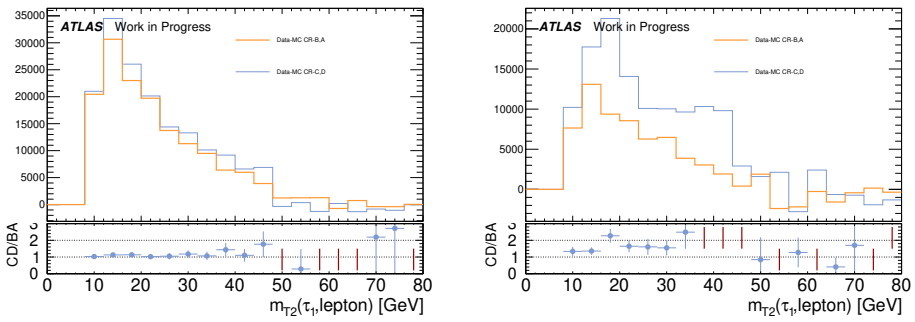


Figure 7.3: Region subdivision in the simplified ABCD-method used for this work. The phase space is spanned by two uncorrelated variables: m_{T2} and the relative sign between the tau and the light lepton (same sign SS and opposite sign OS). Regions are orthogonal to one another.

regions (VR E,F).
For the QCD estimation it is assumed that the difference between data and the MC estimations seen in figure 7.2. are entirely attributable to multijet processes. Therefore, the number of QCD events can be easily calculated from the difference of data and background events in the control regions C and D. With this the number it is then possible to calculate a transfer factor (TF) between the two regions:

$$\mathrm{TF} = \frac{\mathrm{N}^{\mathrm{C}}_{\mathrm{Data}} - \mathrm{N}^{\mathrm{C}}_{\mathrm{MC}}}{\mathrm{N}^{\mathrm{B}}_{\mathrm{Data}} - \mathrm{N}^{\mathrm{B}}_{\mathrm{MC}}} \tag{7.2}$$

Since the control regions are defined to be orthogonal, it can be assumed that CR-C relates



(a) **(b)**

Figure 7.4: Correlation

to CR-B the same way as SR-D relates to CR-A. The same transfer factor can thus be used to calculate the number of QCD events in SR-D:

$$\mathrm{N}^{\mathrm{D}}_{\mathrm{QCD}} = (\mathrm{N}^{\mathrm{A}}_{\mathrm{Data}} - \mathrm{N}^{\mathrm{A}}_{\mathrm{MC}}) \cdot \mathrm{TF} \tag{7.3}$$

When applying the TF to the validation region E it can be calculated how many QCD events are expected in the validation region F. This value can then be compared to the difference in events between data and MC in this region to quantify the uncertainty of the QCD estimation. The ABCD-method used for this work is called *simplified* because it includes only positive event weights in the QCD sample produced in the end. To achieve this, the QCD tree information is written from the data events in the CR-A which all have an event weight with value 1. To ensure a correct estimation of QCD events the tree is written with an event weight which includes a weight factor (WF) which reweights the number of data events to match the difference between data and MC in this region:

$$\mathrm{WF} = \frac{\mathrm{N}^{A}_{\mathrm{Data}} - \mathrm{N}^{A}_{\mathrm{MC}}}{\mathrm{N}^{A}_{\mathrm{Data}}} \tag{7.4}$$

The number of expected QCD events is therefore calculated as:

$$\mathrm{N}^{\mathrm{D}}_{\mathrm{QCD}} = \mathrm{N}^{\mathrm{A}}_{\mathrm{Data}} \cdot \mathrm{WF} \cdot TF \tag{7.5}$$

The two uncorrelated variables chosen for this study are m_{T2} and the relative sign between the selected tau lepton and the light lepton. The uncorrelation of these two variables can be checked with plots that can be seen in figure 7.4. The m_{T2} distribution of the data-MC difference is plotted for the same sign (SS) case (regions A,B) and for the opposite sign (OS) case (regions C,D). It can be confirmed that no significant correlation is present if the ratios of the number of events for OS and SS gives approximately 1 at least at low values for m_{T2}.

Since at least for low m_{T2} value the ratio of the the event number in he two regions does not show a slope. For higher m_{T2} values no reliable statement can be made, as many statistical fluctuations occur.
It has to be noted that the QCD processes relevant for the LepHad channel background estimation, include observations that have mistakenly recognised as leptons (fake leptons). As electrons and muons have different fake rates, the QCD estimation has been done by

splitting the dataset into events including one signal electron and one signal muon. The correlation is therefore also checked separately. Even if the distribution is less precise for muons, no significant correlation is seen since at least for low m_{T2} value the ratio of the the event number in he two regions does not show a slope. For higher m_{T2} values no reliable statement can be made, as many statistical fluctuations occur.
As expected the two transfer factors that resulted from this estimations are different:

$$\mathrm{TF}_e = 1.099$$
$$\mathrm{TF}_\mu = 1.632$$

The number of data and MC events for each region can be found in table 7.2 for electrons and in table 7.3 for muons. The calculated weight factors are:

$$\mathrm{WF}_e = 0.115$$
$$\mathrm{WF}_\mu = 0.016$$

The comparison of expected QCD events and the data-MC difference in the validation region F gave a discrepancy of 0.01 % for electrons and of 0.05 % for muons.

Region	Data	MC	Data-MC	QCD expectation
B	175920	1011825	74095	-
C	424711	343272	81439	-
A	301345	266804	34541	-
D	-	-	-	37964.1
E	132664	93683	38981	-
F	309341	266732	42609	42844

Table 7.2: Number of weighted events for electrons

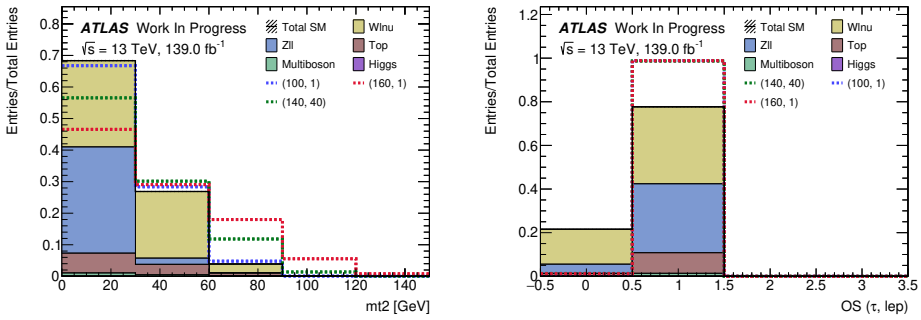
Region	Data	MC	Data-MC	QCD expectation
B	131581	101476	30105	-
C	414737	365594	49143	-
A	343795	338405	5390	-
D	-	-	-	8799.16
E	119261	100130	19131	-
F	310938	281199	29739	31230

Table 7.3: Number of weighted events for muons

7.4 Control of the background estimation

The QCD estimation obtained with the simplified ABCD-method is only valid in the phase space region of $m_{\mathrm{T2}} > 30$ GeV and OS. For this reason these requirements have to be added to the preselection when including the QCD sample. The normalised shape of the distribution of m_{T2} and OS(τ, ℓ) for background and three signal points can be found in figure 7.5.

It can be seen that while the OS cut does not remove any signal event, the m_{T2} cut does remove a big fraction of the signal event, especially for signal points in the compressed region. However, this cut also removes nearly 70 % of the entire background (excluding QCD) and in particular 96 % of the $Z+$jets background. The yields before and after the m_{T2} cut can be seen in table 7.4.



(a) **(b)**

Figure 7.5: Distribution of m_{T2} in (a) and OS(τ, ℓ)(b) after selecting one medium tau and one signal lepton. The value OS(τ, ℓ)= 0 refers to a same sign scenario and OS(τ, ℓ)= 1 to an opposite sign scenario.

Process	Yield	Stat. error	Raw
W+jets	2270025.0	± 11485.4	1748962
Z+jets	2029349.9	± 5593.0	3570645
Top	605016.2	± 289.7	5299115
Multiboson	81782.8	± 107.8	4605503
Higgs	4853.9	± 6.6	1803654
Total SM	4991027.8	± 12778.5	17027879
DS (120,1)	943.5	± 6.4	28389
DS (140,40)	603.0	± 7.8	7102
DS (160,80)	350.1	± 4.5	7544

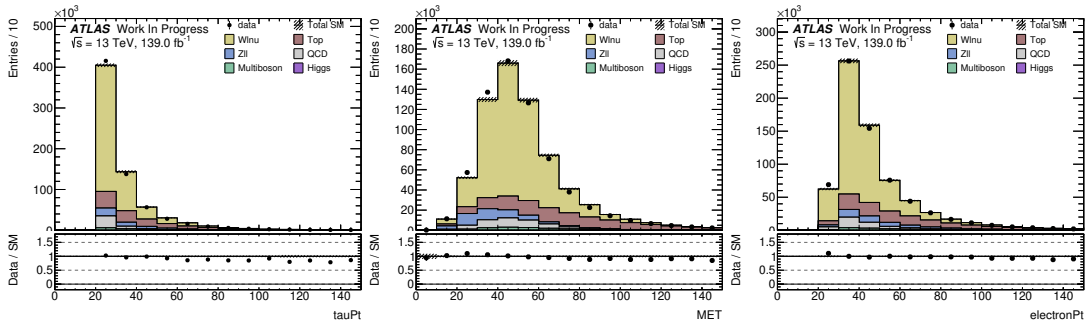
(a)

Process	Yield	Stat. error	Raw
W+jets	1019227.4	± 7815.9	710549
Top	242738.0	± 183.5	2117123
Z+jets	86504.8	± 1146.6	346179
Multiboson	31082.8	± 68.4	1545011
Higgs	323.3	± 1.2	374934
Total SM	1379876.2	± 7902.0	5093796
DS (120,1)	391.5	± 4.0	12087
DS (140,40)	262.2	± 5.1	3163
DS (160,80)	145.9	± 2.9	3221

(b)

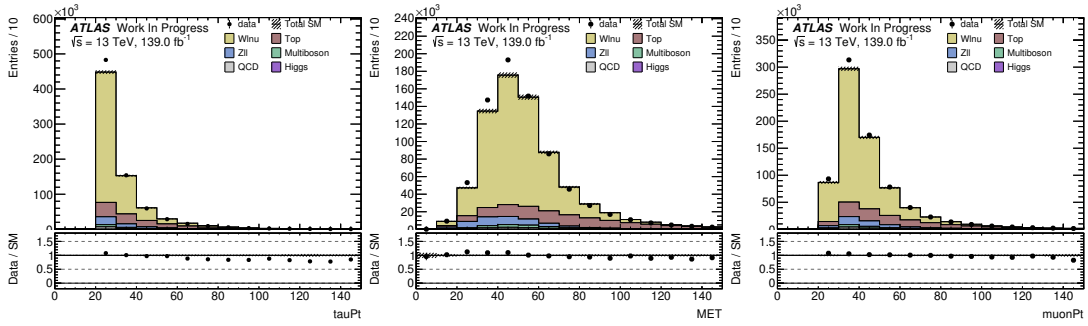
Table 7.4: Number of events after selecting one medium tau and one signal lepton with opposite sign without the m_{T2} cut (a) and with the m_{T2} cut (b)

The comparison of SM expectations and blinded data can be seen in figure 7.6 for electrons and in figure 7.7 for muons. Including the QCD estimation a good agreement between data and MC estimation could be reached.



(a) **(b)** **(c)**

Figure 7.6: Comparison of (blinded) data and MC expectations including QCD estimations for electrons after passing a trigger selection and a preselection of one medium tau and one signal lepton with opposite sign and the cut $m_{\mathrm{T2}}(\ell, \tau) > 30\,\mathrm{GeV}$. The distribution of the tau lepton transverse momentum (a) as well as missing transverse energy (b) and the lepton transverse momentum (c) all show a good agreement of data and SM expectation.



(a) **(b)** **(c)**

Figure 7.7: Comparison of (blinded) data and MC expectations including QCD estimations for muons after passing a trigger selection and a preselection of one medium tau and one signal lepton with opposite sign and the cut $m_{\mathrm{T2}}(\ell, \tau) > 30\,\mathrm{GeV}$. The distribution of the tau lepton transverse momentum (a) as well as missing transverse energy (b) and the lepton transverse momentum (c) all show a good agreement of data and SM expectation.

8 Cut and Count method

The first attempt to define a signal region is done with the socalled Cut-And-Count method. The principle of this method is to define the set of upper or lower requirements for specific variables additionally to the preselection in section 6.2.1 that maximises the significance. As reaching high significances in SUSY searches often requires very strict requirements such a method might lead to an optimisation based on statistical fluctuations due to the finite size of the MC samples. To avoid this effect that would lead to unreliable significances, some requirements have to be fulfilled. The number of background events which are left after applying the selected cuts has to be of at least two and the statistical uncertainty of the background cannot exceed 50%. Additionally, the number of events of each background subsample cannot be negative. A negative number of events might occur due to negative event weights generated for example in Sherpa. For those samples which show a negative number of events, after a cut selection the number is set to 0.
It has to be noted that since every signal point shows different kinematic distributions, the optimisation of the cut selection can be done only for a single signal point. Once an optimal cut combination has been found it is applied to every signal sample, as well as the background samples to finally calculate the significance reached at each signal point.
The combination of variables and cuts that leads to the highest significance can be found in different ways. In the following two methods are presented. The first using a multidimensional cut scan and the second presenting a manual selection of variables and cuts with the help of N-1 plots.

8.1 Multidimensional Cut Scan

A multidimensional cut scan can be done with the help of an algorithm. In this study the scan has been done with the publicly available *ahoi* framework [31]. This algorithm is able to combine different cuts on several variables and calculate the significance for each combination to eventually select the one which delivers the best results for the given sample. Some possible cut values have to be given as input to the algorithm, as well as the variables to be scanned. To keep computation time low, these input values have to be chosen carefully, by checking the overlaps of the variable distributions of signal and background. The shapes of the distribution is represented by the normalised number of events of signal and background. To have a better understanding of the background, the entries of each type are divided by the number of total background events. In figure 8.1 one can see two examples of shape plots for the variables $\mathrm{P_T}(\tau)$and $E_{\mathrm{T}}^{\mathrm{miss}}$.

To have a better overview of the several cut combinations and the relative significance, each one is binned into values of signal efficiencies and only the ones with the highest background rejection are kept. For each signal efficiency the significance and the background rejection (one minus background efficiency) can be calculated. Two plots give the user a

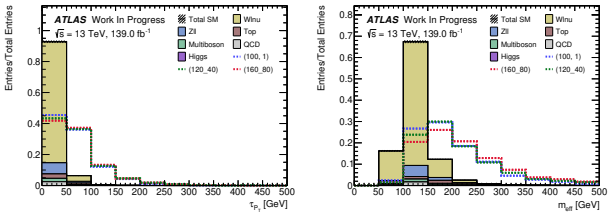


Figure 8.1: Example of shape-plots

better overview over these values. The first one shows the significance for each bin of signal efficiency. For a better understanding, different significances and or different uncertainties can be plotted at the same time. The second plot shows the background rejection reached for each signal efficiency. This is called Receiving Operating Characteristic-curve and in the ideal case of 100% background rejection for 100% signal efficiency it would be rectangular with a 90° angle at the point (1,1).
As already mentioned a multidimensional scan might lead to an unreliable optimisation if based on statistical fluctuation. In order to understand whether such a situation is occurring, a fraction of the input data is kept as *test*. After the scan, two significances, as well as background rejections from two statistically unrelated datasets can be calculated. A discrepancy of the two values would point to an optimisation on statistical fluctuations. This can be easily checked optically for the significance plot and the ROC-Curve.

8.1.1 Results of Multidimensional Cut Scan

Before choosing the possible cut values for the scan it has been decided to add a b-veto to the preselection presented in section 6.2.1. As can be seen in figure 8.2 this cut removes a high fraction of the Top backgrounds loosing nearly no signal. The working point of the tau reconstruction has also been changed from medium to tight, as this help to remove several $W+$ jets events even if it causes a loss of 20 % of the background (see figure 8.2).

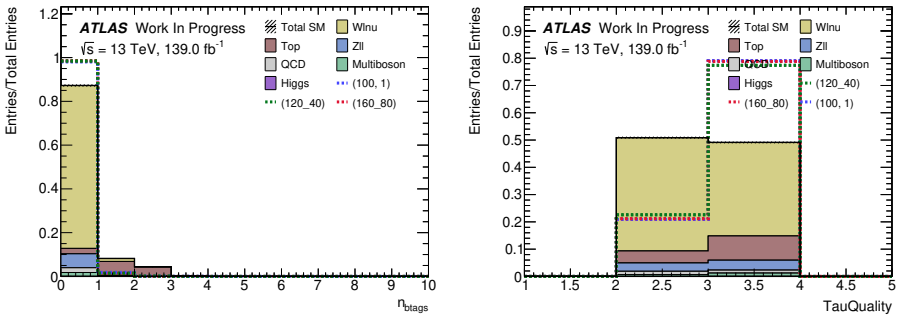


Figure 8.2: Shapes of tau Quality and N(b-tags)

Since the aim of this work is to reach sensitivity at a compressed mass scenario, the signal point that has been chosen to optimise with is for m($\widetilde{\tau}$) = 140 GeV and m($\widetilde{\chi}^0$) = 40 GeV. After looking at the shape plots after this tighter preselection (see appendix) a set of possible

cuts has been given as input to the algorithm. This selection can be seen in table 8.1.

It is not necessary to reach significances of at least 1.6, since the LepHad channel is not

variable	separator	cut-value
$\mathrm{P_T}(\tau)$	$>$	[**0**,50,100,150]
$E_{\mathrm{T}}^{\mathrm{miss}}$	$>$	[0,**50**,100]
$\mathrm{M_T}(\tau)$	$>$	[0,**50**,100,150]
$\Delta R(\tau,\ell)$	$>$	[0,1,**2**]
$\mathrm{M_T}(\ell)$	$>$	[0,50,**100**]
$\Delta\eta(\tau,\ell)$	$<$	[**2**,6]
$\Delta\phi(\tau,\ell)$	$>$	[0,**1**,2, 2.5]
N(jets)	$<$	[**1**,2,3,10]
$\mathrm{m_{inv}}(\tau,\ell)$	$>$	[0,50,**100**]
m_{eff}	$>$	[50,100,**150**,200]
$\mathrm{M_T}(\tau)+\mathrm{M_T}(\ell)$	$>$	[50,100,150,200]
d0sig(ℓ)	$>$	[0,0.5,**1**]

Table 8.1: Possible cut values and variables given as input in the ahoi algorithm

meant to reach sensitivity by itself, but to improve the significances of the HadHad channel. However, in order to combine the significances as will be explained in section 11.1 at least positive significances are required.

The significance plot in figure 8.3 shows that it is very difficult to get even a positive significance for the LepHad channel, especially when including a dummy systematic uncertainty of 30 %. In addition, the ROC-Curve in figure 8.4 shows that a high background rejection can only be reached with a very low signal efficiency, which usually leads to very low significances. As can be inferred from both plots there is a good agreement between test value and the calculated values. However looking at the significance plot it can be noticed that at low signal efficiencies the disagreement becomes higher.

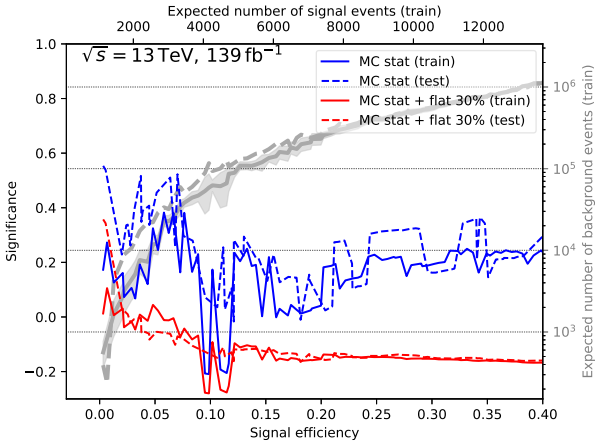


Figure 8.3: Significance plot done with ahoi

In order to attempt a better optimisation, the cut combination given for a signal efficiency of

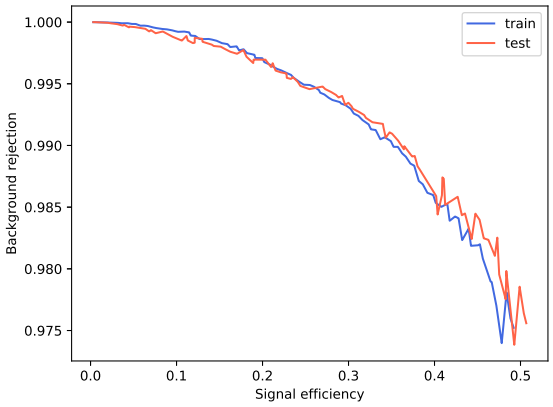


Figure 8.4: ROC-curve done with ahoi

0.06 was selected as starting point for a second attempt explained in the next sections. This efficiency has been chosen because at low values the highest significances can be found, but it is still high enough not to show a big disagreement from the test sample. The selected cut combination can be seen from the values in bold in table 8.1. The variable sumMt has not been considered, since including each cut value of this variable would have lead to the same significance.

8.2 N-1 Plots and Shape Plots

A possible method to improve a preselection is to use N-1 plots. These are plots of all the variable used in the preselection with every cut applied except the one on the plotted variable. This is useful to check whether a single cut is necessary and also if it might be more reasonable to lower or raise a cut value. For the last purpose a significance plot is often put underneath the variable distribution in order to check how the significance behaves depending on the value. In figure 8.5 can be seen a N-1 plot as an example.

After adjusting the preselection given from the multidimensional scan shape plots of other variables have been checked in order to find possible new cuts. As the cut selection is already very tight, this last step is only necessary to find possible cuts which remove a small number of background events but nearly no signal.

8.3 Results

The cut selection that resulted from this last step was:

- $M_T(\tau) > 50$ GeV
- $M_T(\ell) > 100$ GeV

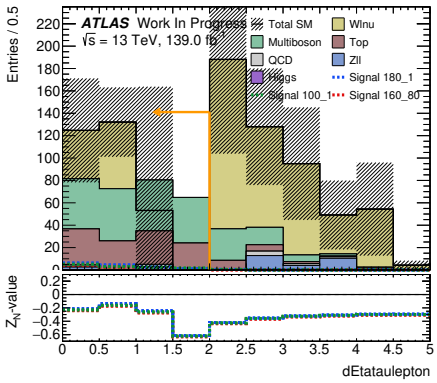


Figure 8.5: Example of N-1 plot for the variable $\Delta\eta(\tau, \ell)$with the preselection given from the multidimensional scan.

- $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV
- N(jets) < 2
- $\mathrm{m_{inv}}(\tau, \ell)>$ 100 GeV
- $m_{\mathrm{eff}} > 150$ GeV
- $\Delta\eta(\tau, \ell)< 2$
- $d_0$sig(ℓ) ¿1
- tauRNNJetScore¿0.97
- $m_{\mathrm{T2}} > 40$ GeV

The number of events that are left after these cuts can be inferred in table 8.2. As the total background statistical uncertainty was already of 40 % no more cuts could be applied. However, this region definition provides no sensitivity as for every signal point available a negative significance is calculated.

Process	**Yield**	**Stat. error**	**Raw**
Top	394.7	± 7.6	3051
Multiboson	297.8	± 6.2	12902
Wlnu	64.5	± 333.7	896
Zll	30.9	± 22.6	161
QCD	17.7	± 1.4	240
Higgs	0.3	± 0.1	152
Total SM	805.9	± 334.6	17402
StauStau.140p0$_4$0p0	21.1	± 1.4	253
StauStau.100p0$_1p$0	20.3	± 1.6	187
StauStau.180p0$_1p$0	20.3	± 1.7	161

Table 8.2: Yield table for the cut selection resulting from the Cut-And-Count method.

9 Multivariate Analysis (MVA)

Since the cut-and-count method gives insufficient results, a multivariate approach has been tried in order to discriminate the LepHad channel from its corresponding SM background. This method allows to combine different discriminating variables into one variable (MVAoutput score) using Boosted Decision Trees (BDTs). This is a concept that comes from supervised learning (sub-discipline of machine learning), which represents the capability of an algorithm to learn from labelled input data to make predictions for unlabelled data. [32]
In the following the BDT method is explained and different results are presented.

9.1 Theory

9.1.1 Boosted Decision Trees (BDT)

With simulated samples one always knows which class (signal or background) an event belongs to. Each sample contains therefore a set of events $(y_1, \mathbf{x}_1)...(y_n, \mathbf{x}_n)$ with the information x_i of the kinematic variables and the information y of the sample class. In the case of this study this last information is given by the variable SUSYFinalState which has a value of 0 for each background and a value greater than zero for the signal samples.
Since recorded data is not labelled, aim of the BDT is to learn how to predict the class of unclassified data with the help of a labelled training sample by developing a prediction rule to correctly categorise a new observation. [32]
The method of BDTs is not that far from the previously presented method. In the cut-and-count method the kinematic variable with the best separation power is chosen and a cut is applied on it. Afterwards the next best variable is chosen and this procedure is repeated until no more cuts can be applied, due to a too small number of events.
A single decision tree is very similar to this method. A yes/no decision for one variable at a time is repeated until a discrimination criterion is fulfilled. This stopping criterion is usually reached when the phase space has been reduced to a signal-like or a background-like region. [33] This procedure leads to a structure like the one seen in figure 9.1.
An algorithm that compares the overlaps of the distributions of signal and background and chooses the best variable to apply a cut on might have a high discrimination power but it is also very sensitive to statistical fluctuations which can lead to complete different tree structures. To avoid this issue multiple decision trees can be combined together to form so-called *forests*.[33]
Single decision trees in a forest which are combined together with a process called *boosting* are called Boosted Decision Trees. *Boosting* is a way of improving the classification performance of an MVA-method by iteratively pushing the focus of the classifier on problematic observations.[32] The method used for shifting this focus depends on the type of BDT.
Finally a total output of the forest is written in a variable called output score. One can then

apply a cut on this score to discriminate signal from background. If the BDT works correctly this variable should have a high separation power, showing opposite peaks of the signal and background distribution.

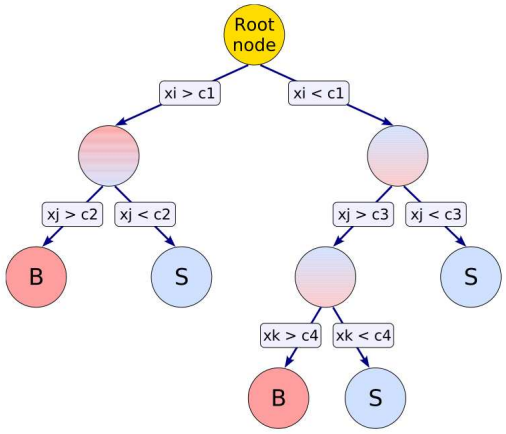


Figure 9.1: Schematic view of a single decision tree [33]

Boosting methods

The two most known methods of boosting are *AdaBoost* and *GradientBoost*. As mentioned before, these two types of boosting differ in how they give observations with a weak discrimination power more relevance in the following iteration.
AdaBoost is called like this because it adapts its parameters based on the performance of the current iteration.[32]
This means that events that were misclassified from a decision tree receive a higher weight in the training of the following tree. Therefore, the following tree is trained with a modified event sample where the actual event weights are multiplied by a *boost weight* α, derived by the misclassification rate *err*.

$$\alpha = \frac{1 - \text{err}}{\text{err}} \tag{9.1}$$

After this step, the weights of the complete samples are then renormalised in order to keep a constant sum of weights. The boosted event classification is so given by:

$$y_{Boost}(x) = \frac{1}{N_{trees}} \cdot \sum_{i}^{N_{trees}} \ln(\alpha_i) \cdot h_i(x) \tag{9.2}$$

where $h_i(x)$ is the result of a classifier which is encoded as -1 for background and +1 for signal and N is the number of combined decision trees. A small value of $y_{Boost}(x)$ represents a background-like event, while a large one a signal-like event.
If we represent the model response by the function F(x) which is a sum of parameterised base functions:

$$F(x; P) = \sum_{m=0}^{M} \beta_m f(x; a_m); \qquad P \in \{\beta_m; a_m\}_0^M \tag{9.3}$$

aim of the boosting procedure is to adjust the parameters P minimising the deviation between F(x) and and the true value y. This deviation is calculated with a *loss function* L(F,y).
AdaBoost is based on an exponential loss function

$$L(F, y) = e^{-F(x)y} \tag{9.4}$$

which lacks reliability with outliers, making this algorithm very sensitive to negative or very high event weights.
A second boosting method called *Gradient Boost* avoids the exponential function. In fact, the algorithm used by GradientBoost tries to overcome the shortcomings originating from the exponential function by classifying events with functions like for example the binomial log-likelihood loss function:

$$L(F, y) = \ln\left(1 + e^{-2F(\mathbf{x})y}\right) \tag{9.5}$$

The minimisation of this function is done by matching the leaf values of the decision tree with the calculated current gradient of the loss function.[33]
GradientBoost is already less inclined to overtraining than *AdaBoost*, but one can also enhance its robustness with a hyperparameter called *learning rate*.
To understand this hyperparameter one can represent the model response depending on the previous iteration:

$$F(x)_m = F(x)_{m-1} + h_m(x) \tag{9.6}$$

where $h_m(x)$ is the function that has to be fitted in order to minimise the loss function. This last contribution can be scaled with a new parameter ν which is called *learning rate*.

$$F(x)_m = F(x)_{m-1} + \nu h_m(x) \tag{9.7}$$

This hyperparameter should be adjusted together with the parameters that controls the number of estimators, i.e. the number of trees. Small values of learning rate require a high number of trees to maintain a constant training error [34].

9.1.2 Technical Settings

When training with boosted decision trees one can decide between different options for certain input settings to reach an efficient training.
As already mentioned it is very important to avoid overtraining, as this would reduce the class prediction of unlabelled data. At the same time one aims at an output score with possibly opposite signal and background shapes. Depending on the input files it can be more or less difficult to reach a good separation. To improve the results one should find the most suitable input settings.
Firstly the choice of boosting algorithm can give very different results. In this work only GradientBoost has been used, since AdaBoost showed unsatisfying results.
As already mentioned, the output of a BDT is a variable that is here called *MVA output score* which should serve as variable with very high discriminating power. To profit from this variable one should evaluate every sample that should be optimised. Evaluating therefore means just writing an extra variable to a sample. To evaluate a sample it is not necessary to use it also as input data for the training.

A first setting which is very important for BDTs is therefore the input data. For evaluation one can take as many signal samples as wanted, as these are not changing the training. The choice of input data instead, is very important. Signal samples differ in the masses of staus and neutralinos, thereby showing different kinematic distributions. Including too many signal samples might lead to a less efficient training caused by a not characteristic signal distribution. However, when training with simulated data one always has to take into account the available amount of statistics. A too low number of raw events could rapidly lead to overtraining, forcing one to choose hyperparameters that are not sufficient to separate signal from background. In case of low statistics one should include different signal points, thereby trying to include points with similar kinematic distributions. It is therefore very important to find a good equilibrium between number of raw events and the shapes of the signal distributions.
As already explained Overtraining means that the algorithm creates a prediction rule based on fine aspects of the input data. This leads to a perfect fit of the input data but also to a very poor prediction of unlabelled data. In order to handle overtraining, the input samples are split into a training and a test sample. In the framework used for this training 80% of the input data is used for training and 20% for testing. Afterwards the training the BDT response is compared to the test sample which has not been trained and is therefore uncorrelated. If the results of the training sample corresponds to the testing sample no overtraining occurs. Therefore, overtraining can be very easily checked optically by plotting training and testing samples together. As already mentioned, one of the output plots of the BDT is the MVA output scores which shows the BDT response to the training. In figure 9.2 one can see an example of an output score of a highly overtrained sample. The shape of this output score is what one aims to, but due to the high overtraining this result can not be accepted.

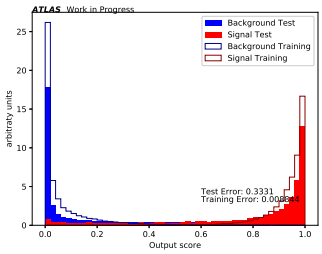


Figure 9.2: Example of an MVA output score distribution with high overtraining

A second output plot one usually looks at is the Receiving Operating Characteristic Curve (ROC curve). This is a representation of the signal efficiency that can be reached for a certain background rejection (i.e 1 minus background efficiency). If the algorithm can perfectly discriminate signal from background the ROC curve would be rectangular, with an angle at the point (1,1), which means 100% signal efficiently for a complete background rejection. In figure 9.3 the curve for a highly overtrained sample is depicted.

A further setting of the BDT is the preselection. The higher the number of training data, the better the training result will be. However, some cuts can be applied in order to reduce the computational time. A good strategy could therefore be to apply cuts on regions that contain only background events and nearly no signal events. Furthermore, depending on the

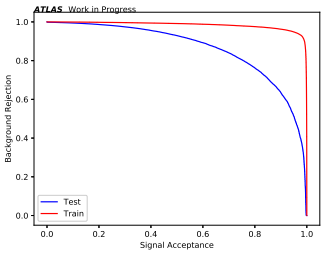


Figure 9.3: Example of a ROC curve with high overtraining

input data, it might make sense to apply some cut to reduce obvious separation of signal and background distributions, to let the BDT learn on non evident signal and background overlaps. These choices has to be made always taking into account that a too low number of input events would rapidly lead to overtraining. For some samples it might be convenient to keep the preselection as loose as possible to improve the discriminating ability of the algorithm and eventually apply a tighter preselection on the evaluated sample. However, it would be of benefit to apply before the training all the cuts which are absolutely necessary. In fact, depending on the strictness of the cut, it could significantly change the kinematic distribution of the variables which would result in a weaker discrimination power of the output score.
Finally, one should also study which kinematic variables to give as input to the BDT. In general, one can choose every variable as input, but it should be empirically tested if there are some combinations of kinematic variables that work better than others. In order to know which variables have the best discriminating power, one can check the ranking of the variables. The ranking is done by counting how many times the variables are chosen to split decision tree nodes and by weighting it with the separation gain of this split [33].
Except the pre-training BDT settings, there are also the already mentioned hyperparameters one can tune to reach the best response as possible. The most important parameters are:

- the maximal depth of the tree
- the number of estimators (trees)
- the learning rate

Each of this parameter should be tuned in order to avoid overtraining, but also trying not to *underfit* the data. For example in this study the maximal depth was chosen to have at least a value of 3. To keep overtraining under control the learning rate is a very important parameter. This parameter can have a value between 0 and 1 and in this study it was kept between 0.01 (for high overtraining) and 0.1 as recommended in the Scikit-learn guide [34]. This parameter strongly interacts with the number of trees. In fact, a small learning rate typically demands a higher number of estimators.[33]
When the samples one is training with are not equal, one can choose to *balance* the dataset. In SUSY searches one has to train with a much smaller number of events of signal compared to background. When choosing the option *balancing*, only a fraction of the background event is taken for training, in order to train with a similar number of events of signal and background.

9.2 Optimisation with different preselections

To get an idea on how the BDT performs for the LepHad channel it was firstly trained with adding only a few cuts to the preselection presented in 6.2.1, coming to the following preselection:

- one tight tau
- one signal lepton with opposite sign to the tau lepton
- veto on b-tagged jets (b-veto)
- $m_{\mathrm{T2}}(\ell, \tau) > 30\,\mathrm{GeV}$

The last cut had to be applied for the validity of the QCD estimation. This preselection will be referred to as *loose preselection* throughout this section.
Signal samples with stau masses between 80-180 GeV and neutralino masses between 1-100 GeV (see signal grid in figure 7.1) as well as all the six backgrounds have been used as input data. In table 9.1 can be found the number of raw and weighted events after this preselection with the sum of every signal input sample defined as *total signal*. Every point of the signal grid in 7.1 and each background sample has been evaluated after the training. With this

Process	**Yield**	**Stat. error**	**Raw**
W+jets	448022.9	± 5099.3	275260
Z+jets	48840.4	± 877.4	176231
Top	29821.6	± 65.2	253342
QCD	20748.8	± 47.1	286846
Multiboson	20018.0	± 52.3	1066184
Higgs	96.2	± 0.9	36441
Total SM	567547.9	± 5175.2	2094304
Total signal	2705.3	± 17.0	42674
DS (120,1)	303.5	± 3.5	9418
DS (140,40)	208.0	± 4.5	2523
DS (160,80)	113.9	± 2.5	2526

Table 9.1: Number of raw and weighted events (yield) with statistical error after adding a b-veto and the cut $m_{\mathrm{T2}} > 30$ GeV to the preselection in 6.2.1

preselection it was possible to try different boosting algorithms and check which gives the best performance.
In figures 9.4, 9.5, 9.7 and 9.6 one can see the output score, as well as the ROC curve and the variable ranking of a training with Adaboost or Gradient Boost both with and without event balancing. Gradient Boost shows a better result than Adaboost and event balancing shows an improved result in both cases. However, the output score still does not show the expected result with a background peak at 0 and a signal peak at 1. It can be noticed that the type of decision tree and the balancing both have an influence on the ranking of the variables. Which variable works best for the training can therefore not be seen easily from shape plots, but it has to be checked with this type of ranking plot.
Since training with Gradient Boost together with a balanced dataset gives the best result, this setup will be used for all the following trainings.

Even though the output score and the ROC curve show a great separation of signal and background when evaluating the files one sees that the BDT does not manage to separate

signal and background enough to get positive significances. This is probably due to the overwhelming number of background events in comparison to the single signal events. It is therefore reasonable to tighten the preselection in order to avoid that the BDT trains on obvious aspects of the distribution of input variables and can concentrate on less obvious ones. In the following three different preselections are shown.

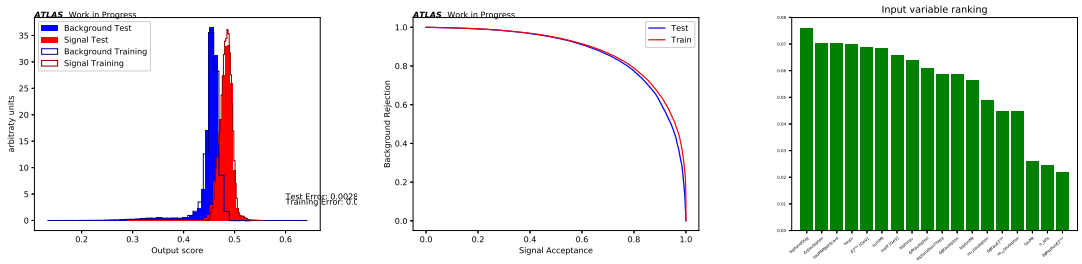


Figure 9.4: Output score, ROC curve and variable ranking for a training with AdaBoost decision tree without event balancing

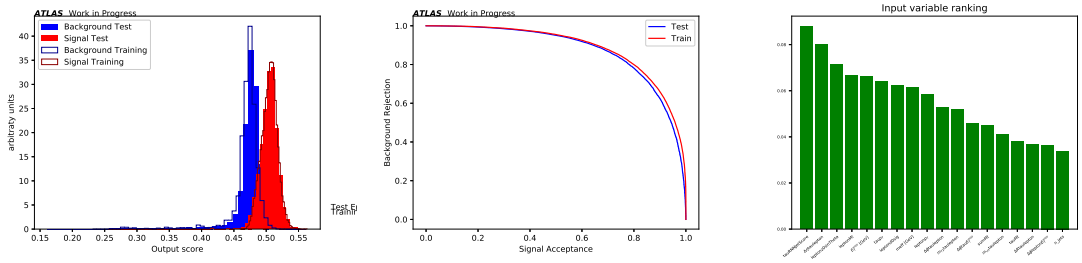


Figure 9.5: Output score, ROC curve and variable ranking for a training with AdaBoost decision tree with event balancing

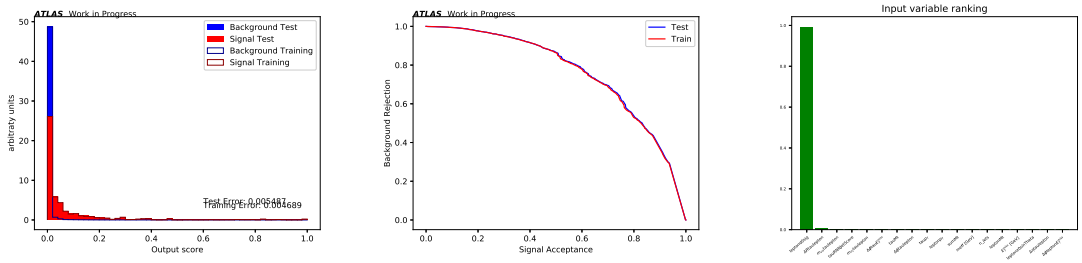


Figure 9.6: Output score, ROC curve and variable ranking for a training with Gradient-Boost decision tree without event balancing

Preselection oriented to CMS paper

A Direct Stau analysis including the LepHad channel has been published by the CMS experiment (see [28]). As a first attempt its preselection has been tried for this study.
It had as pre-training selection following cuts:

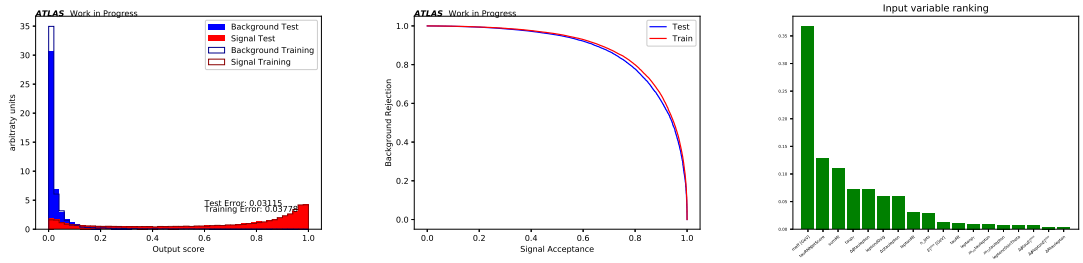


Figure 9.7: Output score, ROC curve and variable ranking for a training with Gradient-Boost decision tree with event balancing

- jetPt $< 20\,\text{GeV}$
- $20\,\text{GeV} < \text{M}_\text{T}(\ell) < 60\,\text{GeV}$ or $\text{M}_\text{T}(\ell) > 120\,\text{GeV}$
- $\text{m}_\text{inv}(\tau, \ell) > 50\,\text{GeV}$
- $2.0 < \Delta R(\tau, \ell) < 4.0$

Since with the current object definition jets are accepted only with transverse momentum greater than 20 GeV, this cut is equal to the cut N(jets)=0.
A training with these cut in addition to the loose preselection resulted in an insufficient separation of signal and background as it delivered negative significances for each evaluated signal point. However, looking at N-1 plots (figure 9.8 of this preselection it shows to be not adequate for the object definition of this analysis.

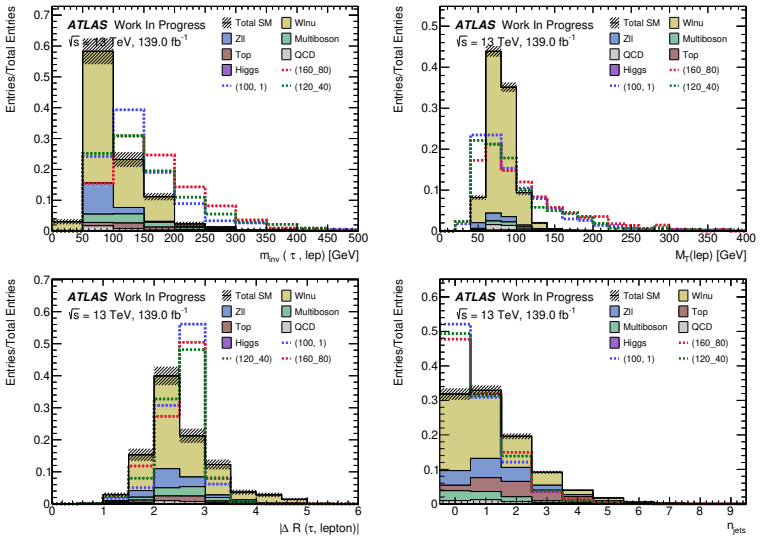


Figure 9.8: N-1 plots for the CMS preselection

Therefore, an adapted preselection is attempted:

- N(jets) < 2
- $\text{M}_\text{T}(\ell) < 60\,\text{GeV}$ or $\text{M}_\text{T}(\ell) > 120\,\text{GeV}$

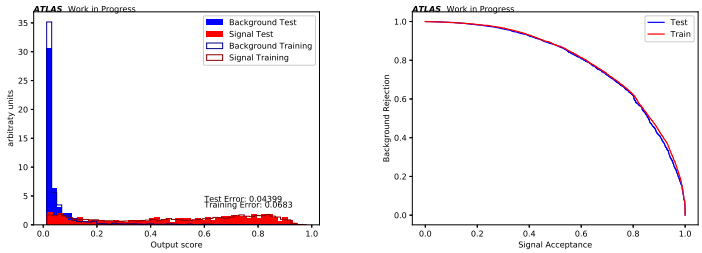


Figure 9.9: Output score and ROC curve for the training with the CMS preselection. The hyperparameter user for this training were maximal depth: 3, learning rate: 0.01, number of estimators: 400

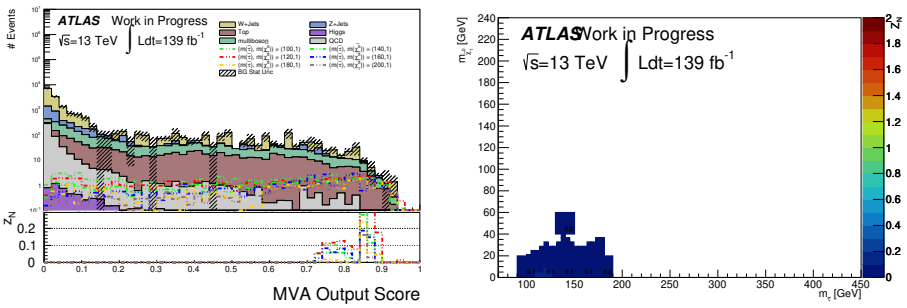


Figure 9.10: MVA output score and significance plot of the training with the CMS preselection. The significance plot has been done with the cut MVAoutput score > 0.76

- $m_{inv}(\tau, \ell) > 100\,\text{GeV}$
- $\Delta R(\tau, \ell) < 4.0$

The new N-1 plots can be seen in figure 9.11.
In figure 9.12 the training plots of a training with this preselection can be seen and the MVA output score and the significance plot in figure 9.13.
Interesting is that the output score in figure 9.12 shows a nearly flat distribution of signal and therefore showing a worse separation than the one reached without additional cuts in figure 9.6. Nevertheless, some signal points show a positive significance. It can therefore be stated that having less background events can be more important than reaching a good separation of signal and background in the output score. These positive significances are also not attributable to overtraining as the ROC curve shows a good agreement between training and test sample. Moreover, the significance pad under the output score shows a smooth slope without jumps, showing therefore that the significance does not raise due to statistical fluctuations.

Even though this preselection does not give good results, at least it shows positive significances. This shows that a tighter preselection is a possible way to reach higher significances with the BDT. However, a better preselection might be achieved.

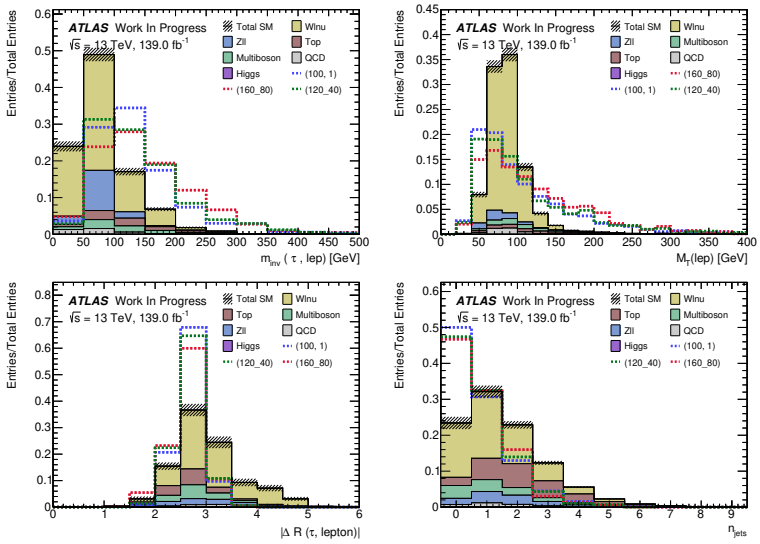


Figure 9.11: N-1 plots for the adapted CMS preselection

Preselection based on variable ranking

To try a completely new cut selection to add to the loose preselection, another cut combination was attempted by cutting on the variables that had the highest ranking for the BDT. Firstly, the preselection explained at the beginning of this section was applied. This training shows that the best variable for the BDT was m_{eff}. Therefore the value of the cut on m_{eff} was chosen looking at shape plots. Afterwards the training was repeated adding this cut to the preselection and the next cut was decided on the new best variable. This procedure was

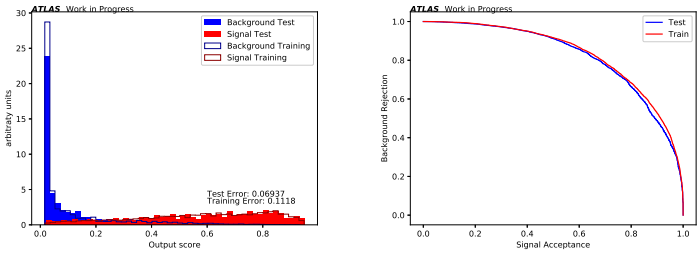


Figure 9.12: Output score and ROC curve for the training with the adapted CMS preselection. The hyperparameter user for this training were maximal depth: 3, learning rate: 0.01, number of estimators: 400

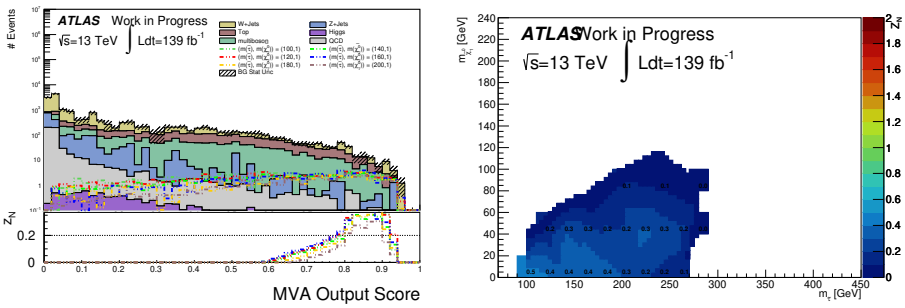
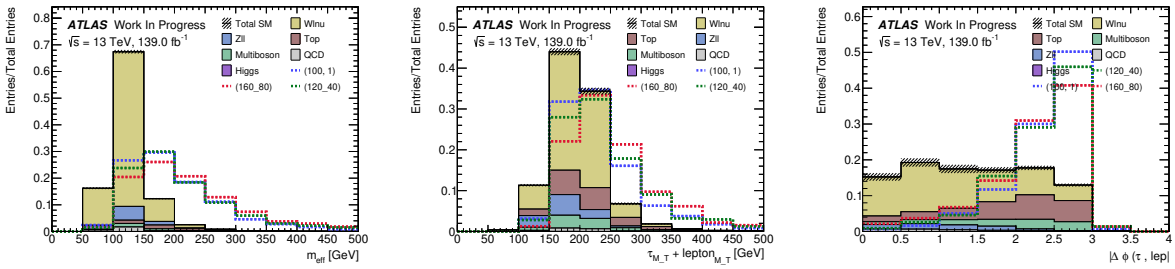


Figure 9.13: MVA output score and significance plot of the training with the adapted CMS preselection. The significance plot has been done with the cut MVAoutputscore > 0.86

repeated three times coming to the following preselection:

- $m_{\text{eff}} > 150$ GeV
- sum$M_T > 200$ GeV
- $\Delta\phi(\tau, \ell) > 1.5$



(a) **(b)** **(c)**

Figure 9.14: Shape plots for the preselection based on the variable with highest ranking. In (a) the distribution of m_{eff} is plotted after applying only the loose preselection. In (b) the distribution of sumM_T is plotted adding also the cut $m_{\text{eff}} > 150$ GeV to the loose preselection and in (c) the distribution of $\Delta\phi(\tau, \ell)$is plotted adding also the cut sum$M_T > 200$ GeV.

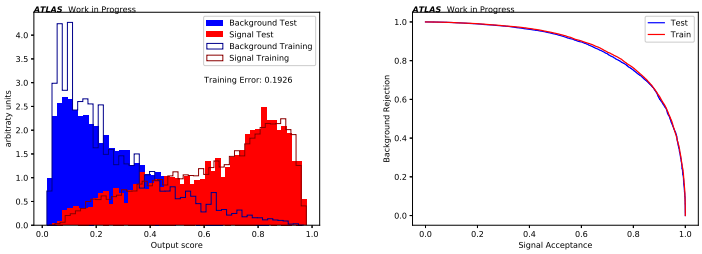


Figure 9.15: Output score and ROC curve for the training with the preselection based on the variable with the highest ranking. The hyperparameter used for this training were maximal depth:3, learning rate:0.01, number of estimators: 700

The shape plots used to decide the cut value can be seen in 9.14.
In 9.15 one can see the output score and the ROC curve resulting from this training, while in in 9.16 the significance plot and the output score with the significance pad.
With this training the output score shows discrimination of signal and background which is noticeably better than the one in the previous training (figure 9.12) and also the significances are slightly higher. The choice of the preselection is therefore relevant, as it has a big influence on the quality of the BDT training. Rejecting those events that the BDT used to choose most to separate signal and background shows good results as it helps boosting the significances by letting the BDT focus on less evident separations. Also in this case the ROC curve does not show a high overtraining and the significance pad under the output score does not show any jumps, making the delivered significance values reliable.
The yields after a the cut MVAOutpuScore>0.92 can be seen in table 9.2. The backgrounds

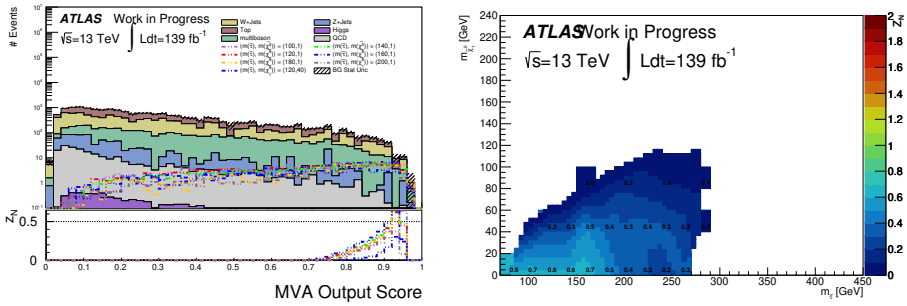


Figure 9.16: MVA output score and significance plot of the training with the preselection based on the variable with the highest ranking. The significance plot has been plotted with the cut MVAoutputscore>0.92

Process	Yield	Stat. error	Raw
Top	9.6	± 1.2	76
Multiboson	6.5	± 0.9	360
W+jets	4.4	± 4.9	54
Z+jets	0.8	± 2.5	10
QCD	0.2	± 0.1	3
Higgs	0.0	± 0.0	5
Total SM	21.4	± 5.7	508
DS (120,1)	9.6	± 0.6	306
DS (140,40)	5.9	± 0.7	70
DS (160,80)	3.1	± 0.4	69

Table 9.2: Number of raw and weighted events after a cut selection based on the variable ranking and and accepting events with an output score value greater than 0.92

that contribute the most after this cut are top, multiboson and $W+$ jets.

Preselection with a low ratio of signal to background

Until now the training preselection was kept around a maximum of 3 cuts except the cuts for the loose preselection. This was done because adding more and more variable cuts is a procedure which is very similar to cut-and-count and was already attempted. Aim of the MVA study was to let the BDT "learn" which variables give the best separation of signal to background. However, as a new strategy it was tried to apply an even tighter preselection to finally let the BDT train on variables which show signal separation which are not obvious when doing cut-and-count. It was therefore chosen a preselection with the same procedure described in the previous paragraph until a ratio of background to signal of 5:1 was reached. As the result the following preselection was obtained:

- $m_{\text{eff}} > 200$ GeV
- sum$M_T > 200$ GeV
- N(jets) < 2
- leptond0sig > 0.5

In figure 9.17 the training plots can be seen and in figure 9.18 can be seen the results.
The output score shows a slightly worse discrimination power than the one in the previ-

ous training (figure 9.15). Also the significances are not better than the previous training. Furthermore, from the ROC curve can be inferred that the sample have been slightly more overtrained. On the contrary the significance pad does not show big jumps until a value of 0.9 and the significances in the plot can be seen as reliable since the cut MVAoutput score>0.86 has been applied.

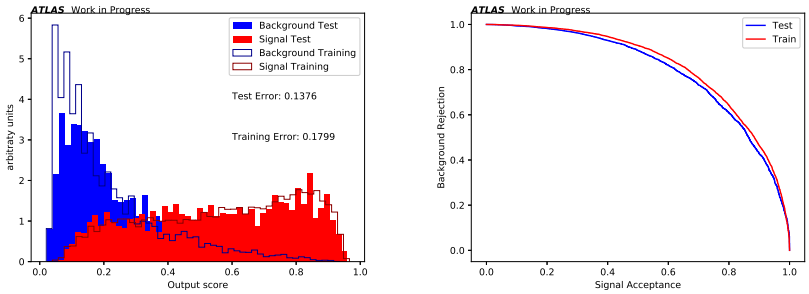


Figure 9.17: Output score, ROC curve for the training with the preselection with a low ratio of signal to background events. The hyperparameter user for this training were maximal depth: 3, learning rate: 0.01, number of estimators: 700

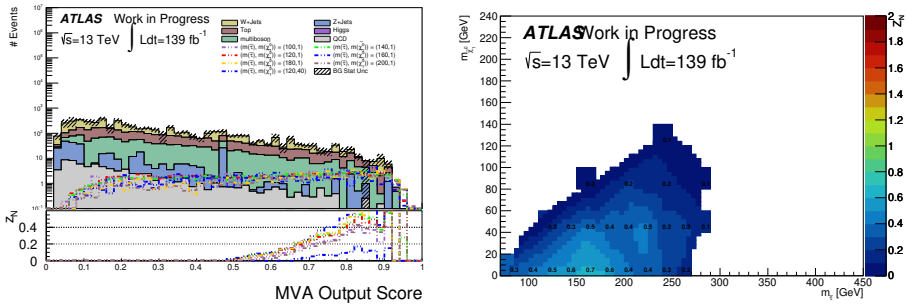


Figure 9.18: MVA output score and significance plot of the training with the preselection with a low ratio of signal to background events. The significance plot has been plotted with the cut MVAoutput score>0.86

9.3 Combining two BDTs Trained on Different Backgrounds

It is important to notice that different backgrounds may have different kinematic distributions. This could possibly lead to a less efficient training, since the BDT would than concentrate on the background with the highest number of events for the choice of variable to cut on, leading to a worse training of the other backgrounds. In order to avoid this, the shapes of singular backgrounds have been analysed in order to check if background might be divided in subgroups and trained separately. In figures 9.19 and 9.20 can be seen that $W+$jets and QCD could be put into one group and multiboson, higgs and top into another. $W+$jets and QCD have similar distributions as they are both reducible backgrounds, which show the same signature as signal because of jet which is misidentified as a tau lepton. $Z+$jets could be

part either of the first or of the second group depending on the variables. More shape plots can be found in the appendix.
The procedure for the training of separate BDTs consists of first choosing the division into subgroups, then take firstly one group as input for the BDT and evaluating it. Next, the already evaluated files should be taken as files to evaluate while taking the second group as input files.
This process creates MC samples with two MVA output scores one can than apply a cut on.

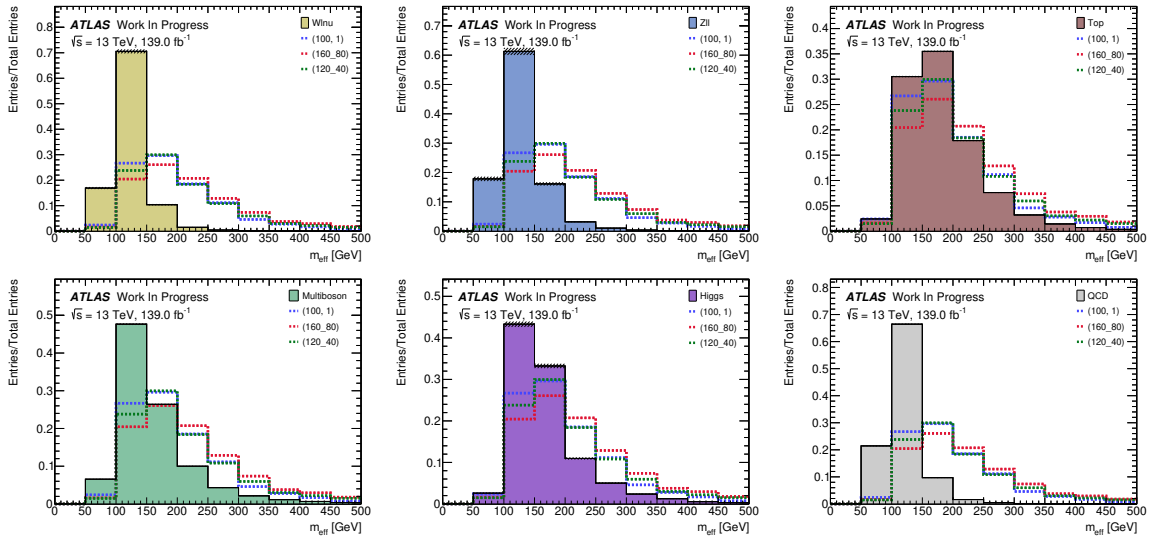


Figure 9.19: Single-background shape plots for the distribution of m_{eff} after the loose preselection. In this example the background $Z+$jets has a similar distribution to $W+$jets and QCD.

In order to choose in which group to place the $Z+$jets background the output scores of the training with separate groups are checked for a loose preselection. Once adding $Z+$jets to $W+$jets and QCD (figures 9.21 ans 9.22) , and once adding it to top, higgs and multiboson (figures 9.23 and 9.24).
It can be noticed that including $Z+$jets into one group or the other does not change the shape of the output score drastically. However, since the trainings of $W+$jets and QCD show evident difficulties in separating signal from background in both cases, it has been decided to train the $Z+$jets background together with higgs, top and multiboson. It is important to notice that the group of top, higgs and multiboson shows more overtraining than QCD and $W+$jets, including $Z+$jets or not.

When training with only a part of the backgrounds less background raw events are available. It is therefore not possible to apply a very strict preselection otherwise overtraining would occur. Only the following cuts have therefore been added to the usual "loose" preselection:

- $m_{\text{eff}} > 150\,\text{GeV}$
- $\text{P}_\text{T}(\tau) > 50\,\text{GeV}$

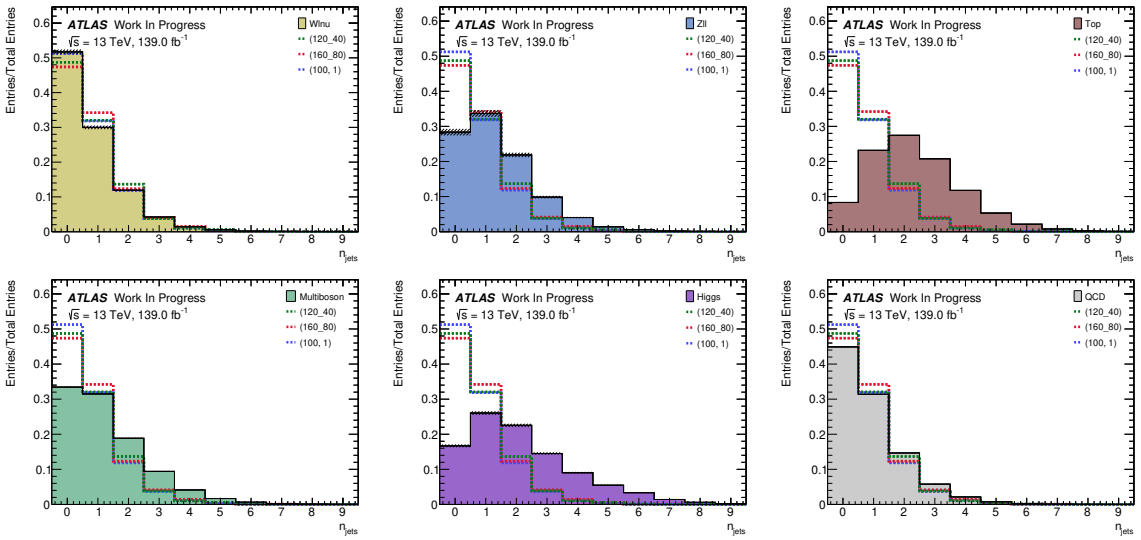


Figure 9.20: Single-background shape plots for the distribution of N(jets) after the loose preselection. In this example the background $Z+$jets has a similar distortion to multiboson, higgs and top.

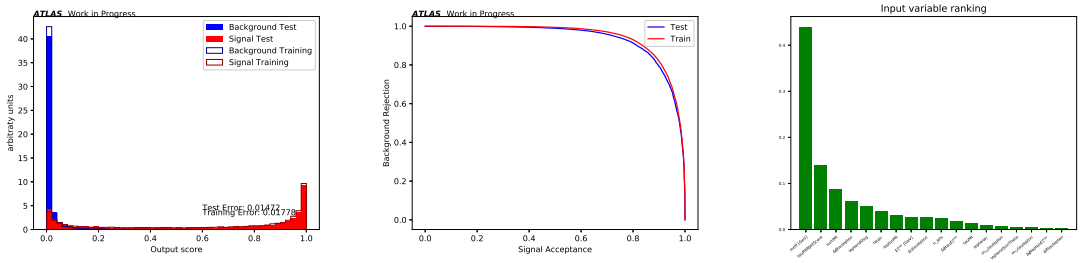


Figure 9.21: Output score, ROC curve and variable ranking for the training with QCD, $W+$jets and $Z+$jets as input. The hyperparameter used for this training were md=3,lr=0.1, n=400

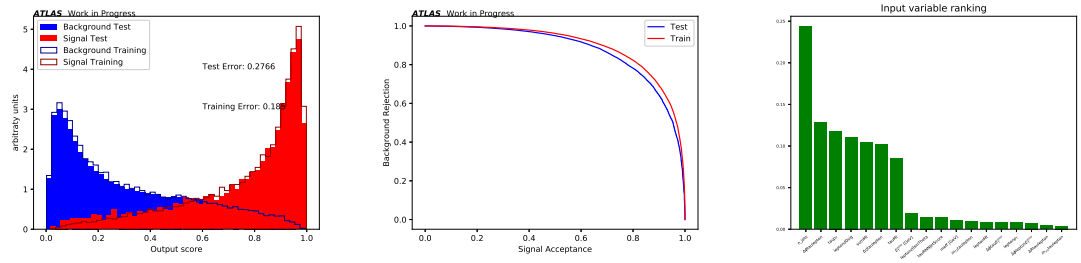


Figure 9.22: Output score, ROC curve and variable ranking for the training with top, higgs and multiboson as input. The hyperparameter used for this training were md=3,lr=0.1, n=400

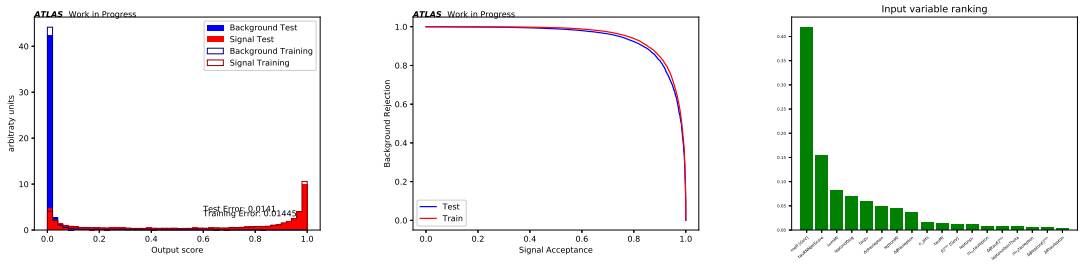


Figure 9.23: Output score, ROC curve and variable ranking for the training with QCD and $W+$jets as input. The hyperparameter used for this training were md=3,lr=0.1, n=400

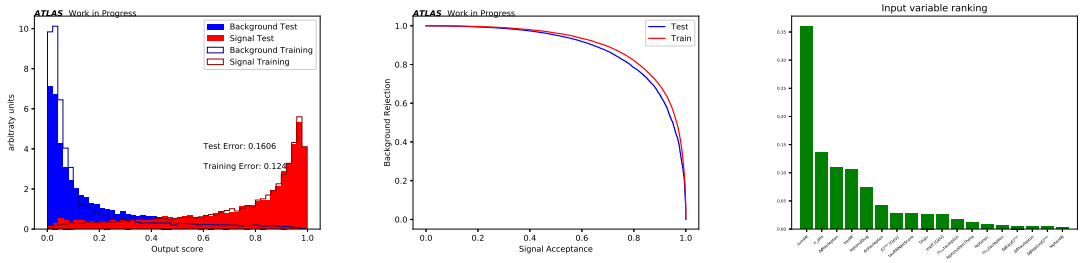


Figure 9.24: Output score, ROC curve and variable ranking for the training with $Z+$jets, top, higgs and multiboson as input. The hyperparameter used for this training were md=3,lr=0.1, n=400

Process	Yield	Stat. error	Raw
W+jets	20223.8	± 680.9	40837
Top	8004.6	± 33.9	67690
Z+jets	4856.9	± 229.3	52850
Multiboson	4029.5	± 21.7	287751
QCD	406.5	± 6.7	5190
Higgs	32.6	± 0.5	13427
Total SM	37553.8	± 719.7	467745
Total signal	1477.9	± 12.8	23119
DS (120,1)	168.2	± 2.7	5168
DS (140,40)	122.6	± 3.5	1457
DS (160,80)	62.1	± 1.9	1370

Table 9.3: Number of weighted and raw events after the preselection used for training two separate BDTs: m_{eff} >150 GeV , $P_T(\tau)$>50 GeV are added to the loose preselection

These cuts have been chosen with the aim of reducing especially the $W+$jets background, coming eventually to the yields written in table 9.3.

Both groups of backgrounds have been trained with a maximal depth of 3 and 400 decision trees. However, the learning rate of 0.1 which was used for training $W+$jets and QCD (figure 9.25) had to be lowered to 0.5 for the second group (figure 9.26) to avoid overtraining.

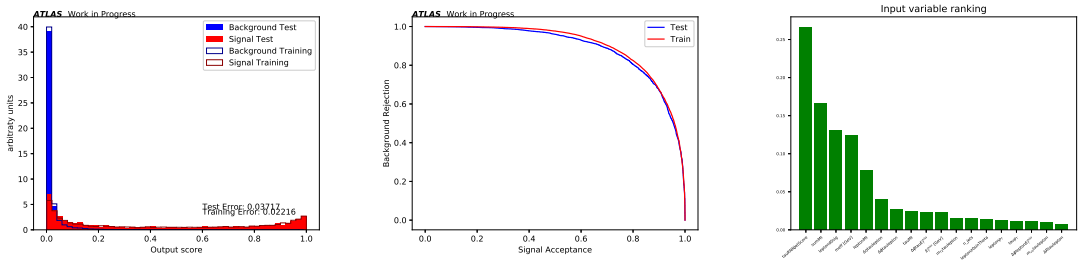


Figure 9.25: Output score, ROC curve and variable ranking for the training with QCD and $W+$jets as input. The hyperparameter used for this training were md=3,lr=0.1, n=400

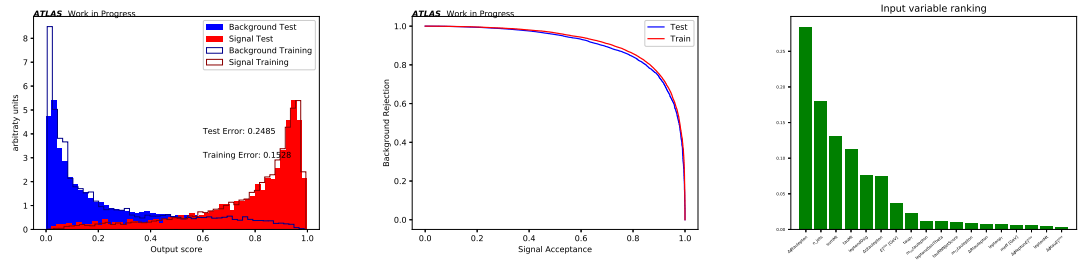


Figure 9.26: Output score, ROC curve and variable ranking for the training with $Z+$jets, higgs, multiboson and top as input. The hyperparameter used for this training were md=3,lr=0.05, n=400

Except the training preselection also the cut MVAoutputscore1>0.9 (resulting from the training of $W+$jets and QCD) has been applied and than the cut on the second MVA output score has been optimised for the point (140,40) coming eventually to the cut MVAoutputscore2>0.96.

As can be seen in figure 9.27, the resulting significances are slightly better than the ones achieved until now with a significance of 0.6 for the point (140,40) and the highest significance of 1.2 for the point (160,1). The number of events left after the cuts on the two output scores can be found in table 9.4. Also in this case the backgrounds that contribute the most are multiboson, top and $W+$jets.

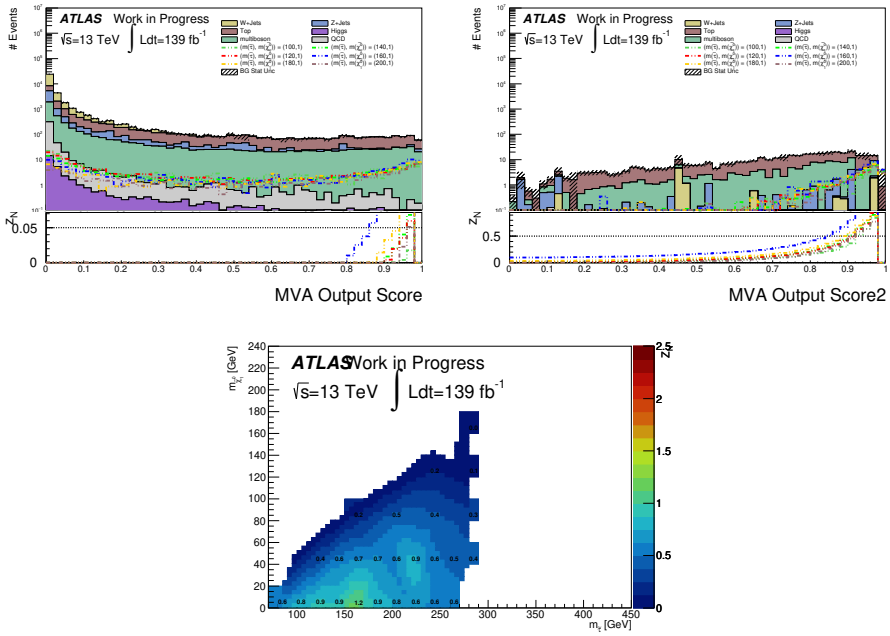


Figure 9.27: MVA Output Score 1 for the training with $W+$jets and QCD and MVA Output Score 2 for the backgrounds $Z+$jets, multiboson, higgs and top. MVA Output Score 2 has been plotted after the cut MVAoutputscore1>0.9. The significance plot has been evaluated for the point (140,40) and the second score and it has been plotted for the cut MVAoutputscore2>0.94

Process	Yield	Stat. error	Raw
Multiboson	23.0	± 1.2	724
Top	16.2	± 1.5	128
W+jets	4.7	± 1.7	17
Z+jets	2.2	± 1.2	10
QCD	0.1	± 0.1	1
Higgs	0.1	± 0.0	33
Total SM	46.4	± 2.9	913
DS (120,1)	13.0	± 0.7	406
DS (140,40)	9.7	± 1.0	111
DS (160,80)	4.5	± 0.5	102

Table 9.4: Number of raw and weighted events after a training of two separate BDTs

9.4 Optimisation with samples without Cut on Impact Parameter

As explained in chapter 4 the samples used until now select only signal leptons. Signal leptons have cuts on the impact parameter. In particular only events with electrond0sig< 5 and

muond0sig< 3 are accepted.
Usually this cut is applied to remove many particles coming from further decays of particles. These particles which come from a second vertex have usually big impact parameter values. However, in the case of leptonically decaying taus this cut might be disadvantageous since the wanted leptons come from a tau decay and therefore from a second vertex. As can be seen in figure 9.28 the d_0 significance of electron and muons show many signal events after the cut value of signal leptons.

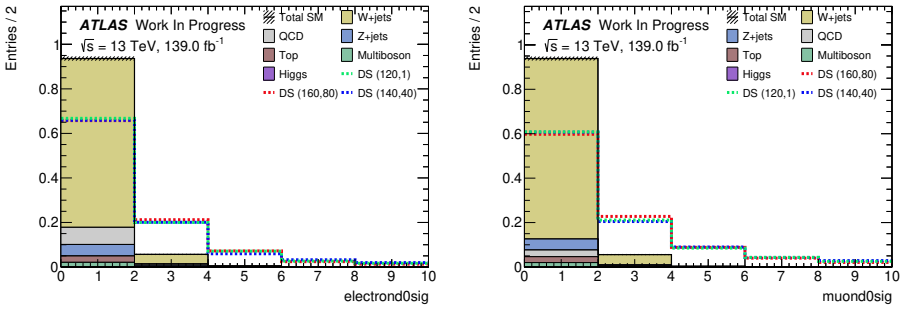


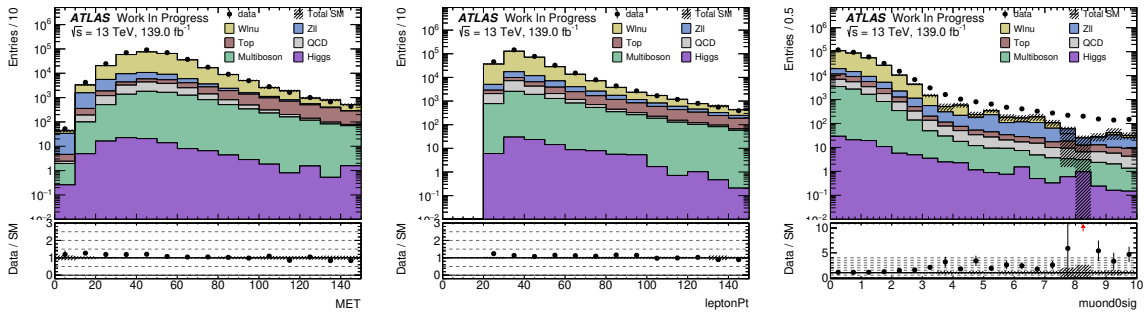
Figure 9.28: Shape plots of signal and background distribution of the electron and muon d_0 significance selecting baseline leptons

The BDT training might therefore profit from the acceptance of events with a high impact parameter value. To check this a sample with baseline leptons has been produced, since baseline leptons do not have any cuts on d0sig.
Before the training the agreement of data to MC should be checked first. As one can see in figure 9.29 there is good agreement for some variables like $E_{\mathrm{T}}^{\mathrm{miss}}$ and the lepton transverse momentum , but a high disagreement for the d0 significance, especially for muons.

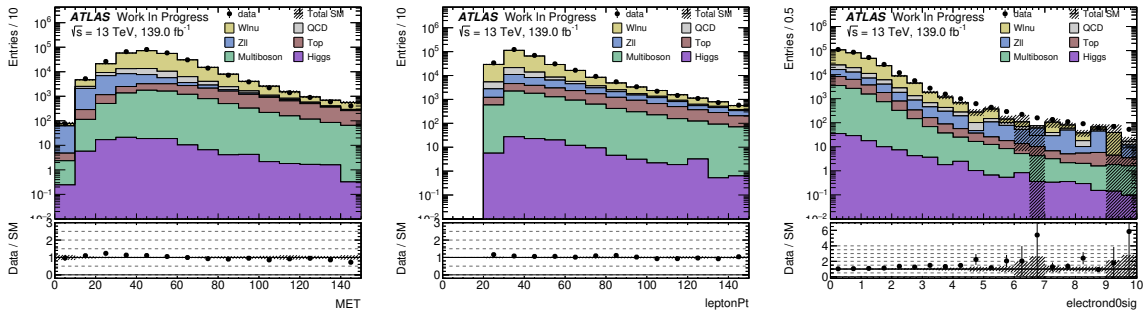
It is important to notice that this mismodeling is still present after the inclusion of QCD and is therefore not caused by a missing background, but from a background underestimation in the regions of d0significance greater than 3 or 5. The background therefore has to be reweighted otherwise it would lead to unrealistic high significance values. Like in figure 9.29 data is blinded at d0sig<10 and divided into 20 bins. With this binning a new weight is given to each subgroup in order to be equal to data. The background events after the blinding cut are not rescaled. The results of this reweighting are in figure 9.30. A comparison of the d0sig distributions before and after the reweighting can be seen in figure 9.31. The impact parameter significance has a good agreement for both electrons and muons and the other two variable kept their good agreement of data to MC also after the reweighting technique.

These rescaled samples have been used for the training. The preselection cuts that have been applied additionally to the usual preselection cuts for the training are:

- $m_{\mathrm{eff}} > 150$ GeV
- sum$M_T > 200$ GeV
- $\Delta\eta(\tau, \ell) < 2$

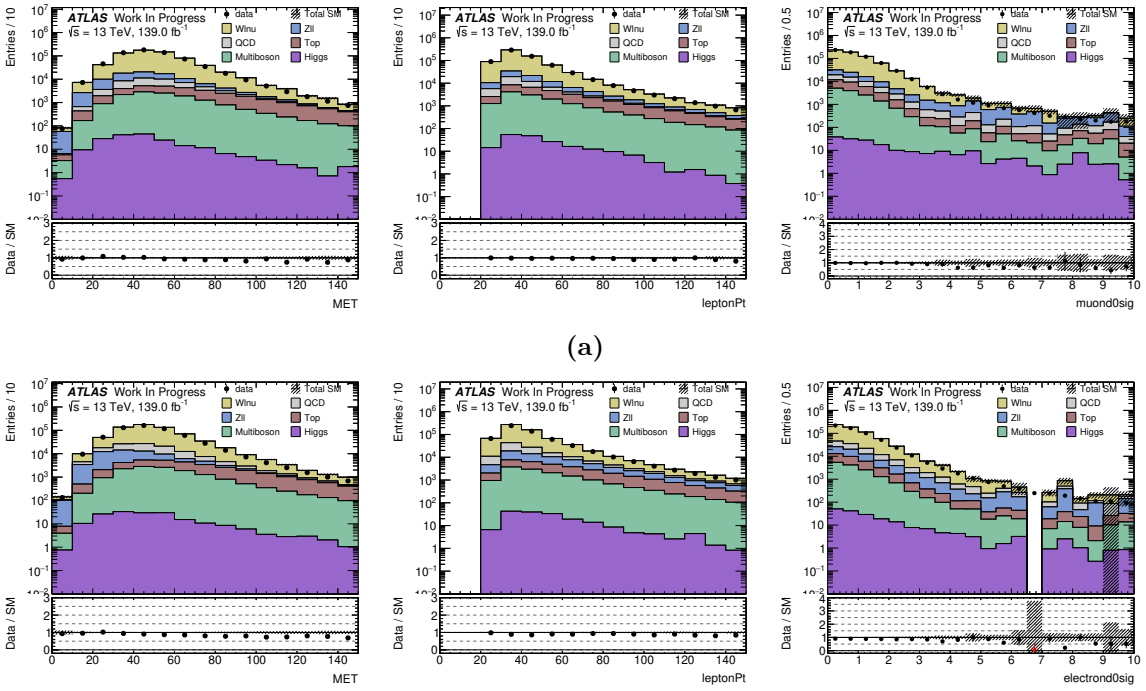


(a)



(b)

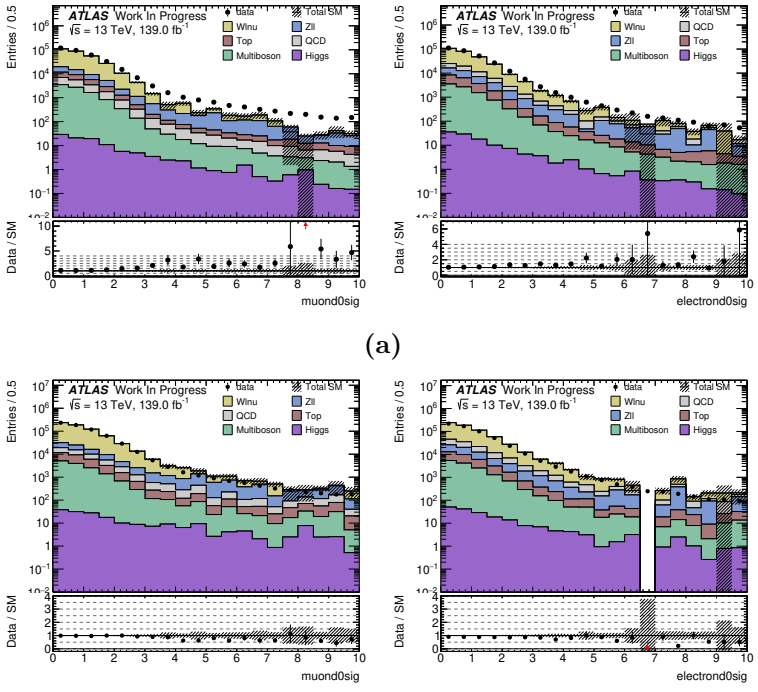
Figure 9.29: Blinded data plotted over background expectation for muons (a) and electrons (b)



(a)

(b)

Figure 9.30: Blinded data plotted over background expectation for muons (a) and electrons (b) after applying a reweight



(a)

(b)

Figure 9.31: Distribution of electron and muon d0sig before (a) and after (b) the reweighting

- $\Delta R(\tau, \ell) < 4$
- d0sig $(\ell) > 2$

The cuts on m_{eff} and sumM_T have been applied as they have always shown a good rejection of background, the cuts on $\Delta\eta(\tau, \ell)$and $\Delta R(\tau, \ell)$rejected only background events and nearly no signal events and finally the d0sig (ℓ) cut be stricter than in other occasions, since the upper requirement on this variable has been removed.
The significances that can be reached with this reweighted sample can be seen in figure 9.32, while the yields are listed in table 9.5. The significances are slightly better than the ones obtained with previously presented methods. However, these samples are also more unreliable than others due to the reweighting. As can be seen in the output score plot in figure 9.33, the significances show a sudden jump which is probably due to a background underestimation.

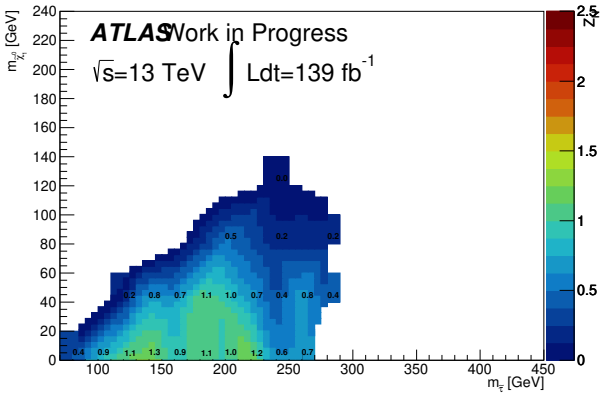


Figure 9.32: Significance plot after a training with a reweighted sample with baseline leptons

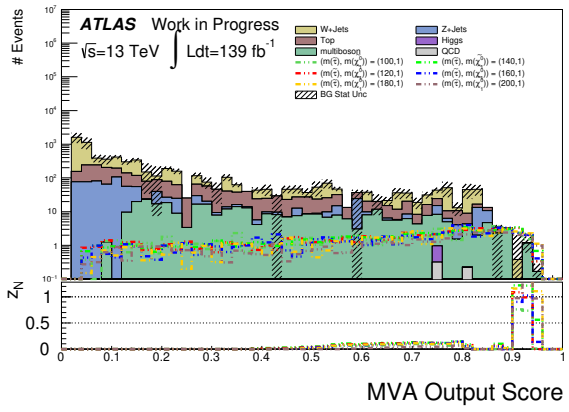


Figure 9.33: Output score of a training with a reweighted sample with baseline leptons

Process	Yield	Stat. error	Raw
Multiboson	4.5	± 0.6	96
Top	2.0	± 0.6	14
W+jets	0.7	± 0.4	4
Z+jets	0.1	± 0.1	2
Higgs	0.0	± 0.0	2
QCD	0.0	± 0.0	0
Total SM	7.3	± 0.9	118
DS (120,1)	3.7	± 0.4	113
DS (140,40)	2.5	± 0.5	29
DS (160,80)	0.9	± 0.2	21

Table 9.5: Number of raw and weighted events after a training on reweighted samples with baseline leptons

9.5 Optimisation with Truth Smearing

A possible reason for the low significances reached with BDTs might be that training with several signal points might lead to a less efficient discrimination power since summing different signal points smears the signal distribution. However, the training with one single signal point is impossible as it would have an insufficient number of raw events and as explained in section 9.1 high statistics are essential to avoid overtraining.
To overcome this problem a signal sample produced with a process of *truth smearing* has been produced. The advantage of using truth smearing samples is that a high number of raw events can be reached in a computational cheap way, avoiding a full detector simulation. A truth smearing sample is in fact a sample with truth events which are corrected with a smearing procedure. This procedure is necessary to approximate the detector response and thereby have a more reliable sample. This type of sample will therefore be very similar to the signal sample with the simulation of the detector response, but not identical.
For the training a single truth smeared sample for the point (120,40) was given as input together with the background samples. Instead, for the evaluation the normal samples were used, including also the same point as the sample used for training. Since triggers were not present in the truth smearing samples the offline cuts had to be included in the preselection. Furthermore the variables tauRNNJetScore and leptond0sig had to be excluded from the input variables, as this information is not present for truth taus. The variable N(jets) was also excluded as it showed a separation which diverged from the sample with reconstructed particles.
To make a proper comparison between the training with truth smearing samples and the usual samples with complete detector simulation a training with the usual set of signal input samples has been performed excluding the previously mentioned variables. As a preselection the cuts $m_{\text{eff}} > 150$ GeV, sum$M_T > 200$ GeV and $\Delta\phi(\tau, \ell) > 1.5$ have been applied. The cut of the output score has been chosen optimising on the point (120,40) where a significance of 0.056 could be reached. In figure 9.34 the output score and the significance plot can be seen. The significance of the point (120,40) has been rounded to 0.1.

When changing the point (120,40) with the truth smearing sample and keeping every other signal sample with full detector reconstruction, the significance of the point (120,40) can be slightly improved to the value of 0.6. This hints to the fact that a higher number of raw events helps increasing the significances of a single point.

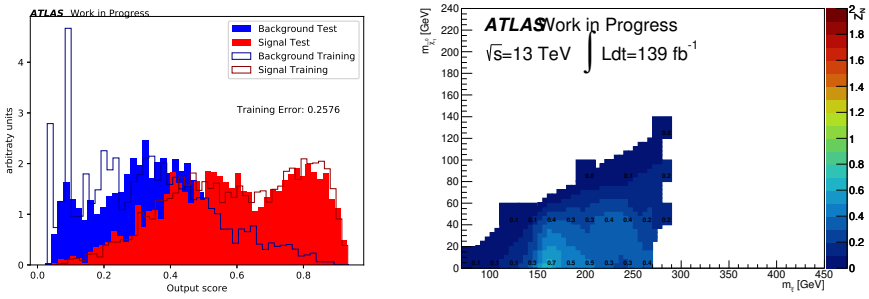


Figure 9.34: Output score and significance plot for a training suing samples with full detector simulation, excluding tauRNNJetScore, N(Jets), leptond0sig from the input variables

However, when training only with the truth smearing sample worse results are delivered, as can be seen in the significance plot in figure 9.35. In the same figure can be seen also the output score.
When comparing this output score with the one in figure 9.34 it can be noticed that training with only one signal point leads to a much weaker discrimination power. This effect might be due to the lower raw events which are left after the preselection. The truth smearing samples has 7809 events after the preselection, while the sum of the input samples with complete detector reconstruction is of 12638 raw events. However, the difference in raw events is very low compared to the difference in discrimination power. It is more likely that including only one signal point is insufficient for a BDT training.

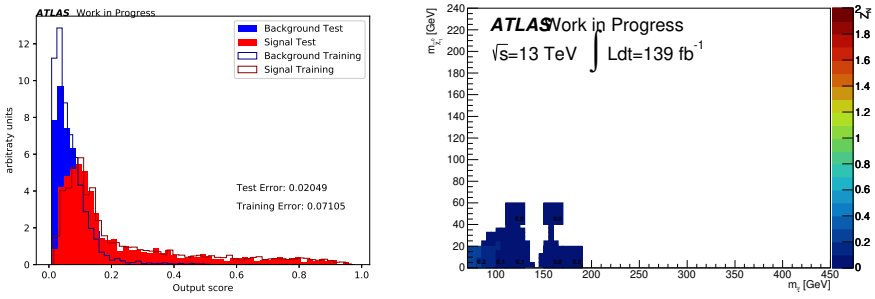


Figure 9.35: Output score and significance plot for a training suing samples with full detector simulation, excluding tauRNNJetScore, N(Jets), leptond0sig from the input variables

9.6 Optimisation with a Signal Grid Extension

After all the training methods presented until now had been performed, an extension of the signal sample grid became available to check whether more signal events could improve the significances delivered until now. The new signal grid can be seen in figure 9.36.
Firstly the preselection including a cut on three variables based on the variable ranking

presented in section 9.2 was used. Training with the same hyperparameters of the previous training but with more signal points led to slightly better results as can be seen in figure 9.37. The number of events left after the cut on the output score can be found in table 9.6. The extension of the sample grid included especially points in the compressed region. The signal points used for training are also all the available signal sample with stau mass between 80-180 GeV (see signal grid in figure 9.36). The cut on the output score is MVAoutputscore>0.92 and it has been chosen by optimising on the signal point (140,40), the same as with the old signal grid. It can therefore be stated that the outcome of a BDT training can be improved including more signal points and therefore more weighted events.
The preselection with a low ratio of signal to background has also been used to compare the results to the old ones, as more training events could reduce the overtraining which was previously present. This can be confirmed looking at the new output score and ROC curve obtained with the grid extension which can be seen in figure 9.38. The comparison of the two results can be found in figure 9.39. Also in this case the grid extension helps to improve the significances, however this training still has less discrimination power than the previously presented one.

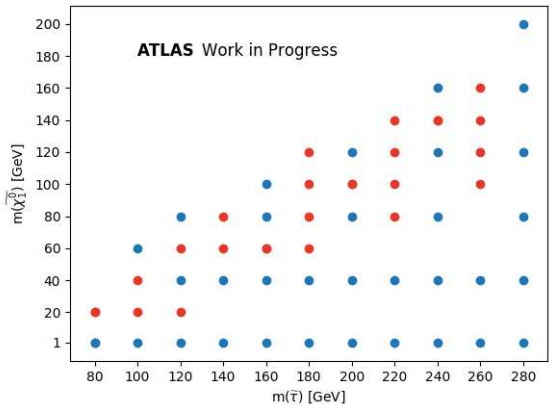
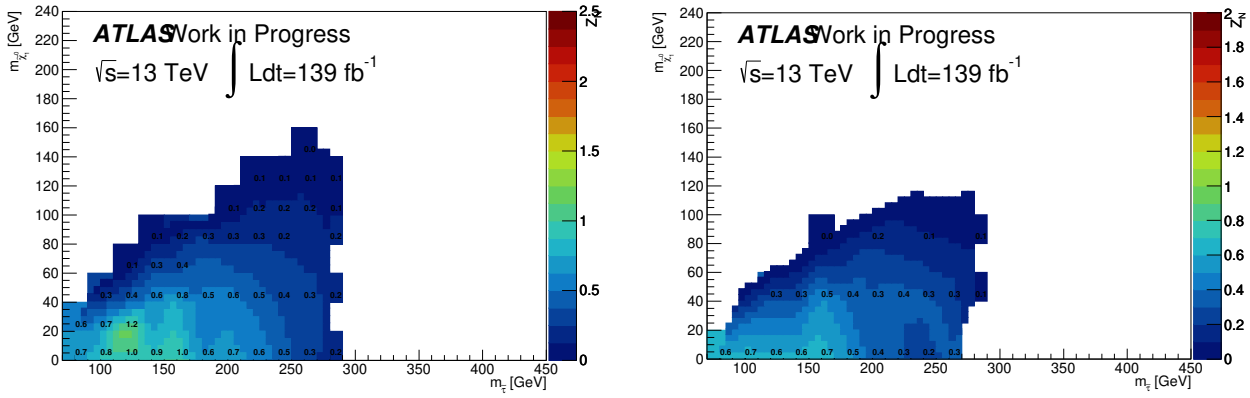


Figure 9.36: Available signal points after grid extension - new points in red

Process	Yield	Stat. error	Raw
W+jets	14.7	± 4.8	31
Multiboson	11.9	± 0.9	322
Top	9.3	± 1.1	74
QCD	0.7	± 0.3	7
Z+jets	0.7	± 0.4	4
Higgs	0.0	± 0.0	2
Total SM	37.2	± 5.1	440
DS (120,1)	14.7	± 0.8	368
DS (140,40)	10.1	± 1.0	108
DS (160,80)	5.0	± 0.5	89

Table 9.6: Number of raw and weighted events after a training with a preselection based on the variable ranking with the grid extension



(a) **(b)**

Figure 9.37: Significance plot after a training with a preselection based on variable ranking with a signal grid extension (a) and without the signal grid extension (b). The significances in picture (a) are delivered after a cut on the output score is MVAoutputscore>0.92

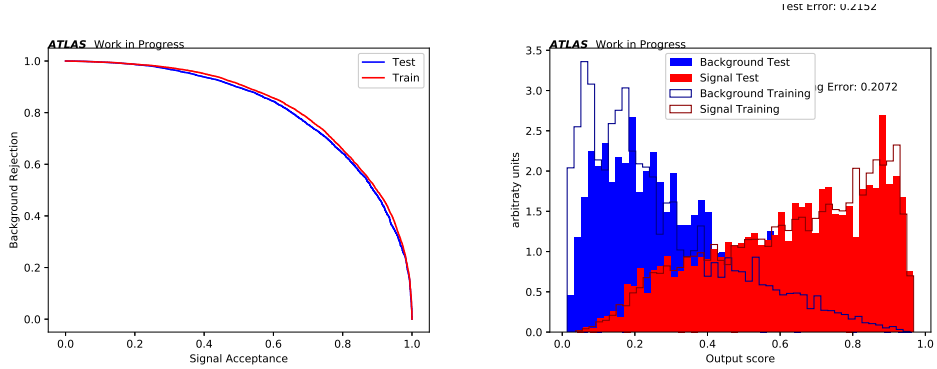
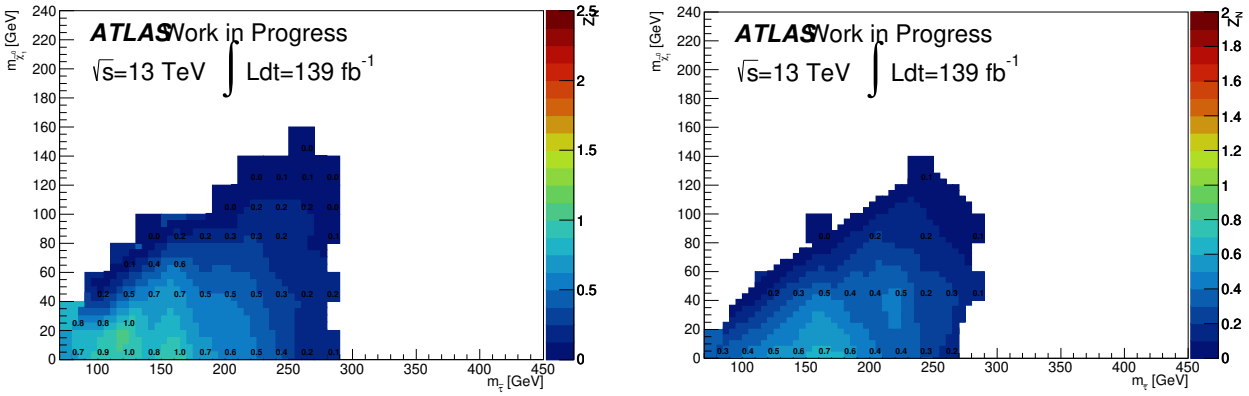


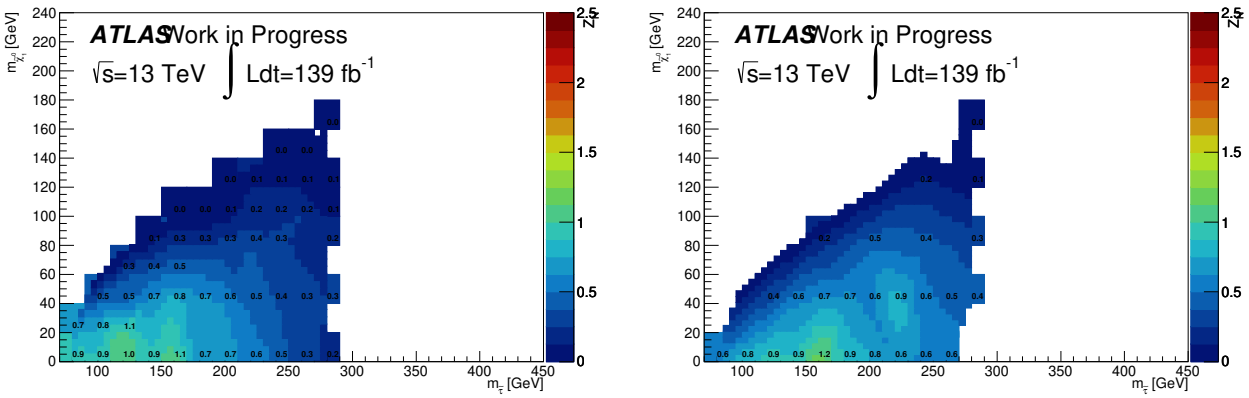
Figure 9.38: output score and ROC curve resulting from a training with signal grid extension



(a) **(b)**

Figure 9.39: Significance plot after a training with the preselection with low ration from signal to background of section 9.2 with a signal grid extension (a) and without the signal grid extension (b)

When training two separate BDTs under with the same conditions (same hyperparameters, same division in groups and preselection cuts) as in section 9.3, but with the grid extension, one obtains the results in figure 9.40. The yields resulting from this training are listed in table 9.7.



(a) **(b)**

Figure 9.40: Significance plot after a training with two separate BDTs with the signal grid extension (a) and without the signal grid extension (b). The significances in plot (a) have been delivered after the cuts MVAOutputScore>0.9 and MVAOutputScore2>0.96

In both cases the results with the signal grid extension show slightly better results.

Process	Yield	Stat. error	Raw
Multiboson	23.1	± 1.3	609
Top	16.7	± 1.5	130
W+jets	10.4	± 2.7	33
Z+jets	3.4	± 1.7	9
QCD	0.3	± 0.2	3
Higgs	0.1	± 0.0	25
Total SM	54.0	± 3.8	809
DS (120,1)	19.9	± 0.9	496
DS (140,40)	13.3	± 1.2	136
DS (160,80)	7.3	± 0.7	131

Table 9.7: Number of raw and weighted events after a training of two separate BDTs with the grid extension

10 Control Regions

When estimating the background present in the signal region it is more advisable to define a control region (CR) to check if the estimation is realistic. A CR has to be rich of background events and very poor of signal events. Once this region is defined the number of events is compared to data and a scaling factor is calculated. This scaling factor should then be applied to the background estimation in the signal region.
A control region should be defined for the backgrounds that contribute the most to a signal region. For every attempt in defining a signal region the three most contributing backgrounds are always $W+$jets, Top and Multiboson. For these three backgrounds a control region is therefore defined.
To calculate a scaling factor that is realistic also in the SR the CR must have a similar cut selection. Every CR defined in the following section will therefore select events with:

- one medium tau
- one signal lepton with opposite sign to the tau
- $m_{\mathrm{T2}} > 30$ GeV

Further cuts will be applied to define a CR of a specific background type. The additional cuts are defined aiming to a maximal *purity* of the region, therefore trying to reach the highest ratio of the selected background to the total background. The scaling factor is calculated by dividing the number of data by the total background events in the CR.

10.1 W+jets Control Region

The $W+$jets CR is defined by adding the cut $\Delta\phi(\tau, E_T^{miss}) > 2$ and a b-veto. From the N-1 plots of these variables in figure 10.1 it can be inferred that the $\Delta\phi(\tau, E_T^{miss})$ cut defines a region rich of $W+$jets events and the b-veto reduces the contribution of Top events.

The yields resulting from this cut selection can be found in table 10.1. It can be calculated that the ratio of weighted $W+$jets events to the total background is of 85 %, making this a reliable $W+$jets CR. When looking at the distributions of the lepton and tau transverse momentum, as well ass the missing transverse energy in this region (figure 10.2) a good agreement of SM estimation and (blinded) data can be confirmed. This is important to check the validity of the scaling factor.
With the yields in table 10.1 the scaling factor is calculated to be 1.02, so the $W+$jets background is slightly underestimated.

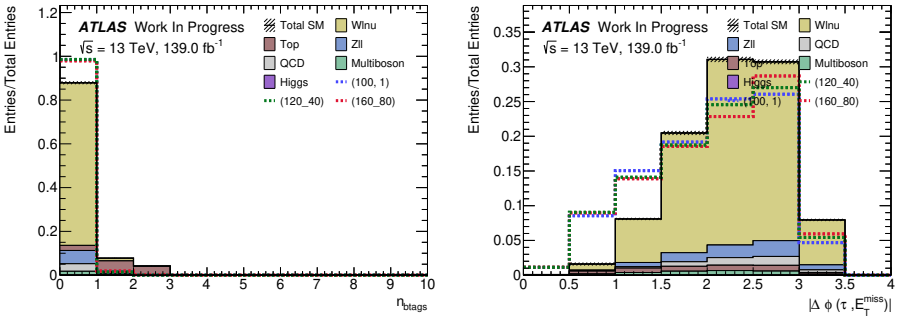


Figure 10.1: N-1 plots for the $W+$jets CR definition which is defined by the cuts $\Delta\phi(\tau, E_T^{miss}) > 2$ and N(b-tags) ==0 in addition to the preselection of 1 medium tau, 1 signal lepton with opposite sign and $m_{\mathrm{T2}} > 30$ GeV .

Process	Yield	Stat. error	Raw
W+jets	707951.3	± 6268.2	405935
Z+jets	57773.9	± 971.1	162697
QCD	32988.0	± 59.4	456450
Top	22084.7	± 55.8	193094
Multiboson	16517.8	± 51.5	728113
Higgs	74.4	± 0.8	25301
Total SM	837390.2	± 6343.7	1971590
Data	854716.0	± 924.5	854716
DS (120,1)	222.5	± 3.0	6965
DS (140,40)	147.7	± 3.7	1818
DS (160,80)	85.4	± 2.2	1907

Table 10.1: Number of events after the definition of the $W+$jets CR

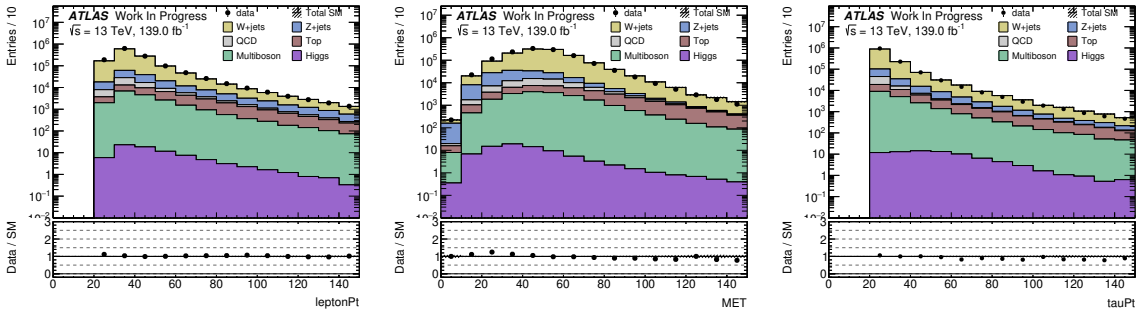


Figure 10.2: Distributions of $\mathrm{P_T}(\ell)$, $\mathrm{P_T}(\tau)$and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the $W+$jets CR with blinded data

10.2 Top Control Region

For the Top CR the number of b-tags has to be greater than 0. As can be seen in figure 10.3 this cut not only allows for a definition of a Top CR with high purity but it also removes almost every signal event. In table 10.2 can be seen the yields of the Top CR. It can be calculated that 88 % of the total background events come from Top events. Also for the Top CR a good agreement between data and SM expectations can be seen in figure 10.4. The scaling factor is 0.97. Therefore the Top background is slightly overestimated.

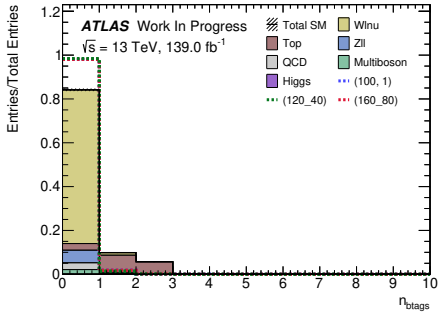


Figure 10.3: Shape plot of N(b-tags) after selecting 1 medium tau, 1 signal lepton with opposite sign and $m_{\mathrm{T2}} > 30$ GeV

Process	Yield	Stat. error	Raw
Top	199067.8	± 165.7	1739979
W+jets	18121.6	± 564.4	78885
Z+jets	4492.1	± 150.7	75397
QCD	2955.5	± 18.0	38594
Multiboson	1064.4	± 12.4	82311
Higgs	196.8	± 0.5	326292
Total SM	225898.1	± 607.6	2341458
Data	219215.0	± 468.2	219215
DS (120,1)	6.1	± 0.6	154
DS (140,40)	5.2	± 0.9	49
DS (160,80)	2.0	± 0.4	37

Table 10.2: Number of events after the definition of the Top CR

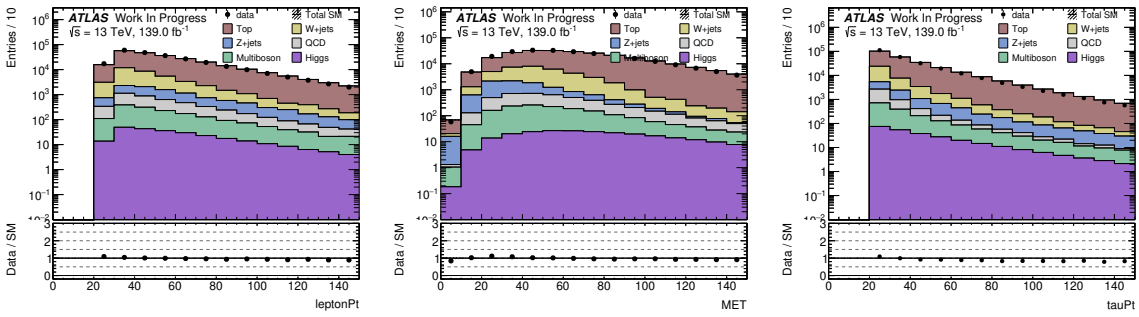


Figure 10.4: Distributions of $\mathrm{P_T}(\ell)$, $\mathrm{P_T}(\tau)$and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the Top CR with blinded data

10.3 Multiboson Control Region

The definition of a Multiboson CR requires more cuts due to the lower number of events compared to $W+$jets and the similarity of its distribution to the one of the Top background. It is defined by adding the cuts $\Delta\phi(\tau, E_T^{miss})$ <0.5 and selecting tight taus to reduce the number of $W+$jets events and N(jets)<2 and $m_{inv}(\tau, \ell)$> 270 GeV to reduce the number of Top events. The N-1 shape plots for these three cuts can be seen in figure 10.5. The final yields can be seen in table 10.3. With these results it can be calculated that the contribution of multiboson to the total background is of 55%. The purity of the Multiboson CR is therefore much lower compared to the other two CR, but a higher purity could have only be reached which more cuts. However, more cuts cannot be applied, otherwise the statistical uncertainty of the background would rise too much. To check the agreement between data and SM expectations the number of bins had to be reduced due to the low number of total background events. The distributions show an agreement within the statistical uncertainties in the missing transverse energy and the lepton transverse momentum, while in the tau lepton P_T a slight overestimation can be seen (see figure 10.6).
In fact, the scaling that can be calculated with this region is 0.86, which represents a slight overestimation.

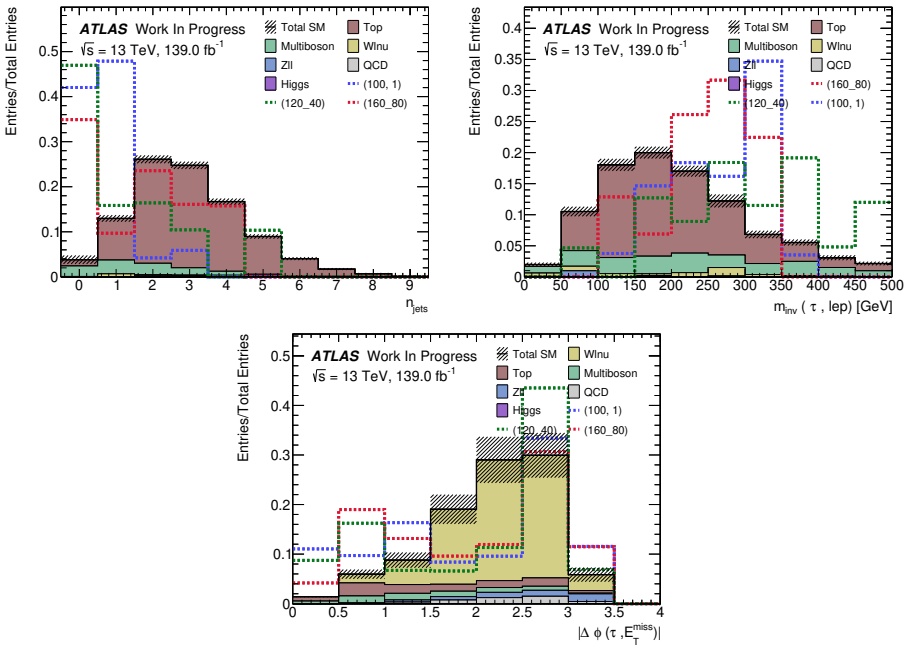


Figure 10.5: N-1 plots of the cuts defining the Multiboson CR after selecting 1 medium tau, 1 signal lepton with opposite sign and $m_{\mathrm{T2}} > 30$ GeV

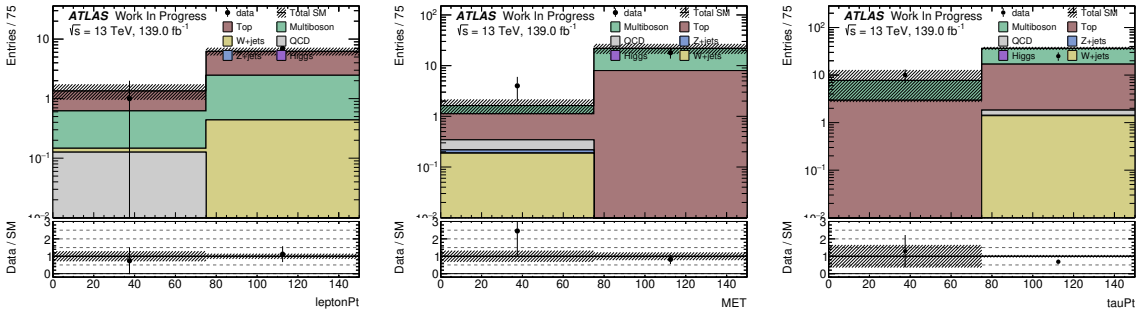


Figure 10.6: Distributions of $P_T(\ell)$, $P_T(\tau)$and E_T^{miss} in the Multiboson CR with blinded data

Process	Yield	Stat. error	Raw
Multiboson	31.2	± 2.2	1571
Top	24.9	± 2.0	220
QCD	0.7	± 0.3	8
Z+jets	0.1	± 0.1	5
Higgs	0.0	± 0.0	1
W+jets	-0.2	± 4.9	17
Total SM	56.6	± 5.8	1822
Data	46.0	± 6.8	46
DS (120,1)	1.4	± 0.2	41
DS (140,40)	0.5	± 0.2	5
DS (160,80)	0.3	± 0.1	5

Table 10.3: Number of events after the definition of the Multiboson CR

11 Conclusion

11.1 Combination with the HadHad Channel

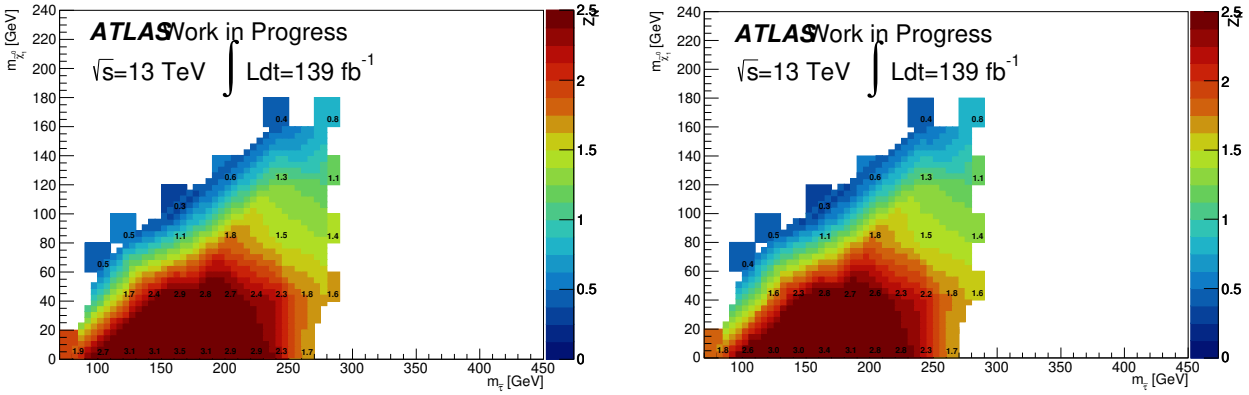
The significances obtained with the LepHad channel might not be sufficient by themselves to obtain sensitivity, but the might be helpful to enlarge the area with sensitivity obtained with the HadHad channel. In order to check this possibility the significances of the two channels are combined as the following:

$$Z_N^{\text{Comb}} = \sqrt{Z_N^{LH\,2} + Z_N^{HH\,2}} \tag{11.1}$$

As an example two of the results obtained with the signal grid extension are combined with the results of the HadHad channel to compare them with the significance plot obtained including only the HadHad channel.
In figure 11.1 can be seen the combination of the significances of the HadHad channel and the significances obtained with a preselection with cuts on the three variables with the best ranking trained with the signal grid extension. In figure 11.2 can be seen the same plot but with the LepHad significances obtained with a training of two separate BDTs and the signal grid extension. Both LepHad significances have been presented in section 9.6. In the same figures can be seen also the results obtained only from the HadHad channel. When comparing these two results with the significances obtained only with the HadHad channel it can be inferred that only a few signal points show an improved significance. In both cases the points with a stau mass of 140 GeV and 120 GeV both with a neutralino mass of 40 GeV ((140,40) and (120,40)) have a significance with a value which is of 0.1 higher than the one obtained including only the HadHad channel. However, none of the two possible signal region definition succeeds in improving the significance for the signal point (160,80) or (100,60) which are more distant from the exclusion limits set with the HadHad channel. However, without a fit of the significance plot it cannnot be stated whether the phase space region with a calculated significance of at least 1.6 can be enlarged. This would be the region where in case of the absence of a data excess over SM expectations exclusion limits with a 95% CL could be set, as explained in chapter 5.

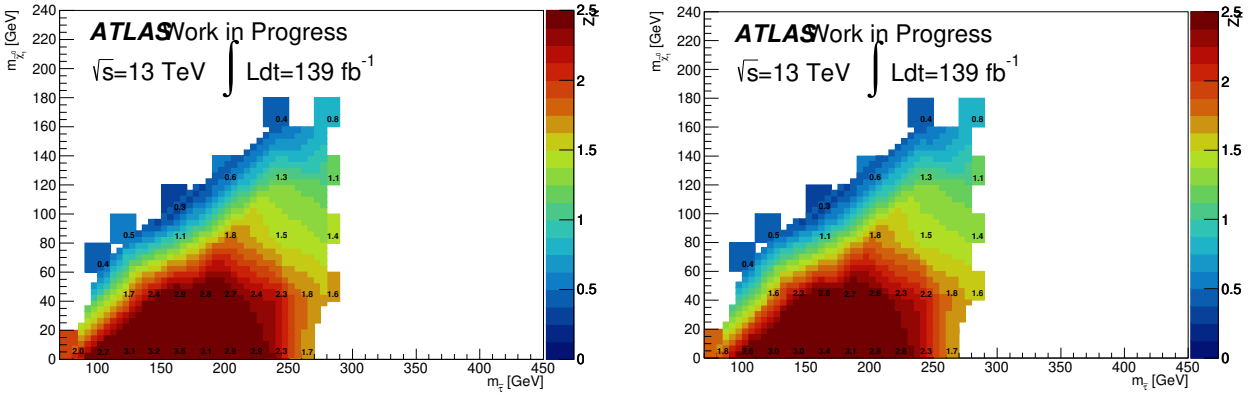
11.2 Interpretation of Results

As the Cut-and-Count method presented in chapter 8 delivered only negative significances which cannot improve the results obtained by the HadHad channel, only the results presented in chapter 9 using the MVA method can be considered for a possible signal region definition. Several approaches have been attempted.
In all of these cases it has been shown that rejecting a high number of background events



(a) (b)

Figure 11.1: in (a) combination of the significances of the LepHad and the HadHad channel using equation 11.1. The results of the LepHad channel are those presented in section 9.6 with the preselection containing cuts on the three variables with the best BDT ranking. In (b) the results obtained including only the HadHad channel



(a) (b)

Figure 11.2: in (a) combination of the significances of the LepHad and the HadHad channel using equation 11.1. The results of the LepHad channel are those presented in section 9.6 with the training of two separate BDTs. In (b) the results obtained including only the HadHad channel

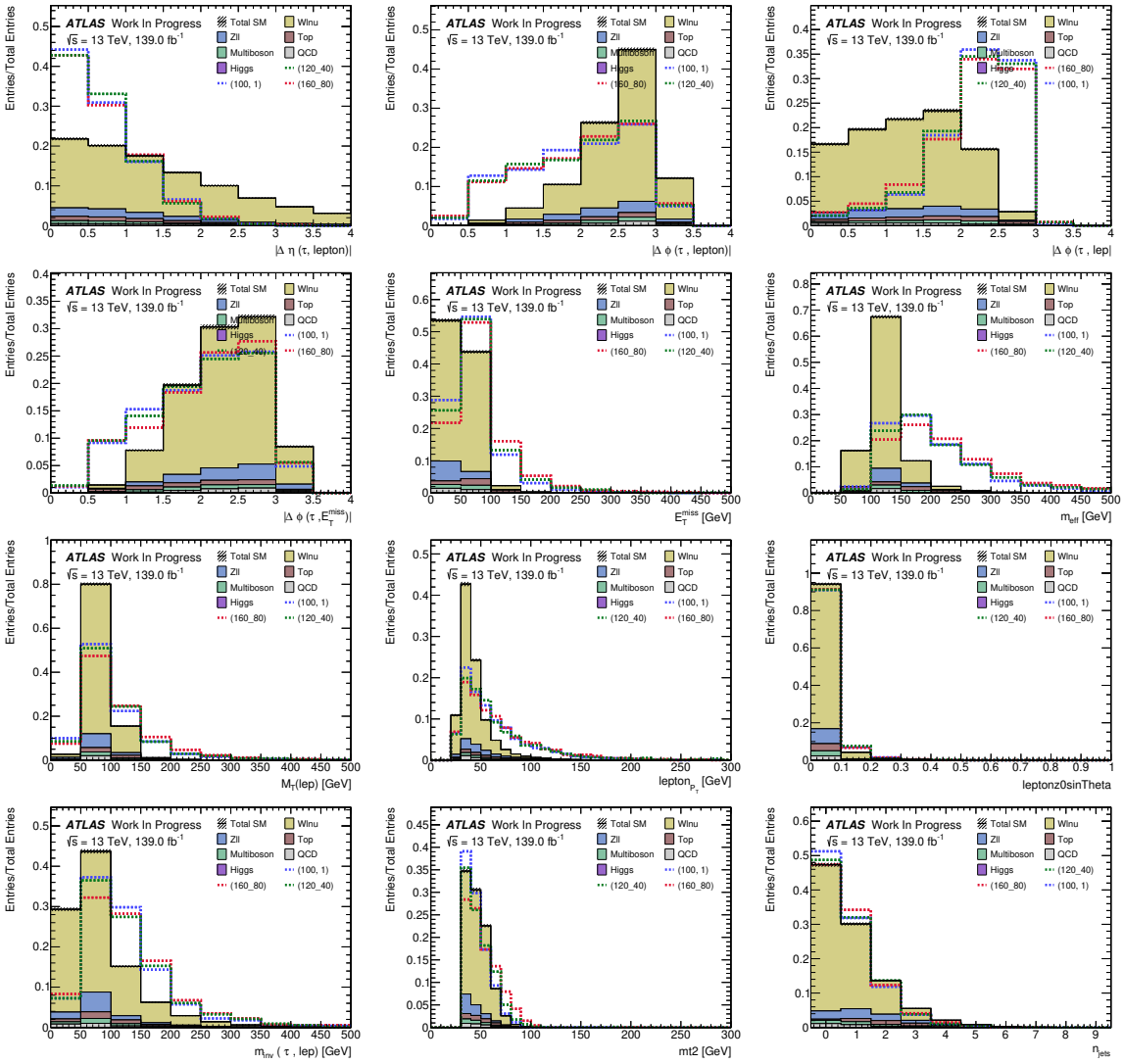
before the training is essential to reach positive significances with the LepHad channel. However, it is important that this preselection is not to strict, otherwise overtraining would occur. It should be trained with a number of events that still allows for a reduction of overtraining with a learning rate that is not lower than 0.01.
Furthermore, the training of separate BDTs delivered slightly better results than the ones obtained with one training and a strict preselection. However, when combining the significances with the HadHad channel, both approaches deliver similar results.
With the samples with baseline leptons higher significances could be calculated. However, the calculations with these samples are also less realistic due to the necessary reweighting technique.
The training with signal samples produced with a truth smearing technique instead of a full detector simulation didn't deliver sufficient results. Most likely due to the inability of the BDT to discriminate signal and background with only one signal point.
Moreover, the extension of the signal grid led to a slight improvement of the results. This is due to the higher number of events that could be given as input to the BDT. A higher number of events in the same phase space region could therefore help to increase the sensitivity in that region.

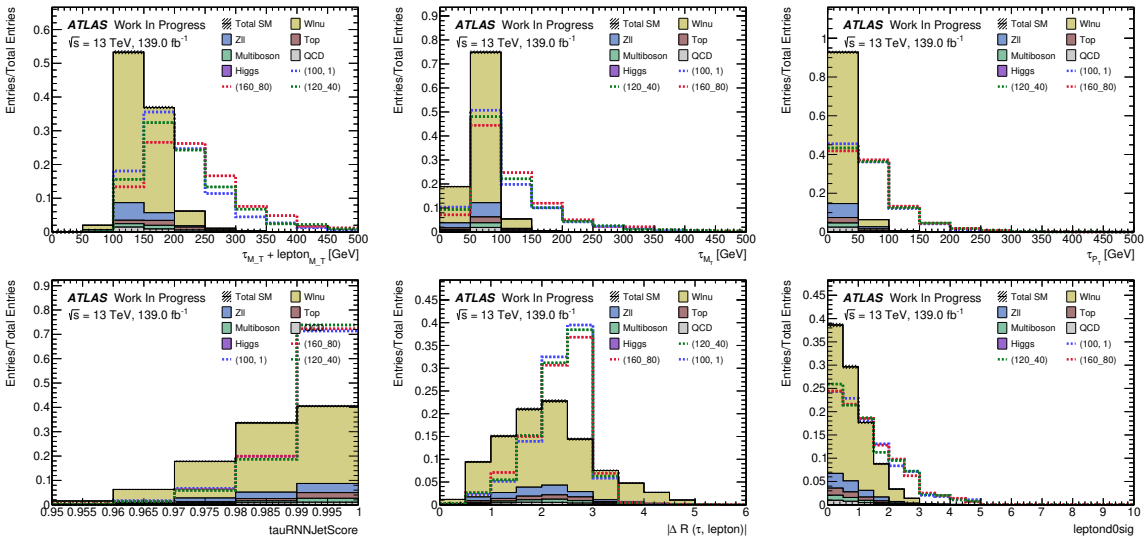
The results given with a training with a grid extension show that it is possible to improve the significances of the HadHad channel when combining it with the LepHad channel. To complete ana analysis of the LepHad channel one of the presented approaches to the MVA method should be chosen and validation regions should be defined.

Appendix

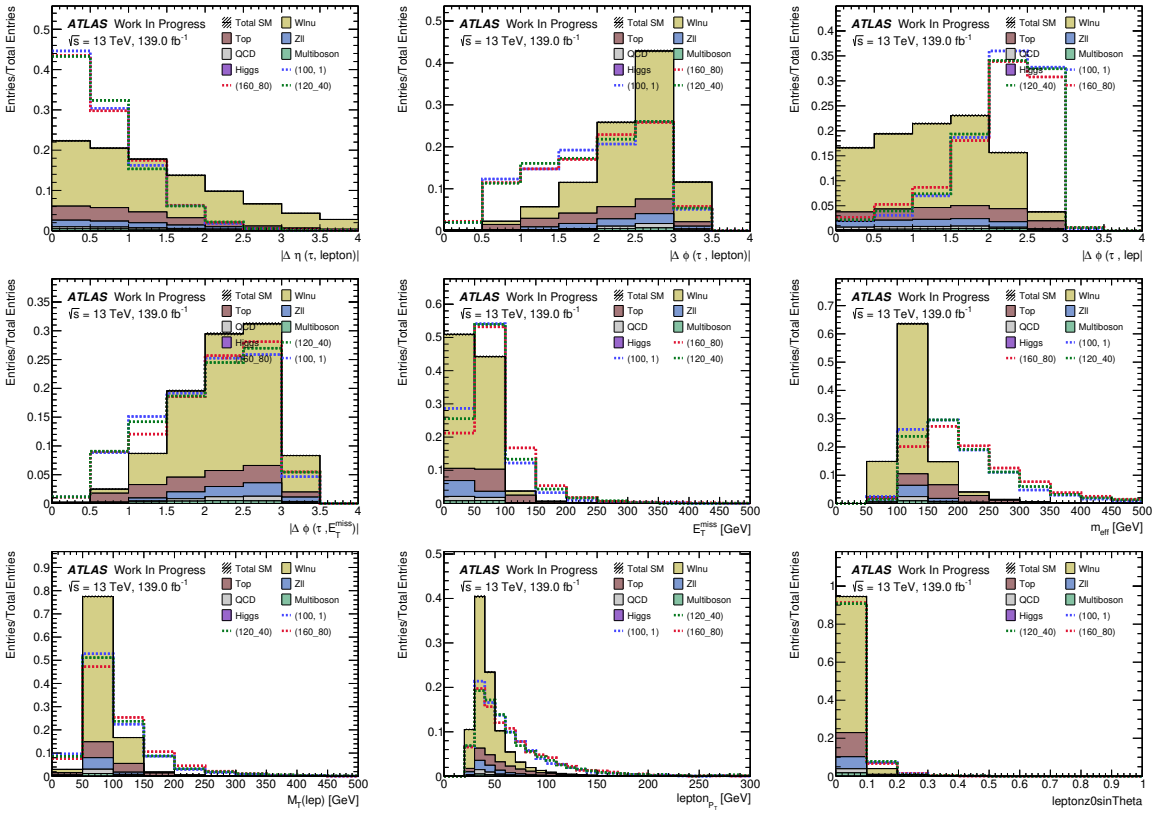
Shape plots of input variables after loose preselection

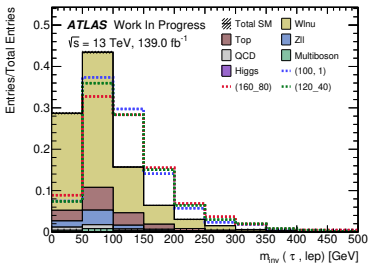
Preselection = 1 tight tau 1 signal lepton, $m_{\mathrm{T2}}(\ell,\tau) > 30\,\mathrm{GeV}$ and b veto

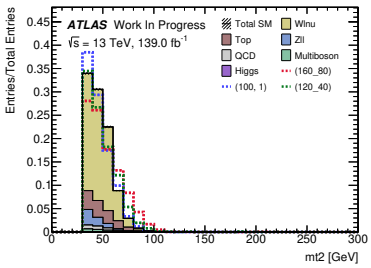


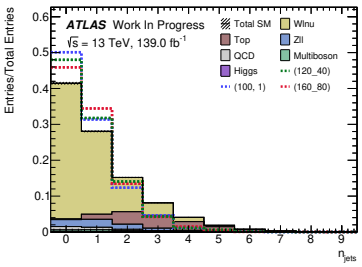


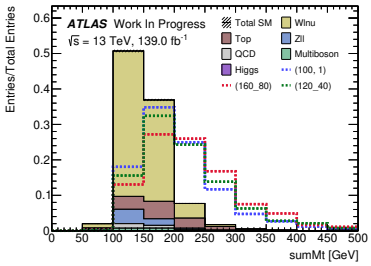
Shape plots of input variables with medium taus and no b-veto

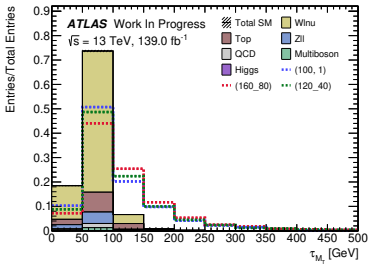


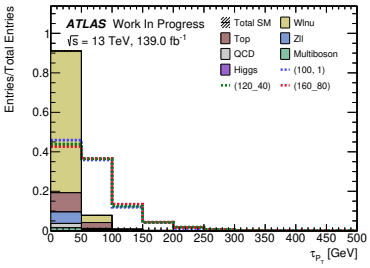

ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(100, 1)
(160_80)
(120_40)
Entries/Total Entries
m_inv (τ , lep) [GeV]

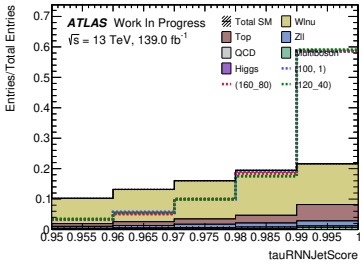

ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(100, 1)
(160_80)
(120_40)
Entries/Total Entries
mt2 [GeV]

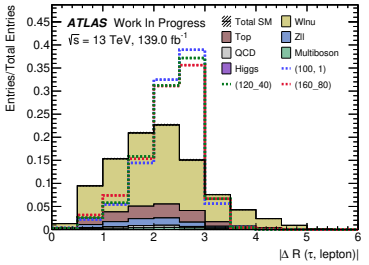

ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(100, 1)
(120_40)
(160_80)
Entries/Total Entries
n_jets

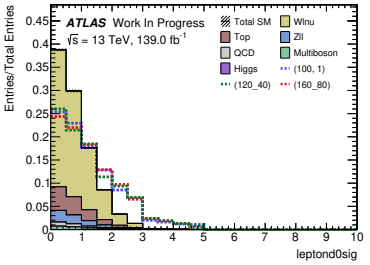

ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(100, 1)
(160_80)
(120_40)
Entries/Total Entries
sumMt [GeV]

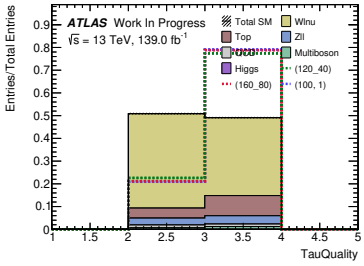

ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(100, 1)
(160_80)
(120_40)
Entries/Total Entries
τ_M_T [GeV]

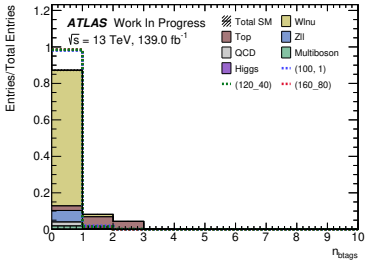

ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(120_40)
(160_80)
(100, 1)
Entries/Total Entries
τ_P_T [GeV]


ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(100, 1)
(160_80)
(120_40)
Entries/Total Entries
tauRNNJetScore


ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(100, 1)
(120_40)
(160_80)
Entries/Total Entries
|Δ R (τ, lepton)|


ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(100, 1)
(120_40)
(160_80)
Entries/Total Entries
leptond0sig

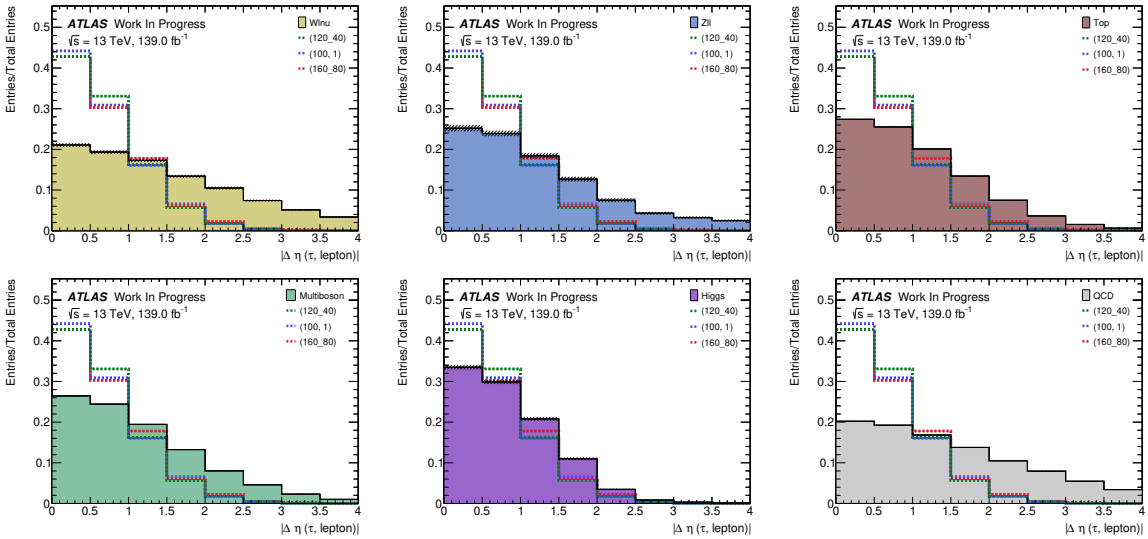

ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(120_40)
(160_80)
(100, 1)
Entries/Total Entries
TauQuality


ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Total SM
Top
QCD
Higgs
Wlnu
Zll
Multiboson
(120_40)
(100, 1)
(160_80)
Entries/Total Entries
n_btags

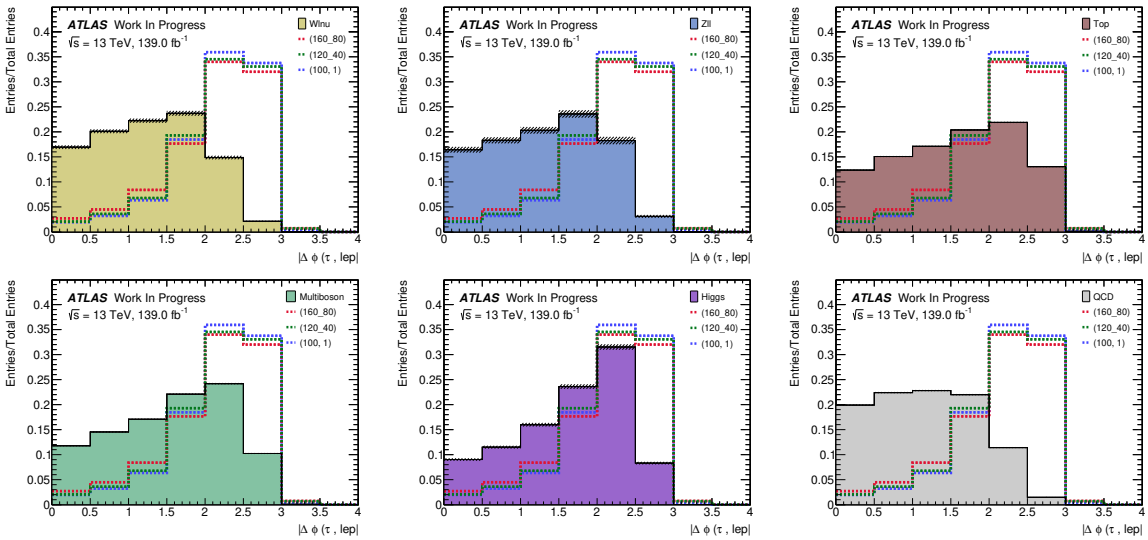
Single Background shape plots

Preselection = 1 tight tau 1 signal lepton, $m_{\mathrm{T2}}(\ell, \tau) > 30\,\mathrm{GeV}$ and b veto

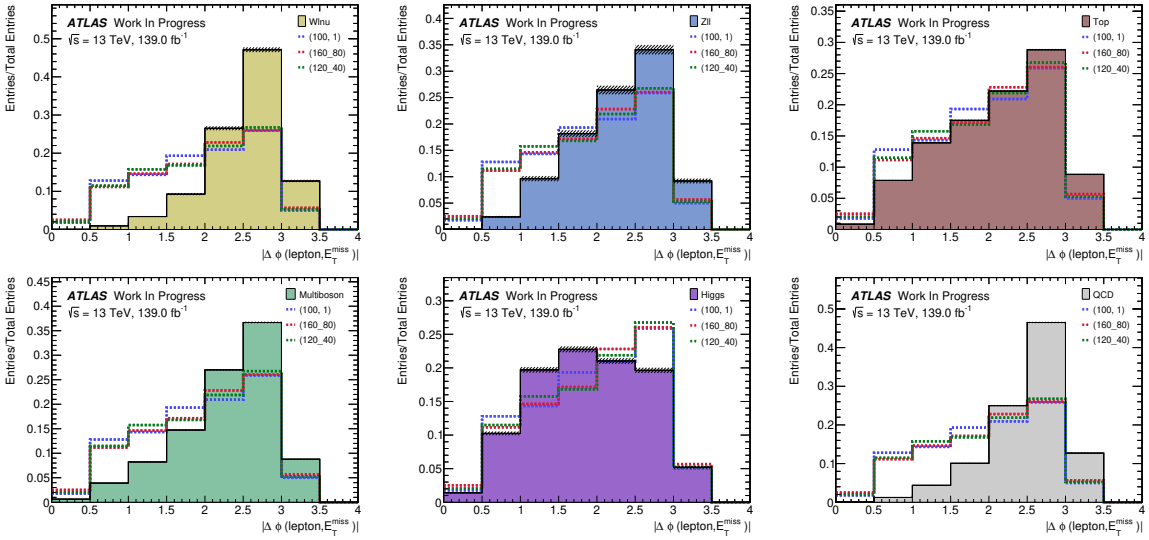
dEtataulepton



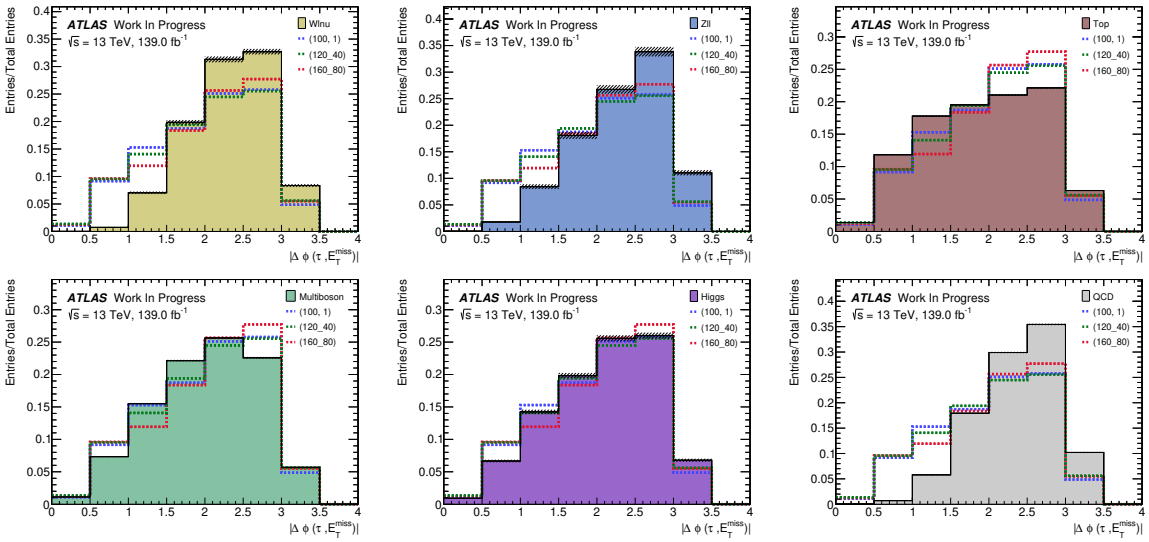
dPhitaulepton



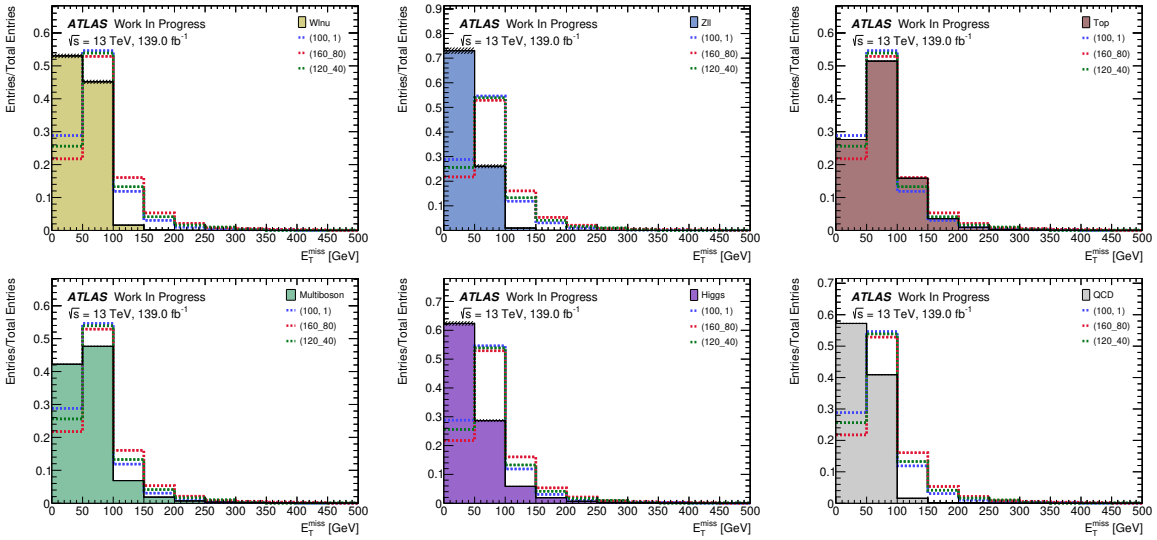
dPhileptonMET



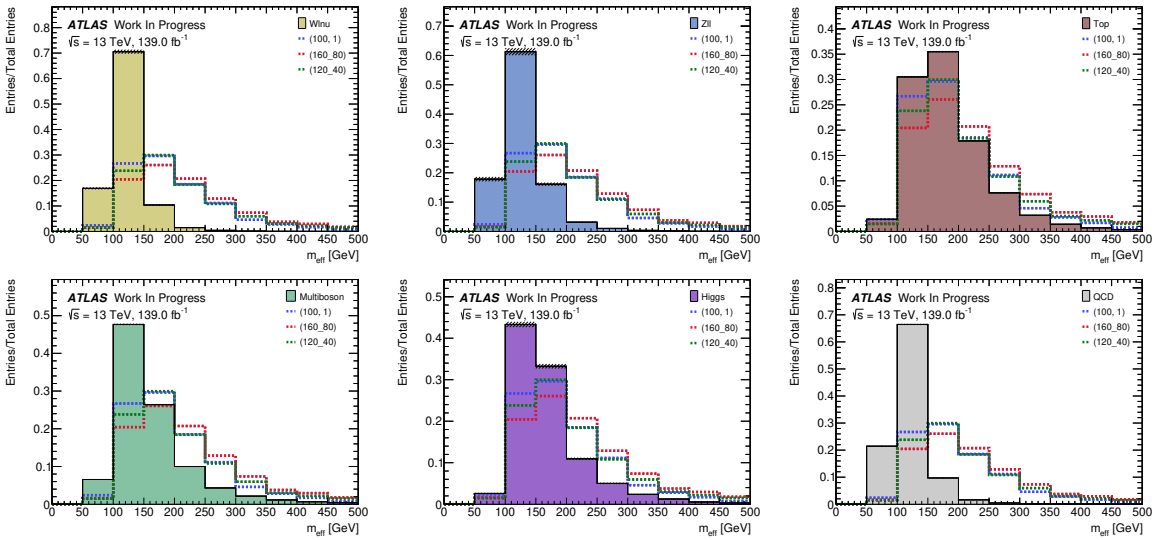
dPhitauMET



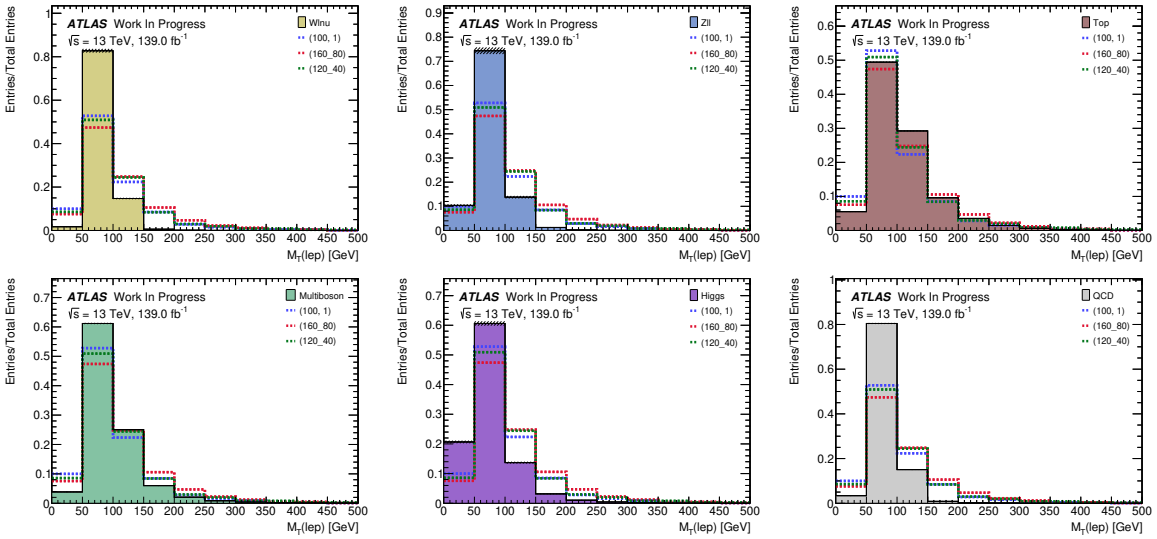
MET


ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Winu
Zll
Top
Multiboson
Higgs
QCD
(100, 1)
(160_80)
(120_40)
Entries/Total Entries
E_T^miss [GeV]

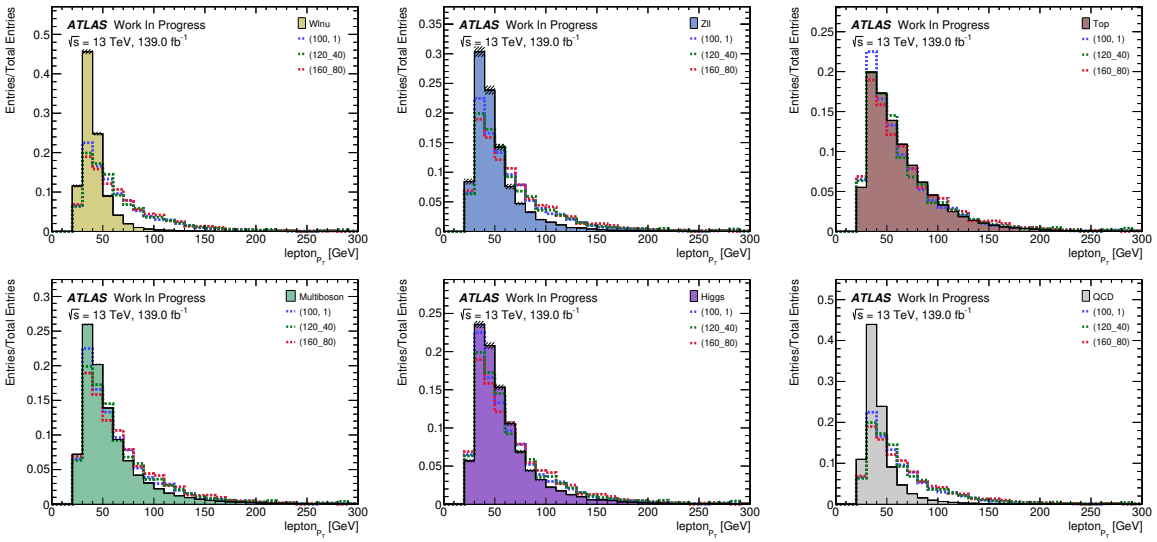
meff


ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Winu
Zll
Top
Multiboson
Higgs
QCD
(100, 1)
(160_80)
(120_40)
Entries/Total Entries
m_eff [GeV]

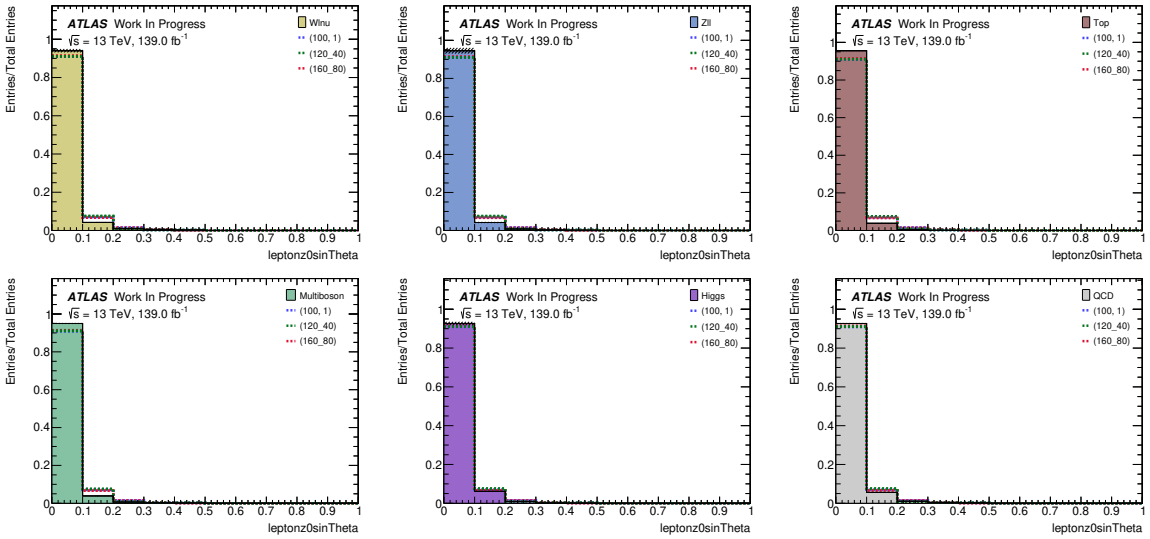
leptonMt



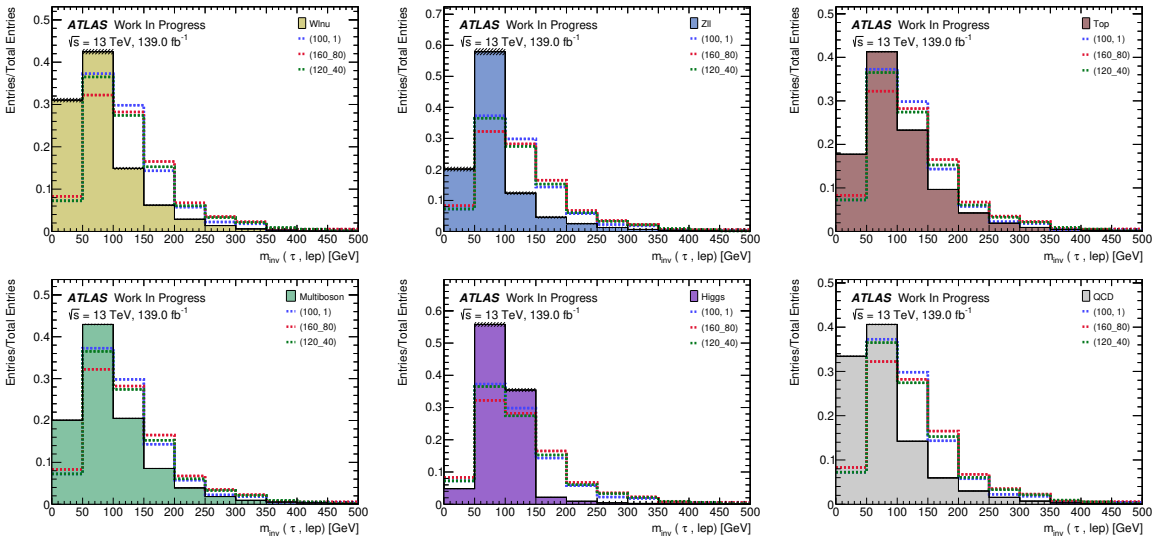
leptonPt



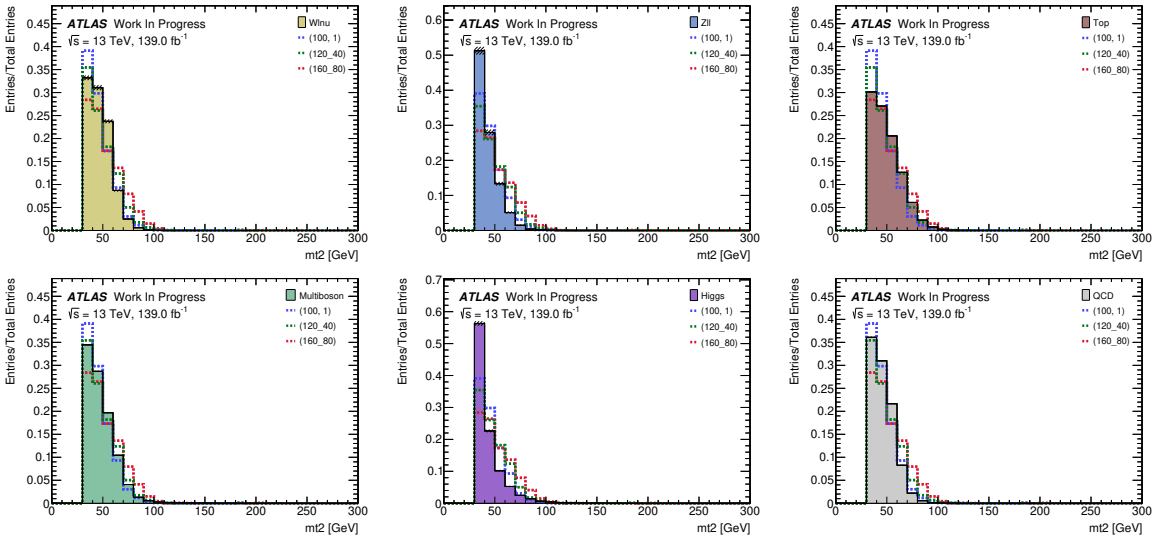
leptonz0sinTheta



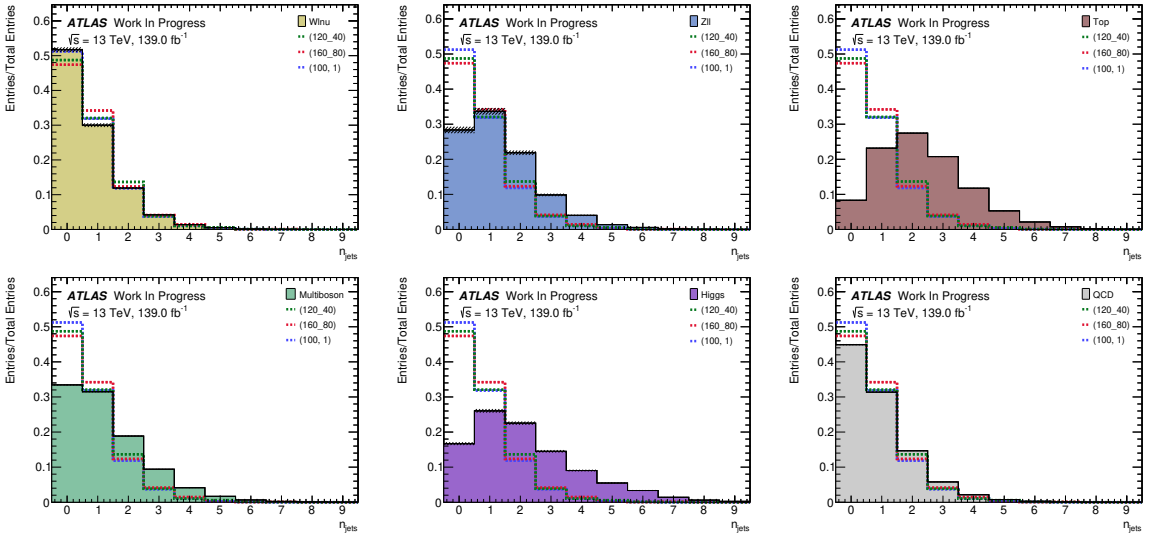
minvtaulepton



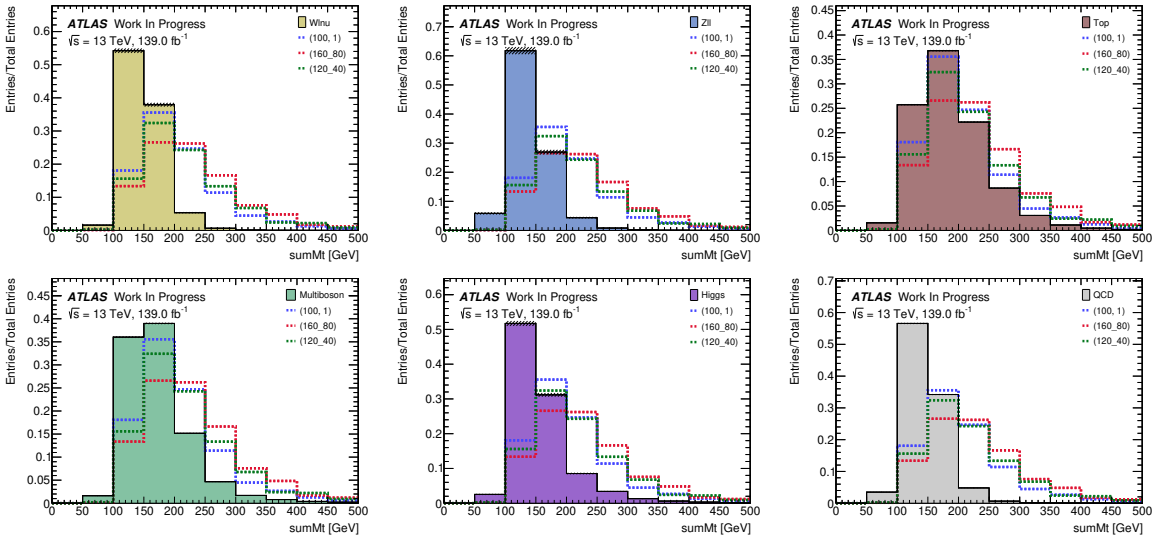
mt2taulepton



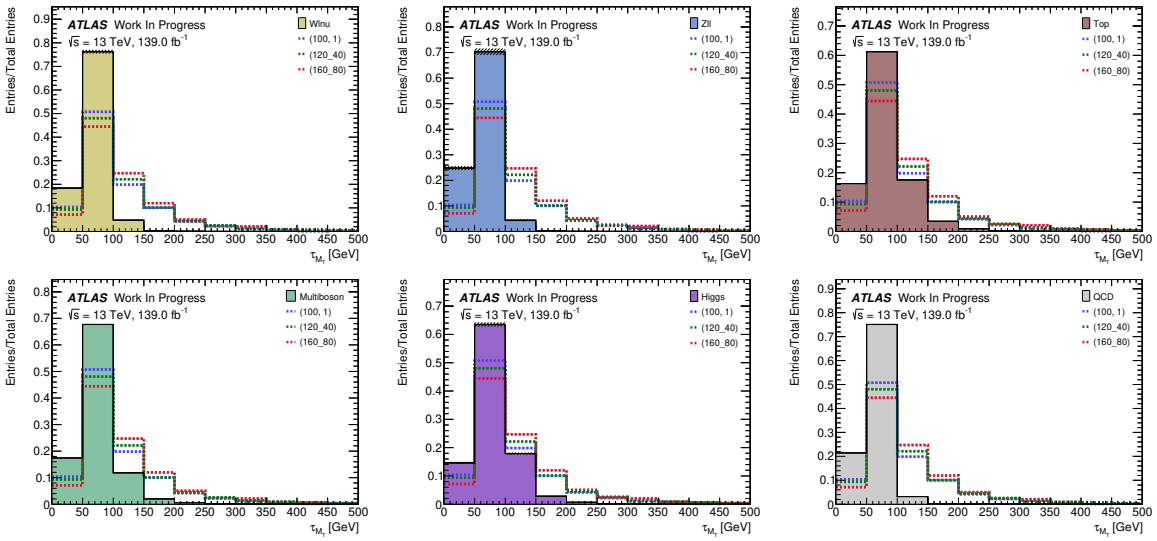
n_jets



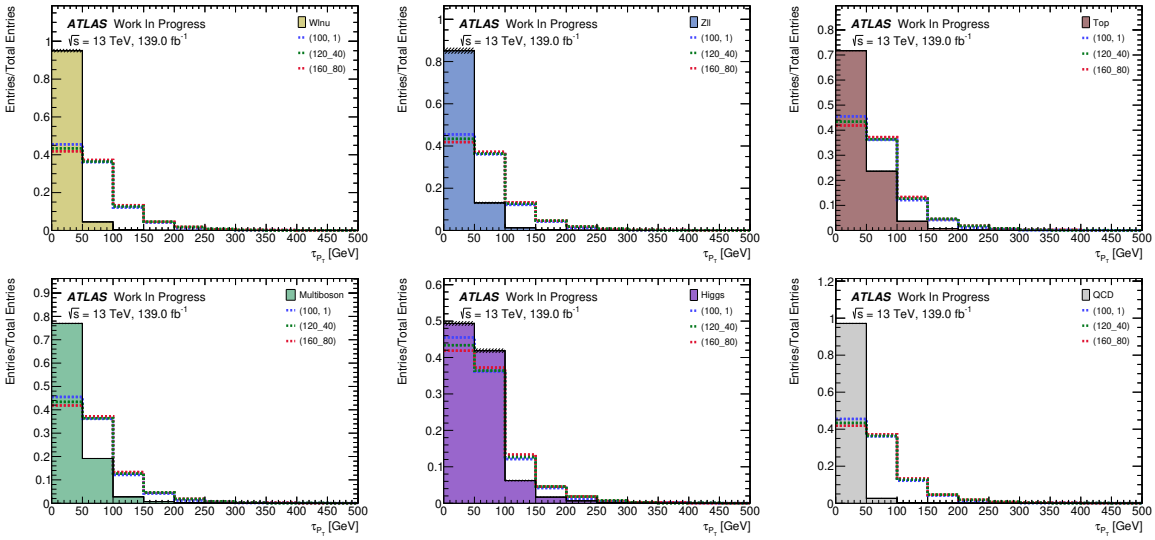
sumMt


ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Wlnu
Zll
Top
Multiboson
Higgs
QCD
(100, 1)
(160_80)
(120_40)
Entries/Total Entries
sumMt [GeV]

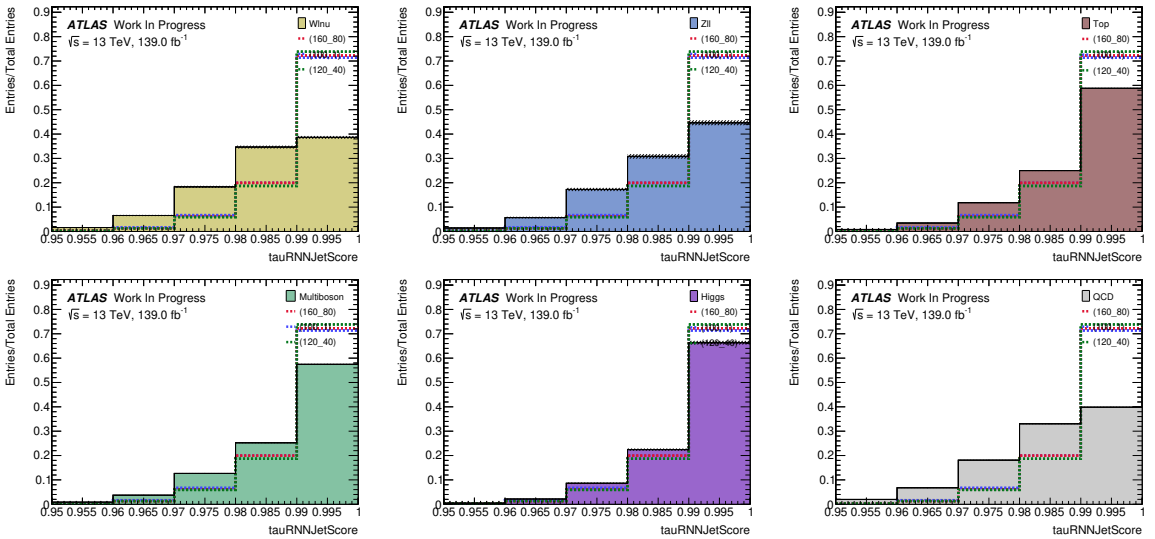
tauMt


ATLAS Work In Progress
√s = 13 TeV, 139.0 fb-1
Wlnu
Zll
Top
Multiboson
Higgs
QCD
(100, 1)
(120_40)
(160_80)
Entries/Total Entries
τMt [GeV]

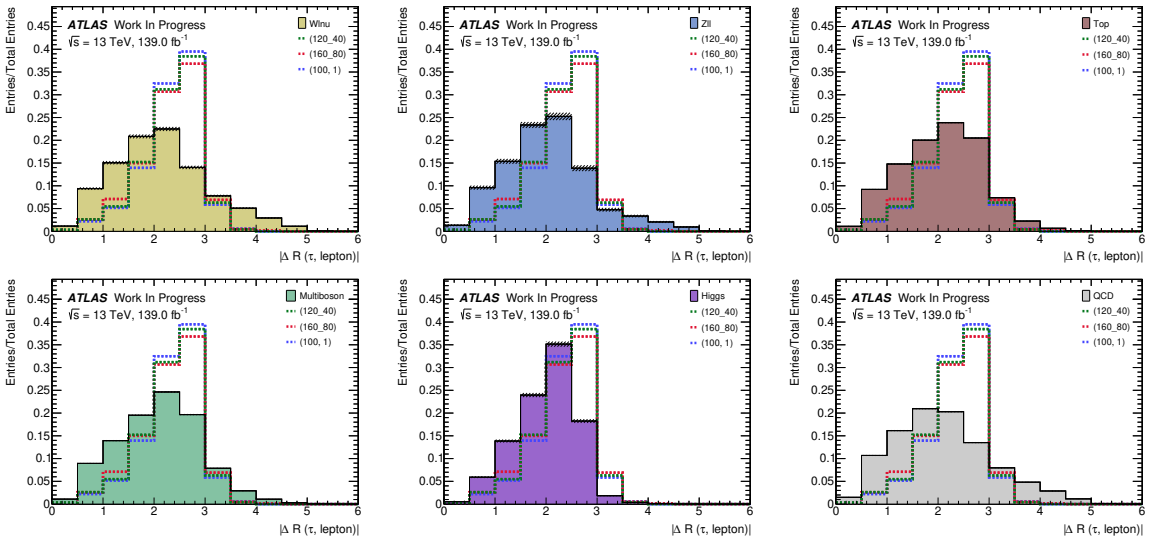
tauPt



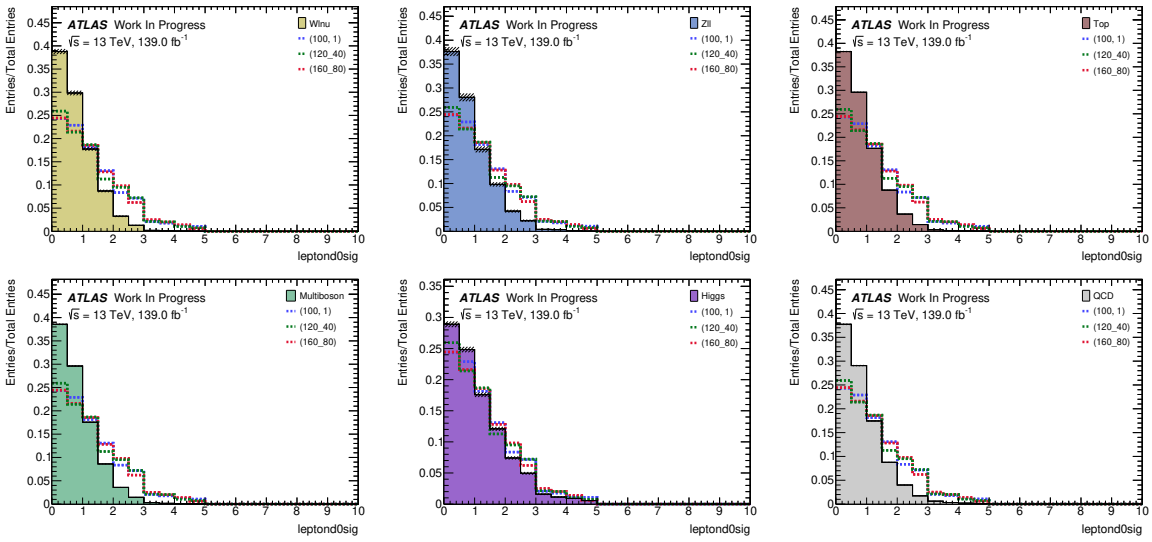
tauRNNJetScore



dRtaulepton



leptond0sig



Bibliography

[1] **Super-Kamiokande Collaboration** Collaboration, Y. Fukuda and T. e. a. Hayakawa, "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.* **81** (Aug, 1998) 1562–1567. https://link.aps.org/doi/10.1103/PhysRevLett.81.1562.

[2] **SNO** Collaboration, Q. Ahmad *et al.*, "Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory," *Phys. Rev. Lett.* **87** (2001) 071301, arXiv:nucl-ex/0106015.

[3] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008. https://cds.cern.ch/record/111880.

[4] S. P. MARTIN, "A supersymmetry primer," *Advanced Series on Directions in High Energy Physics* (Jul, 1998) 1–98. http://dx.doi.org/10.1142/9789812839657_0001.

[5] "The 2004 nobel prize in physics - popular information." https://www.nobelprize.org/prizes/physics/2004/popular-information/.

[6] K. Garrett and G. Duda, "Dark matter: A primer," *Advances in Astronomy* **2011** (2011) 1–22. http://dx.doi.org/10.1155/2011/968283.

[7] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.

[8] K. Müller, "Einführung supersymmetrie,".

[9] "The large hadron collider - timeline." https://timeline.web.cern.ch/timeline-header/93.

[10] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation* **3** no. 08, (Aug, 2008) S08001–S08001. https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001.

[11] R. Steerenberg, M. Albert, R. Alemany-Fernández, T. Argyropoulos, E. Bravin, G. Crockford, J.-C. Dumont, K. Fuchsberger, R. Giachino, M. Giovannozzi, G.-H. Hemelsoet, W. Höfle, D. Jacquet, M. Lamont, E. Métral, D. Nisbet, G. Papotti, M. Pojer, L. Ponce, S. Redaelli, B. Salvachua, M. Schaumann, M. Solfaroli Camillocci, R. Suykerbuyk, G. Trad, J. Uythoven, S. Uznanski, D. Walsh, J. Wenninger, and M. Zerlauth, "Operation and performance of the CERN Large Hadron Collider during proton Run 2,". http://cds.cern.ch/record/2696126.

[12] E. Mobs, "The CERN accelerator complex. Complexe des accélérateurs du CERN,". https://cds.cern.ch/record/2197559. General Photo.

[13] J. Pequenao, "Computer generated image of the whole ATLAS detector." Mar, 2008.

[14] T. A. Collaboration, G. Aad, E. Abat, and J. A. et al., "The ATLAS experiment at the CERN large hadron collider," *Journal of Instrumentation* **3** no. 08, (Aug, 2008) S08003–S08003. https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003.

[15] **ATLAS Collaboration** Collaboration, *ATLAS inner detector: Technical Design Report, 1.* Technical Design Report ATLAS. CERN, Geneva, 1997. http://cds.cern.ch/record/331063.

[16] V. A. MITSOU, "The atlas transition radiation tracker," *Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications* (Jul, 2004) . http://dx.doi.org/10.1142/9789812702708_0073.

[17] A. Ruiz-Martinez and A. Collaboration, "The Run-2 ATLAS Trigger System," Tech. Rep. ATL-DAQ-PROC-2016-003, CERN, Geneva, Feb, 2016. https://cds.cern.ch/record/2133909.

[18] **ATLAS Collaboration** Collaboration, M. zur Nedden, "The Run-2 ATLAS Trigger System: Design, Performance and Plan," Tech. Rep. ATL-DAQ-PROC-2016-039, CERN, Geneva, Dec, 2016. https://cds.cern.ch/record/2238679.

[19] **ATLAS Collaboration** Collaboration, S. Schramm, "ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects," Tech. Rep. ATL-PHYS-PROC-2017-236, CERN, Geneva, Nov, 2017. https://cds.cern.ch/record/2291608.

[20] M. Aaboud, G. Aad, B. Abbott, D. C. Abbott, O. Abdinov, B. Abeloos, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, and et al., "Electron reconstruction and identification in the atlas experiment using the 2015 and 2016 lhc proton–proton collision data at $\sqrt{s} = 13$ tev," *The European Physical Journal C* **79** no. 8, (Aug, 2019) . http://dx.doi.org/10.1140/epjc/s10052-019-7140-6.

[21] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, M. Abolins, O. S. AbouZeid, N. L. Abraham, and et al., "Muon reconstruction performance of the atlas detector in proton–proton collision data at $\sqrt{s}$ s =13 tev," *The European Physical Journal C* **76** no. 5, (May, 2016) . http://dx.doi.org/10.1140/epjc/s10052-016-4120-y.

[22] P. D. Group and P. Z. et al., "Review of Particle Physics," *Progress of Theoretical and Experimental Physics* **2020** no. 8, (08, 2020) , https://academic.oup.com/ptep/article-pdf/2020/8/083C01/33653179/ptaa104.pdf. https://doi.org/10.1093/ptep/ptaa104. 083C01.

[23] **ATLAS Collaboration** Collaboration, "Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment," Tech. Rep. ATL-PHYS-PUB-2019-033, CERN, Geneva, Aug, 2019. http://cds.cern.ch/record/2688062.

[24] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, and et al., "Performance of missing transverse momentum reconstruction with the atlas detector using proton–proton collisions at

$\sqrt{s}$=13 tev," *The European Physical Journal C* **78** no. 11, (Nov, 2018) . http://dx.doi.org/10.1140/epjc/s10052-018-6288-9.

[25] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, "The roostats project," 2011.

[26] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C* **71** no. 2, (Feb, 2011) . http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[27] A. Collaboration, "Search for direct stau production in events with two hadronic τ-leptons in $\sqrt{s} = 13$ tev pp collisions with the atlas detector,".

[28] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. E. Del Valle, and et al., "Search for direct pair production of supersymmetric partners to the τ lepton in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$," *The European Physical Journal C* **80** no. 3, (Mar, 2020) . http://dx.doi.org/10.1140/epjc/s10052-020-7739-7.

[29] C. Lester and D. Summers, "Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders," *Physics Letters B* **463** no. 1, (Sep, 1999) 99–103. http://dx.doi.org/10.1016/S0370-2693(99)00945-4.

[30] M. H. Seymour and M. Marx, "Monte carlo event generators," 2013.

[31] "ahoi framework by nikolai hartmann." https://gitlab.com/nikoladze/ahoi.

[32] H. Binder, O. Gefeller, M. Schmid, and A. Mayr, "The evolution of boosting algorithms," *Methods of Information in Medicine* **53** no. 06, (2014) 419–427. http://dx.doi.org/10.3414/ME13-01-0122.

[33] A. H. et al., "Tmva - toolkit for multivariate data analysis,". https://arxiv.org/abs/physics/0703039.

[34] "Scikit - learn, user guide,". https://scikit-learn.org/stable/user_guide.html.

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

München, 2. November 2020