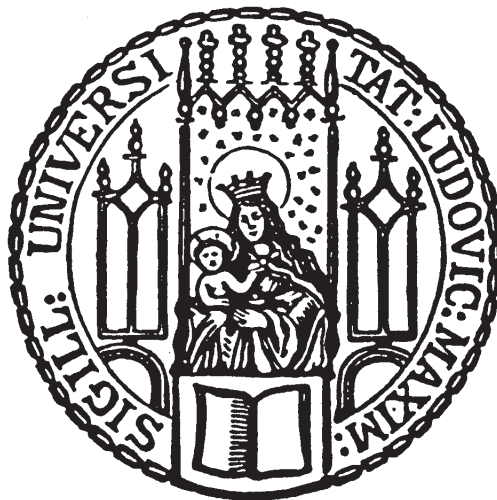


Search for Direct Production of Supersymmetric Scalar Tau  
Leptons Using 13 TeV Data Taken with the ATLAS  
Detector at the LHC

Suche nach direkter Produktion von supersymmetrischen  
skalaren Tau-Leptonen unter Verwendung von  
13-TeV-Daten des ATLAS-Detektors am LHC



Masterarbeit an der Fakultät für Physik  
der  
Ludwig-Maximilians Universität München

vorgelegt von

**Clara Leitgeb**

geboren am 27. August 1994 in Salzburg

München, den 25.10.2017



Gutachterin: Prof. Dr. Dorothee Schaile





# Abstract

The Standard Model of particle physics is very successful in describing the known particles and their fundamental interactions (except gravity). However, there remain some issues to be solved like the Hierarchy problem or the nature of Dark Matter. A possible solution is provided by a (broken) supersymmetric extension of the set of particles. In some supersymmetric models the lightest of these additional particles is assumed to be the lightest of four neutralinos, followed by the supersymmetric partners of the  $\tau$ -lepton. In this thesis, the sensitivity of the ATLAS detector at the LHC for a pair production process of scalar  $\tau$ -leptons with a subsequent decay into a Standard Model  $\tau$ -lepton and a lightest neutralino with a center-of-mass energy of 13 TeV, scaled to an integrated luminosity corresponding to the data collected by ATLAS in 2015 and 2016 ( $36.1 \text{ fb}^{-1}$ ), is studied. Furthermore, two distinct decay channels are taken into account – the hadronic decay of both  $\tau$ -leptons and the case of one hadronic and one leptonic  $\tau$ -decay. For discriminating signal and background a standard cut-and-count approach is used as well as boosted decision trees. A reweighting technique is applied to reduce the statistical uncertainties.



# Contents

<b>1. Introduction</b>	<b>1</b>
<b>2. Theoretical Overview</b>	<b>3</b>
2.1. Standard Model of Particle Physics . . . . .	3
2.2. Supersymmetry . . . . .	7
2.2.1. Dark Matter . . . . .	7
2.2.2. Unification of Interactions . . . . .	8
2.2.3. The Hierarchy Problem . . . . .	8
2.2.4. Construction of a Minimal Supersymmetric Theory . . . . .	9
2.3. Signal Model: Direct Stau Production . . . . .	12
2.3.1. Previous Searches for Direct Stau Production . . . . .	13
<b>3. Experimental Setup</b>	<b>15</b>
3.1. Large Hadron Collider (LHC) . . . . .	15
3.2. ATLAS Experiment . . . . .	17
3.2.1. ATLAS Coordinate System . . . . .	17
3.2.2. Inner Detector (ID) . . . . .	18
3.2.3. Electromagnetic Calorimeter (ECAL) . . . . .	18
3.2.4. Hadronic Calorimeter (HCAL) . . . . .	18
3.2.5. Muon Spectrometer (MS) . . . . .	19
3.2.6. Trigger System . . . . .	19
<b>4. Data and Monte Carlo Simulation</b>	<b>21</b>
4.1. Data . . . . .	21
4.2. Monte Carlo Event Generation . . . . .	21
<b>5. Object Reconstruction and Identification</b>	<b>23</b>
5.1. Electrons . . . . .	23
5.2. Muons . . . . .	24
5.3. Jets . . . . .	25
5.4. Tau Leptons . . . . .	26
5.5. Missing Transverse Energy (MET) . . . . .	27
5.6. Overlap Removal . . . . .	27
5.7. Variable Definitions . . . . .	28
<b>6. Samples</b>	<b>33</b>
6.1. Signal Samples . . . . .	33
6.2. Background Samples . . . . .	34
6.2.1. $W$ +jets . . . . .	34
6.2.2. $Z$ +jets . . . . .	34
6.2.3. Dibosonic Processes . . . . .	34

6.2.4. Top Quarks . . . . .	35
6.2.5. QCD Multijet Processes . . . . .	35
<b>7. Preselection and Triggers</b>	<b>43</b>
7.1. Hadronic Decay of both $\tau$ -Leptons (HadHad-Channel) . . . . .	43
7.2. One Leptonic and One Hadronic $\tau$ -Decay (LepHad-Channel) . . . . .	44
7.3. Comparison of HadHad- and LepHad-Channels . . . . .	44
<b>8. Signal Region Optimization</b>	<b>51</b>
8.1. Statistical Significance . . . . .	51
8.2. Cut-and-Count Method . . . . .	52
8.2.1. HadHad-Channel . . . . .	52
8.2.2. LepHad-Channel . . . . .	58
8.3. Multivariate Analysis . . . . .	61
8.3.1. Boosted Decision Trees . . . . .	61
8.3.2. General Settings for BDTs . . . . .	65
8.3.3. Overtraining . . . . .	66
8.3.4. BDT trained on ( $m_{\tilde{\tau}} = 200$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV) . . . . .	67
8.3.5. BDT trained with High Statistics Sample . . . . .	68
8.3.6. Parameterized BDTs . . . . .	71
8.3.7. Results . . . . .	77
<b>9. Improvement of the <math>W</math>+jets Statistics</b>	<b>87</b>
9.1. Container Tau Promotion Method . . . . .	87
9.2. Performance of the Tau Promotion Method . . . . .	90
9.2.1. $W$ +jets Sample . . . . .	90
9.2.2. QCD Sample . . . . .	90
9.3. Results with Tau Promotion . . . . .	100
9.3.1. Cut-and-Count Method . . . . .	100
9.3.2. Boosted Decision Trees . . . . .	102
<b>10. Summary and Interpretation of Results</b>	<b>111</b>
<b>11. Conclusion</b>	<b>113</b>
<b>A. Appendix I: List of Samples</b>	<b>115</b>
A.1. Signal Samples . . . . .	115
A.2. Background Samples . . . . .	117
<b>B. Appendix II: N-1 Plots</b>	<b>123</b>
B.1. SR DS HadHad Low Mass . . . . .	123
B.2. SR DS HadHad Intermediate Mass . . . . .	125
B.3. SR DS HadHad High Mass . . . . .	127
B.4. SR DS LepHad . . . . .	129
<b>C. Appendix III: Correlation Plots</b>	<b>131</b>
C.1. BDT trained on ( $m_{\tilde{\tau}} = 200$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV) . . . . .	131
C.2. Low Stau Mass BDT . . . . .	134
C.3. High Stau Mass BDT . . . . .	137

---

<b>D. Appendix IV: N-1 Plots with Tau Promotion</b>	<b>141</b>
D.1. SR DS HadHad High Mass with Tau Promotion in $W$ +jets and QCD . . .	141
D.2. SR DS HadHad High Mass with Tau Promotion in $W$ +jets . . . . .	143



# 1. Introduction

The early foundations of the Standard Model of particle physics have already been laid in the late 19<sup>th</sup> century, e.g. with the discovery of the electron by J. J. Thomson in 1897. Since then, many other particles have been postulated and found at various experiments, the most recent triumph being the discovery of the Higgs boson in 2012 by the LHC experiments ATLAS and CMS. However, the search still continues, as there are still many open questions, such as "How can the electroweak force and the strong force be united?", "Why is the Higgs mass not larger than the experimentally determined value?" and "What does Dark Matter consist of?". Answers are provided amongst others by supersymmetric models through introducing partner particles for the Standard Model particles. These partners are assumed to have higher masses than the already known particles, otherwise they would have been within detection reach for a long time at several colliders, e.g. PEP, PETRA, LEP and Tevatron.

Searches at the LHC for strong production mechanisms for SUSY particles so far could shift the exclusion limits for squark and gluino masses beyond 1 TeV [1, 2]. Contrary, the electroweak production of SUSY particles suffers from low cross-sections, which means that the current limits for the masses of charginos, neutralinos and sleptons are mostly considerably lower [3–5]. This is especially the case for direct stau production, as there was almost no sensitivity for the search with the ATLAS detector in run 1. There, difficulties do not only arise due to the small cross-sections but also because of very large Standard Model backgrounds. In run 2 of the LHC, the center-of-mass energy has increased from 8 TeV to 13 TeV and by the end of 2016, 36.1 fb<sup>-1</sup> of data have been collected with the ATLAS detector. These are promising conditions for another search for direct stau pair production. However, there is a drawback, namely the large instantaneous luminosity which leads to high trigger thresholds. Together with the low cross-sections, the small amount of simulated events for the signal but also for some background processes will turn out to be the most limiting factor for sensitivity.





## 2. Theoretical Overview

### 2.1. Standard Model of Particle Physics

In particle physics the main focus is to understand elementary particles and their interactions. The theoretical framework in which all this knowledge is collected is called the Standard Model of particle physics. It is somewhat remarkable to which extent the Standard Model has proven to be accurate in the description of all experimental data so far.

The world around us can be described with just four different particles and four possible interactions between them. In each atom, there are negatively charged electrons surrounding heavy nuclei, which consist of positively charged protons and electrically neutral neutrons. The nucleus is held together by the strong interaction between protons and neutrons. Another particle, the neutrino, is emitted e.g. in a  $\beta$ -decay, when a neutron decays via the weak interaction. The electromagnetic force can be observed in the interaction of charged particles. The fourth force is gravity, which is negligible at atomic scales as it is a lot weaker than the other interactions. When moving to higher energy scales, however, it turns out that only the electron and the neutrino are fundamental particles, as the proton and the neutron have a substructure in the form of three valence quarks, bound via the strong force. The proton consists of two up-quarks and one down-quark, the neutron of two down-quarks and one up-quark. These two quarks, together with the electron and the electron-neutrino form what is called the first generation of particles. In fact, there are two more generations: The second, containing the strange-quark, the charm-quark and the muon together with the muon-neutrino, and the third, with the bottom-quark, the top-quark, the  $\tau$ -lepton and the  $\tau$ -neutrino. These particles are listed in table 2.1 and table 2.2 together with their electric charges and masses.

Quarks and leptons are fermions with spin  $\frac{1}{2}$ . In relativistic quantum mechanics, these particles are described by the Dirac equation which can be derived by factorizing the

generation	particle	el. charge [ $e$ ]	mass [GeV]
1	electron	-1	0.0005
	electron neutrino	0	$< 10^{-9}$
2	muon	-1	0.106
	muon neutrino	0	$< 10^{-9}$
3	$\tau$ -lepton	-1	1.78
	$\tau$ -neutrino	0	$< 10^{-9}$

Table 2.1.: The Standard Model leptons with their charges (in units of the electron charge  $e$ ) and masses [6].

generation	particle	el. charge [ $e$ ]	mass [GeV]
1	down quark	$-1/3$	0.003
	up quark	$+2/3$	0.005
2	strange quark	$-1/3$	0.1
	charm quark	$+2/3$	1.3
3	bottom quark	$-1/3$	4.5
	top quark	$+2/3$	174

Table 2.2.: The Standard Model quarks with their charges (in units of the electron charge  $e$ ) and masses [6].

relativistic energy-momentum relation (with the convention  $c = \hbar = 1$ ):

$$0 = p^\mu p_\mu - m^2 \stackrel{!}{=} (\beta^\kappa p_\kappa + m)(\gamma^\lambda p_\lambda - m) \quad (2.1)$$

Due to the fact that there are no linear terms in  $p$  on the left side of the equation the factors  $\beta^\kappa$  and  $\gamma^\lambda$  must be equal. From the squared terms in  $p$  follow certain conditions for the four components of  $\gamma$ :

$$(\gamma^0)^2 = 1, \quad (2.2)$$

$$(\gamma^{1,2,3})^2 = -1, \quad (2.3)$$

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}, \quad (2.4)$$

where  $g^{\mu\nu}$  is the Minkovski metric. These requirements are fulfilled by a set of four matrices, referred to as  $\gamma$ -matrices:

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, \quad i \in \{1, 2, 3\} \quad (2.5)$$

Here, 1 stands for the  $2 \times 2$  unit matrix, 0 the  $2 \times 2$  matrix of zeros and  $\sigma_i$  are the three Pauli matrices. Equation 2.1 then has two possible solutions:  $\gamma^\kappa p_\kappa + m = 0$  or  $\gamma^\kappa p_\kappa - m = 0$ . Conventionally the latter solution is chosen. Substituting  $p_\mu$  by its quantum mechanical equivalent  $i\hbar\partial_\mu$ , the Dirac equation is obtained, acting on a fermionic wave function  $\psi_e$ :

$$(i\gamma^\mu \partial_\mu - m)\psi_e = 0, \quad (2.6)$$

The quarks and leptons form the fermionic part of the Standard Model. The bosonic part arises through the fields that correspond to the fundamental interactions. The forces acting on the mentioned particles are described by quantum field theories.

All of these quantum field theories are built upon the principle of local gauge invariance. The Dirac equation is invariant under global gauge transformation

$$\psi_e(x) \rightarrow \psi'_e = e^{ie\rho} \cdot \psi_e(x), \quad (2.7)$$

where  $\rho$  is a constant in  $x$ . However, it is not invariant under a so-called local gauge transformation, i.e. if  $\rho = \rho(x)$ . By introducing terms to guarantee local gauge invariance the principle of an interaction arises somewhat naturally. The following term is inserted:

$$\partial^\mu \rightarrow D^\mu = \partial^\mu + iqA^\mu(x). \quad (2.8)$$

If now the wavefunction is transformed the new field  $A^\mu$  receives a transformation at the same time, like the following:

$$\psi_e(x) \rightarrow \psi'_e(x) = e^{ie\rho(x)}\psi_e(x) \quad (2.9)$$

$$A^\mu(x) \rightarrow A'^\mu(x) = A^\mu(x) - \partial^\mu\rho(x) \quad (2.10)$$

with equations 2.9 and 2.10 the new Dirac equation

$$(i\gamma^\mu D_\mu - m)\psi_e = 0, \quad (2.11)$$

preseves its form under local gauge transformation [7].

The field  $A^\mu$  which had to be introduced can be interpreted as a new interaction or a particle that couples to the initial particle  $\psi_e$ . This principle of constructing interaction by introducing local gauge invariance is what every quantum field theory in the Standard Model is based on.

The oldest quantum field theory is quantum electrodynamics (QED) which was developed in the 1940s by Tomonaga, Feynman and Schwinger and is a theory of electromagnetic interaction. This interaction is mediated via the photon and only affects particles which carry electric charge [6].

The theory for the strong interaction is provided by quantum chromodynamics (QCD). Instead of positive and negative charge as in QED, the strong interaction couples to the so-called color-charge which is only carried by quarks and gluons. In nature, only color-neutral states are allowed, which is the reason for quarks always being in a bound state. There are three different colors (red ( $r$ ), blue ( $b$ ) and green ( $g$ )). Therefore also more than one mediator particle is needed. The gluons themselves carry (unlike the electrically neutral photon) one color and one anti-color each. This leads to the "color-octet": Eight gluon states, which correspond to each combination of color and anti-color. The gluon singlet state ( $|r\bar{r}\rangle + |b\bar{b}\rangle + |g\bar{g}\rangle$ ) is not realised in nature. It would be color-neutral and therefore a freely propagating particle allowing for long-range strong interactions, which has not been observed [7].

The weak interaction is mediated through three bosons:  $W_1$ ,  $W_2$  and  $W_3$ , which couple to the weak isospin which is only carried by left-handed particles. The weak isospin of each particle has three components of which each couples to the corresponding W-boson. The linear combinations of  $W_1$  and  $W_2$  form two electrically charged mass-eigenstates  $W^+$  and  $W^-$ .

However, this is not exactly the way the weak force is realised. Instead the weak and the electromagnetic force can be combined in the electroweak interaction. The electric charge  $q_{em}$  and the third component of the weak isospin  $I_3$  are then combined in the weak hypercharge:

$$Y = 2(q_{em} - I_3) \quad (2.12)$$

The bosons coupling to this hypercharge are not simply  $W_3$  and the photon. Instead the QED gauge theory has to be redefined to give rise to a neutral boson  $B_\mu$  which is not identical to the photon  $A_\mu$ . Then the  $W_3$  and the  $B_\mu$  bosons combine to two different mass eigenstates, the photon  $A_\mu$  and the  $Z$ -boson  $Z_\mu$ , defined by the mixing angle  $\theta_W$  (Weinberg angle) [6, 7]:

$$A_\mu = B_\mu \cdot \cos \theta_W + W_\mu^3 \cdot \sin \theta_W \quad (2.13)$$

$$Z_\mu = -B_\mu \cdot \sin \theta_W + W_\mu^3 \cdot \cos \theta_W \quad (2.14)$$

The photon (as well as the gluon) has been found to be massless. In contrast, the  $W^\pm$ - and  $Z^0$ -bosons have masses of 80 GeV and 91 GeV, respectively. The Standard Model as it has been described up to here cannot explain the origin of these masses. What solves this problem, is called the Higgs mechanism, based on spontaneous symmetry breaking. The Higgs field is a complex scalar doublet  $\Phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$  whose components can be generally written as

$$\phi(x) = \frac{1}{\sqrt{2}}(\phi_1(x) + i\phi_2(x)). \quad (2.15)$$

The idea of spontaneous symmetry breaking is now to find a potential with the feature that the ground state obeys a different symmetry than the system itself. A potential which fulfills this condition is this one:

$$V(\Phi) = -\mu^2|\Phi|^2 + \lambda^2|\Phi|^4 \quad (2.16)$$

This potential has a global minimum at a value of  $x$  other than 0 whereas at 0 there is a local maximum. There are now two possibilities for fluctuations around the ground state: Fluctuations in the azimuthal angle and in the radius. Angular fluctuations do not need additional energy, therefore they correspond to a massless Goldstone-boson. Radial fluctuations, in contrast, require additional energy. The particle which arises from this effect is the Higgs boson which, with the above definitions, has a mass of  $m_H = \sqrt{2}\mu$ . The Higgs boson has been found at the LHC in 2012. The reason that up to now no Goldstone-boson has been found is that the corresponding three degrees of freedom are "swallowed" by mixing with three of the electroweak bosons  $W^\pm$  and  $Z^0$ , thereby generating their masses. All other massive particles in the Standard Model, i.e. the fermions, receive their masses by coupling to the Higgs field (Yukawa coupling). However, the Higgs mechanism does not explain the mass hierarchy of these particles [7].

## 2.2. Supersymmetry

Although the Standard Model of particle physics provides a very accurate description of the known phenomena in high-energy physics there are still some open questions. There are many hints that the Standard Model has to be modified and extended to give a precise description of physics at and beyond the Planck scale ( $M_P = (8\pi G_{Newton})^{-\frac{1}{2}} = 2.4 \cdot 10^{18}$  GeV) [8].

### 2.2.1. Dark Matter

In astronomy it has been assumed for a long time, that there was a relation between the light which is received from an object and its mass. The first hint that this assumption does not necessarily have to be true, was shown in the early 1930s by J.H. Oort when he studied the Doppler shifts of stars in the center of the Milky Way. The velocities he concluded on turned out to be too high for the stars to remain in their orbits under the gravitational force of the luminous mass of the galaxy. A similar observation was made by F. Zwicky, who observed the Doppler shifts of the galaxies in the Coma cluster. Certainly, there was the possibility that the observations were not precise enough. However, in the 1970s V. Rubin studied 60 isolated galaxies and saw an astounding deviation from the expected "Keplerian" rotation profile of the galaxies [9].

Also in the late 1970s D. Walsh found "missing matter" in a different way: Gravitational lensing. There, the light from a distant luminous object is deflected by a closer, massive object, leading to so-called Einstein rings. From the radii of these rings one can conclude on the mass that causes the deflection. However, Walsh found that the calculated mass was much larger than one would expect from the luminous matter forming the gravitational lense [9].

Finally, it can be stated that from a cosmological point of view there is enough evidence for the existence of dark matter. What is still unclear is the nature of dark matter. The Standard Model does not provide any candidates for dark matter particles. Such particles should be electrically neutral, only interacting via the gravitational and the weak force and be sufficiently heavy. The Standard Model neutrino is indeed electrically uncharged and interacting weakly but is not able to explain structure formation in the early universe [9].

However, three additional dark matter candidates are provided by supersymmetry: The neutralino (Section 2.2.4) and the supersymmetric partners of the neutrino and the graviton (for a quantum field theory of gravity). All three of them are electrically neutral and weakly interacting. But the sneutrino is expected to have too high annihilation rates to be considered as a dark matter particle. Gravitinos (according to some models) may also be too light to explain dark matter [9]. In the model considered in this thesis, the best dark matter candidate is the (lightest) neutralino, as it is expected to be sufficiently stable (because it is assumed to be the lightest supersymmetric particle (see sections 2.2.4 and 2.3)).

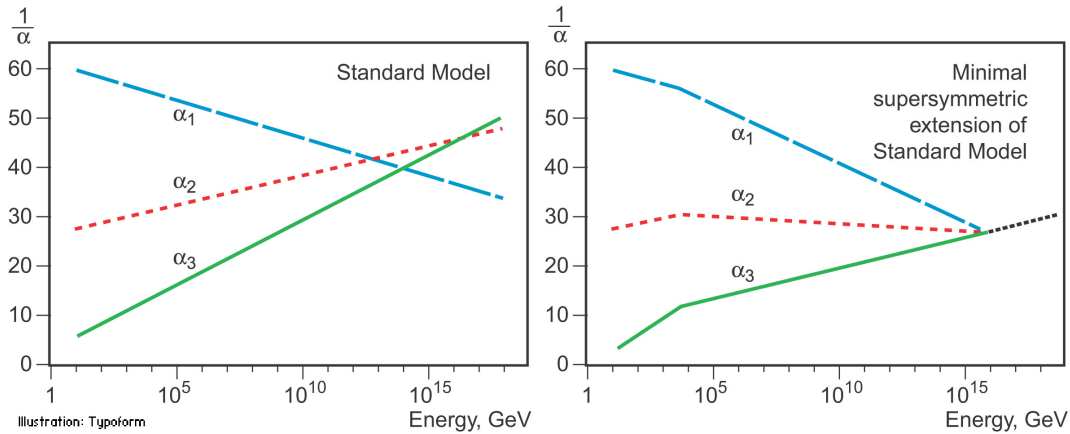


Figure 2.1.: The left diagram shows the energy dependence of the coupling parameters of the three fundamental interactions (electromagnetic  $\alpha_1$ , weak  $\alpha_2$  and strong  $\alpha_3$ ) as predicted by the Standard Model. There is no energy scale at which all of the couplings intersect, therefore not allowing for a common description. In the right figure the situation is shown for a supersymmetric extension of the Standard Model. Here, a unification of the interactions becomes possible [10].

### 2.2.2. Unification of Interactions

As already mentioned in the previous section, the Standard Model relies on three forces being responsible for the interaction between particles. The couplings of these forces can be very different. However, the couplings are no constants but depend on the energy scale. This is referred to as the running of the coupling constants. At some point of the energy scale, the coupling constants can merge, possibly resulting in the unification of the corresponding forces. So it is possible to combine the electromagnetic and the weak force as described previously. Of course, it would be a further success for particle physics to combine the electromagnetic, the weak and the strong force all in one interaction. As it can be seen in figure 2.1 this is not possible, as the three couplings as predicted by the Standard Model do not intersect at one single energy scale. However, if there was physics beyond the Standard Model (like Supersymmetry), it could possibly correct for this, allowing for a unification of all three fundamental forces.

### 2.2.3. The Hierarchy Problem

In the Standard Model all particles receive their masses through interaction with a complex scalar field, the Higgs field  $\Phi$ , whose classical potential can be written as in equation 2.16.

In the Standard Model this potential has an expectation value  $\langle \Phi \rangle = \sqrt{-\mu^2/2\lambda} > 0$  at its minimum. In this case the parameters  $\mu$  and  $\lambda$  have to obey the following relations:

$$\mu^2 < 0 \tag{2.17}$$

$$\lambda > 0 \tag{2.18}$$

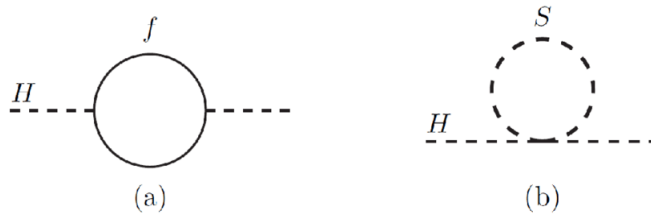


Figure 2.2.: Depicted here are the quantum loop corrections to the parameter  $\mu^2$  caused by a Dirac fermion (a) and a scalar boson (b) [11].

In 2012 an electrically neutral, scalar particle was found at the LHC which fulfills all the expected properties of the theoretically predicted Higgs boson. As its mass is found to be approximately at 125 GeV the parameters of the Higgs potential are calculated as  $\mu^2 = -(92.9 \text{ GeV})^2$  and  $\lambda = 0.126$ . But, taking higher orders of Feynman diagrams into account, the parameter  $\mu^2$  is corrected by loop diagrams stemming from each particle that couples either directly or indirectly to the Higgs field. These corrections to  $\mu^2$  are proportional to the coupling  $\lambda_f^2$  to a fermion  $f$  and therefore also proportional to  $m_f^2$ . Taking all of these corrections together and given a cutoff of the loop integrals in the order of the Planck scale  $M_P$  the resulting Higgs squared mass parameter  $\mu^2$  is about 30 orders of magnitude larger than the measured  $-(92.9 \text{ GeV})^2$ . This is referred to as the hierarchy problem.

One rather ingenious solution to the hierarchy problem is to assume the existence of new particles whose quantum corrections cancel out the loop integrals induced by the known Standard Model particles. As fermionic loop diagrams add positive contributions to  $\mu^2$  and complex scalar particles lead to negative contributions, the fermionic loop diagrams could be cancelled out by scalar particles and vice versa (see figure 2.2). This could be achieved by introducing a new symmetry relating fermions and bosons, as it is done with supersymmetry [8].

#### 2.2.4. Construction of a Minimal Supersymmetric Theory

The symmetry between bosons and fermions is constructed by introducing a transformation  $Q$  which converts a fermionic state into a bosonic state and vice versa:

$$Q|\text{Fermion}\rangle = |\text{Boson}\rangle \quad (2.19)$$

$$Q|\text{Boson}\rangle = |\text{Fermion}\rangle \quad (2.20)$$

Furthermore,  $Q$  must satisfy the following relations:

$$\{Q, Q^\dagger\} = P^\mu \quad (2.21)$$

$$\{Q, Q\} = \{Q^\dagger, Q^\dagger\} = 0 \quad (2.22)$$

$$[P^\mu, Q] = [P^\mu, Q^\dagger] = 0 \quad (2.23)$$

Here,  $P^\mu$  is the four-momentum generator of translations in spacetime. These relations define the supersymmetric algebra. One can now construct irreducible representations

Names	spin 0	spin 1/2	Names	spin 1/2	spin 1
squarks, quarks	$(\tilde{u}_L, \tilde{d}_L)$	$(u_L, d_L)$	gluino, gluon	$\tilde{g}$	$g$
	$\tilde{u}_{R^*}$ $\tilde{d}_{R^*}$	$u_R^\dagger$ $d_R^\dagger$			
sleptons, leptons	$(\tilde{\nu}, \tilde{e}_L)$	$(\nu, e_L)$	winos, W bosons	$\tilde{W}^\pm, \tilde{W}^0$	$W^\pm, W^0$
	$\tilde{e}_{R^*}$	$e_R^\dagger$			
Higgs, Higgsinos	$(H_u^+, H_u^0)$ $(H_d^0, H_d^-)$	$(\tilde{H}_u^+, \tilde{H}_u^0)$ $(\tilde{H}_d^0, \tilde{H}_d^-)$	bino, B boson	$\tilde{B}^0$	$B^0$

Table 2.3.: The chiral and gauge supermultiplets in the minimal supersymmetric extension of the Standard Model (MSSM). It is important to keep in mind that the helicity indices ( $R$ ,  $L$ ) of the squarks and sleptons are simply inherited from their Standard Model partners and do not refer to any helicity state as they are spin-0 particles [8].

of the supersymmetric algebra, which are called supermultiplets and contain a Standard Model particle and its supersymmetric partner (superpartner) each. The fermionic and the bosonic states inside the supermultiplet are then related by  $Q$  and  $Q^\dagger$ . Because  $-P^2$  commutes with both  $Q$  and  $Q^\dagger$ , both states have the same eigenvalues of  $-P^2$  and consequently their masses should be equal. As  $Q$  and  $Q^\dagger$  also commute with the gauge transformation generators, the supersymmetric particles also have the same electric charge, weak isospin and color as their Standard Model partners [8].

One can distinguish two main forms of supermultiplets: The chiral and the gauge supermultiplets. Chiral supermultiplets consist of one Weyl fermion and two real scalar particles. The Standard Model fermions and the extended Higgs sector together with their superpartners are assembled in this kind of supermultiplet. Gauge supermultiplets are composed of one spin-1 vector boson and a spin-1/2 Weyl fermion. The Standard Model gauge bosons together with their supersymmetric partners, the gauginos, are combined in a gauge supermultiplet. The supermultiplets in the minimal supersymmetric extension of the Standard Model (MSSM) are summarized in table 2.3.

As already mentioned, the Standard Model Higgs boson is the scalar part of a chiral supermultiplet. One reason to introduce two Higgs supermultiplets is that otherwise a gauge anomaly would arise in the electroweak gauge symmetry [8]. The fermionic partner of the scalar Higgs must be a weak isodoublet like the leptonic isodoublet in the Standard Model, with one part carrying a weak hypercharge of  $Y = +1/2$  ( $H_u$ ) and the other  $Y = -1/2$  ( $H_d$ ). The  $H_u$  supermultiplet then provides the Yukawa couplings for the up-type quarks and the supermultiplet with  $H_d$  couples to the down-type quarks and to the charged leptons. The weak isospin components of  $H_u$  carry electric charges  $(+1, 0)$  whereas for  $H_d$  the weak isospin components have electric charges  $(0, -1)$ . The neutral, scalar Higgs boson  $h$  which is part of the established Standard Model is a linear combination of the neutral parts of both supermultiplets. There are four additional degrees of freedom in the extended Higgs sector: Two charged Higgs bosons,  $H^\pm$ , and two neutral Higgs bosons, denoted as  $H^0$  and  $A^0$  [8]. The supersymmetric partners of the Higgs bosons can mix with the binos and winos, resulting in eight new mass eigenstates, called charginos  $\tilde{\chi}_{1,2}^\pm$  and neutralinos  $\tilde{\chi}_{1,2,3,4}^0$ .



An important statement is that the supersymmetric partners of the Standard Model fermions listed in table 2.3 are not necessarily identical to the mass eigenstates. In fact, there can be mixing between the partners of the right- and left-handed quarks and leptons, leading to pairs of squarks and sleptons, e.g.  $\tilde{t}_1$  and  $\tilde{t}_2$  or  $\tilde{\tau}_1$  and  $\tilde{\tau}_2$ .

Taking the MSSM, the corresponding superpotential would be [8]:

$$W_{\text{MSSM}} = \bar{u}\mathbf{y}_u QH_u - \bar{d}\mathbf{y}_d QH_d - \bar{e}\mathbf{y}_e LH_d + \mu H_u H_d \quad (2.24)$$

with  $H_u$ ,  $H_d$ ,  $Q$ ,  $L$ ,  $\bar{u}$ ,  $\bar{d}$  and  $\bar{e}$  as the superfields which correspond to the chiral supermultiplets in table 2.3.  $\mathbf{y}_{u,d,e}$  are  $3 \times 3$  matrices containing the Yukawa couplings for each particle family. Phenomenologically, the superpotential in equation 2.24 is sufficient, however, additional terms are allowed, some of them violating baryon number ( $B$ ) or lepton number ( $L$ ). This could lead e.g. to allowed decay channels for the proton, whose decay is not observed in nature. Therefore, the existence of those terms is avoided by introducing another symmetry, called R-parity. R-parity is a multiplicatively conserved quantum number and defined as in the following:

$$P_R = (-1)^{3(B-L)+2s} \quad (2.25)$$

where  $s$  is the particle's spin. This definition is useful in the sense that all Standard Model particles and the Higgs bosons have  $P_R = +1$ , whereas all supersymmetric particles have  $P_R = -1$ . This fact implies that the lightest particle with  $P_R = -1$ , i.e. the lightest supersymmetric particle (LSP), has to be stable. If it is electrically neutral as well, it provides a suitable dark matter candidate. Furthermore, heavier supersymmetric particles can only decay via another supersymmetric particle, in order to conserve R-parity. Finally, supersymmetric particles can only be produced in pairs in collisions of Standard Model particles [8].

It has been stated before that in supersymmetry as it is, the supersymmetric particles are expected to have the same mass as their Standard Model partners. However, up to the submission of this thesis, no supersymmetric particle has been found. This leads to the conclusion that supersymmetry must be a broken symmetry. In order to still provide a solution to the hierarchy problem, the relations of the couplings for each pair of particles has to stay the same. That means that supersymmetry breaking has to be introduced in a "soft" way, leading to an effective Lagrangian:

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}} \quad (2.26)$$

where  $\mathcal{L}_{\text{SUSY}}$  is conserving supersymmetry and the whole supersymmetry breaking happens in the  $\mathcal{L}_{\text{soft}}$ -part. This could be implemented via a spontaneous symmetry breaking, i.e.  $\mathcal{L}_{\text{soft}}$  is invariant under supersymmetry but its ground state is not. So, if the largest mass of the soft terms in  $\mathcal{L}_{\text{soft}}$  is  $m_{\text{soft}}$  then the loop corrections of the squared Higgs mass parameter  $\mu^2$  must vanish in the limit  $m_{\text{soft}} \rightarrow 0$ , so the hierarchy problem is still solved. This also means that the masses of the supersymmetric particles should not be much larger than the Standard Model particles, if soft supersymmetry breaking is given. For this reason there is still a good chance to discover supersymmetry at the Large Hadron Collider (LHC) at CERN [8].

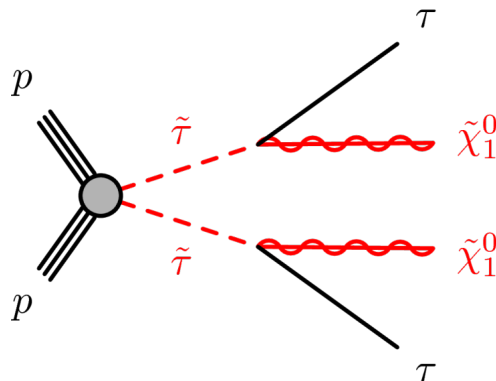


Figure 2.3.: After the direct production of two scalar  $\tau$ -leptons each of them decays into a Standard Model  $\tau$ -lepton and a lightest neutralino, which is assumed to be the LSP, and contributes to missing transverse energy. The particular decay channel depends on the decays of the two  $\tau$ -leptons [13].

### 2.3. Signal Model: Direct Stau Production

Many SUSY searches are based on so-called simplified models. This means, the considered process is limited to a certain number of steps in the decay-chain and the branching ratio is assumed to be 100%, i.e. no other decays are taken into account. The process which is discussed in the following is direct pair production of staus ( $\tilde{\tau}$ ) which is part of the pMSSM (phenomenological minimal supersymmetric extension to the Standard Model [12]). The production of stau pairs can happen via a Drell-Yan process. The staus then decay into one Standard Model  $\tau$ -lepton and one lightest neutralino ( $\tilde{\chi}_1^0$ ) each (see figure 2.3). The neutralino cannot decay further, as it is assumed to be the LSP. Because it is neutral and can only interact via the weak interaction it will pass the detector unnoticed, causing an imbalance of the total momentum of all produced particles (also called missing transverse energy, see section 5.5). For the  $\tau$ -leptons there are three different possible decay-modes. Either both  $\tau$ -leptons decay hadronically or both leptonically or one decays leptonically and the other hadronically. These decay-modes are referred to as the HadHad-channel for the doubly hadronic decay and the LepHad-channel for the one-leptonic and one-hadronic decay. The doubly leptonic decay will not be discussed in these studies as it has the smallest branching fraction.

As already mentioned, the left-handed  $\tau$ -leptons have a SUSY partner as well as the right-handed  $\tau$ -leptons. Because the handedness is no property of the scalar staus, both can mix, depending on a mixing angle  $\theta_\tau$ :

$$\tilde{\tau}_1 = \cos \theta_\tau \cdot \tilde{\tau}_L + \sin \theta_\tau \cdot \tilde{\tau}_R \quad (2.27)$$

$$\tilde{\tau}_2 = \cos \theta_\tau \cdot \tilde{\tau}_R - \sin \theta_\tau \cdot \tilde{\tau}_L \quad (2.28)$$

In the model considered here, it is assumed that  $\tilde{\tau}_1$  and  $\tilde{\tau}_2$  are produced with different cross-sections but are mass degenerate. In this scenario it is assumed that  $\theta_\tau = 0$ , i.e.  $\tilde{\tau}_1$  is identical to the superpartner of the left-handed  $\tau$ -lepton and  $\tilde{\tau}_2$  to that of the right-handed  $\tau$ -lepton. The cross-sections themselves are dependent on the stau mass and range from

0.07 pb to 0.8 pb for stau masses between 80 GeV and 260 GeV. These are listed in table A.1 and table A.2.

There are several Standard Model processes which either have or fake the same signature as the signal-process (two  $\tau$ -leptons and missing transverse energy). The production of two vector bosons ( $WW$  or  $ZZ$ ) with a subsequent decay ( $W \rightarrow \tau\nu_\tau$ ,  $Z \rightarrow \tau\tau/\nu\nu$ ) results in the same detector signature as the direct stau process. These processes are combined into a class of background processes called diboson-processes. Other important processes are  $W + \text{jets}$  and  $Z + \text{jets}$ , meaning that one  $W$ - or  $Z$ -boson is produced together with a certain number of jets, which in the case of a  $W$ -boson can imitate a  $\tau$ -lepton. Processes with top-quarks are also taken into account, as well as multijet processes.

There are multiple motivations for a search for staus. In some SUSY models the stau is the next-to-lightest supersymmetric particle and therefore would be one of the first supersymmetric particles in the reach of the LHC. Furthermore, if the charginos and the three heavier neutralinos have too high masses to be produced at the LHC, the stau pair production would be the dominant electroweak supersymmetric production process. Moreover, there are models for light-neutralino dark matter, in which the dark matter relic density depends on the annihilation of neutralinos via a t-channel stau exchange, making the stau mass one of the defining parameters of these models [14].

### 2.3.1. Previous Searches for Direct Stau Production

The direct production of a stau pair has also been searched for with the ATLAS detector in LHC run 1 at a center-of-mass energy of 8 TeV. There already it was difficult to choose a selection providing sufficient discrimination power to the Standard Model background processes, due to the low cross-sections. As no exclusion of mass points could be made, upper limits on the signal strength were calculated for  $\tilde{\tau}_L$  and  $\tilde{\tau}_R$  separately (see figure 2.4). The signal strength is a factor the theoretical cross-section has to be multiplied with in order to exclude the mass point at 95% confidence level (CL) (see section 8.1). The best upper limit was observed for the mass point  $m(\tilde{\tau}_R) = 90.6$  GeV ( $m(\tilde{\tau}_L) = 93.1$  GeV) and  $m(\tilde{\chi}_1^0) = 0$  GeV. But even for these mass points the cross-section would have to be a factor 3.1 (1.6) higher to exclude them [5].

Furthermore, the CMS experiment performed a search for direct stau production in run 2 at a center-of-mass energy of 13 TeV, which could not exclude any signal mass points, instead setting an upper limit on the cross-section for the production of staus with a mass of 125 GeV and nearly massless neutralino at 1.5 times the theoretical cross-section [15].

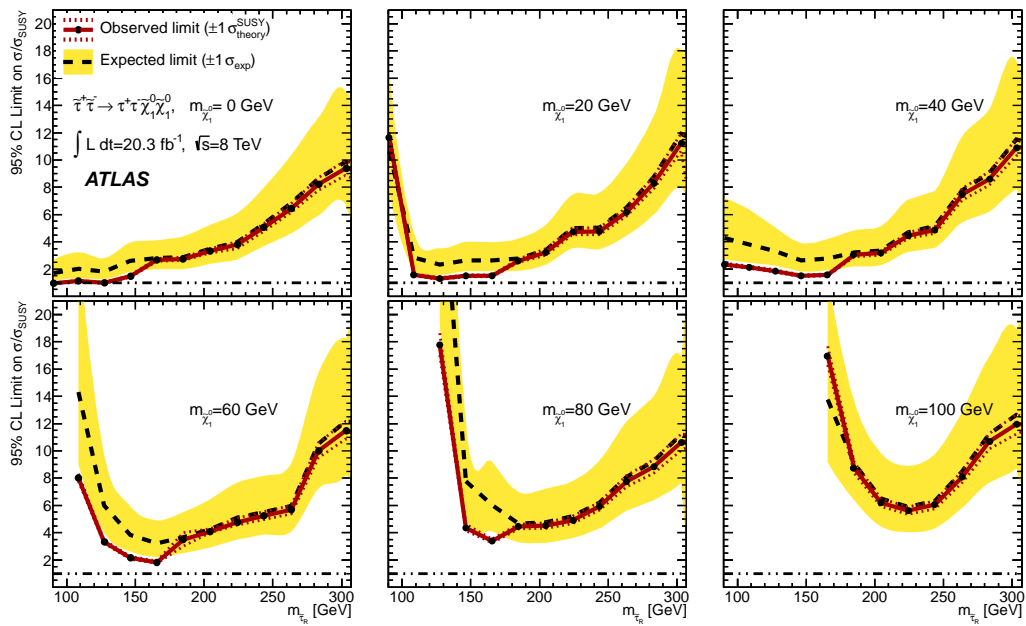


Figure 2.4.: Depicted here are the upper limits on the cross-sections for several mass points used in the analysis in run 1. As can be seen, the observed limit is above 1 for each point, preventing an exclusion of any of them. The best limit is obtained for a massless neutralino and  $m_{\tilde{\tau}} \approx 90$  GeV [5].

## 3. Experimental Setup

### 3.1. Large Hadron Collider (LHC)

The LHC is a hadron accelerator and collider, located at the CERN site near Geneva (see figure 3.1). With its 26.7 km circumference its underground extension reaches also French territory. The tunnel hosting the ring was originally built for the Large Electron-Positron collider ring (LEP). The ring itself consists two beam pipes, leading through eight straight parts and eight arcs. Unlike the Tevatron, the LHC accelerates only protons and collides them at four main interaction points. At each of them a detector is located. The protons are injected via a complex of linear and circular accelerators. Because only protons are accelerated, two separate rings are needed for the LHC. The acceleration and the deflection of the protons is done with superconducting magnetic coils which have to be cooled down to below 2 K with superfluid helium and produce magnetic fields of up to 8 T [16].

The LHC is designed to make collisions at a center-of-mass energy of 14 TeV possible. Currently, it is operating at a center-of-mass energy of 13 TeV in what is called run 2. The number of events showing a certain process is given by

$$N_{\text{event}} = \sigma_{\text{event}} \cdot \int L dt. \quad (3.1)$$

$\sigma_{\text{event}}$  is the cross-section of the process.  $L$  is a collider-specific observable called machine luminosity and is a measure for the statistics of collisions provided by the collider. The luminosity is calculated as

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} \cdot F. \quad (3.2)$$

Here,  $N_b$  is the number of protons per bunch and  $n_b$  the number of bunches per beam,  $f_{\text{rev}}$  is the rotation frequency of the bunches,  $\gamma_r$  is the relativistic Lorentz factor,  $\epsilon_n$  is the normalized beam emittance,  $\beta^*$  is the beta-function and  $F$  a geometrical factor to account for the crossing-angle of the beams at the interaction points. The number of protons per bunch is about  $1.1 \cdot 10^{11}$ , the number of bunches per beam is around 2800 with a spacing of 25 ns between them. The luminosity the LHC is designed for is  $10^{34} \text{ cm}^2\text{s}$ . Commonly, the luminosity is given in  $\text{fb}^{-1}$ . In 2016 and 2015 the LHC has provided approximately  $40 \text{ fb}^{-1}$ , of which  $36.1 \text{ fb}^{-1}$  have been collected by the ATLAS detector (see figure 4.1) [16].

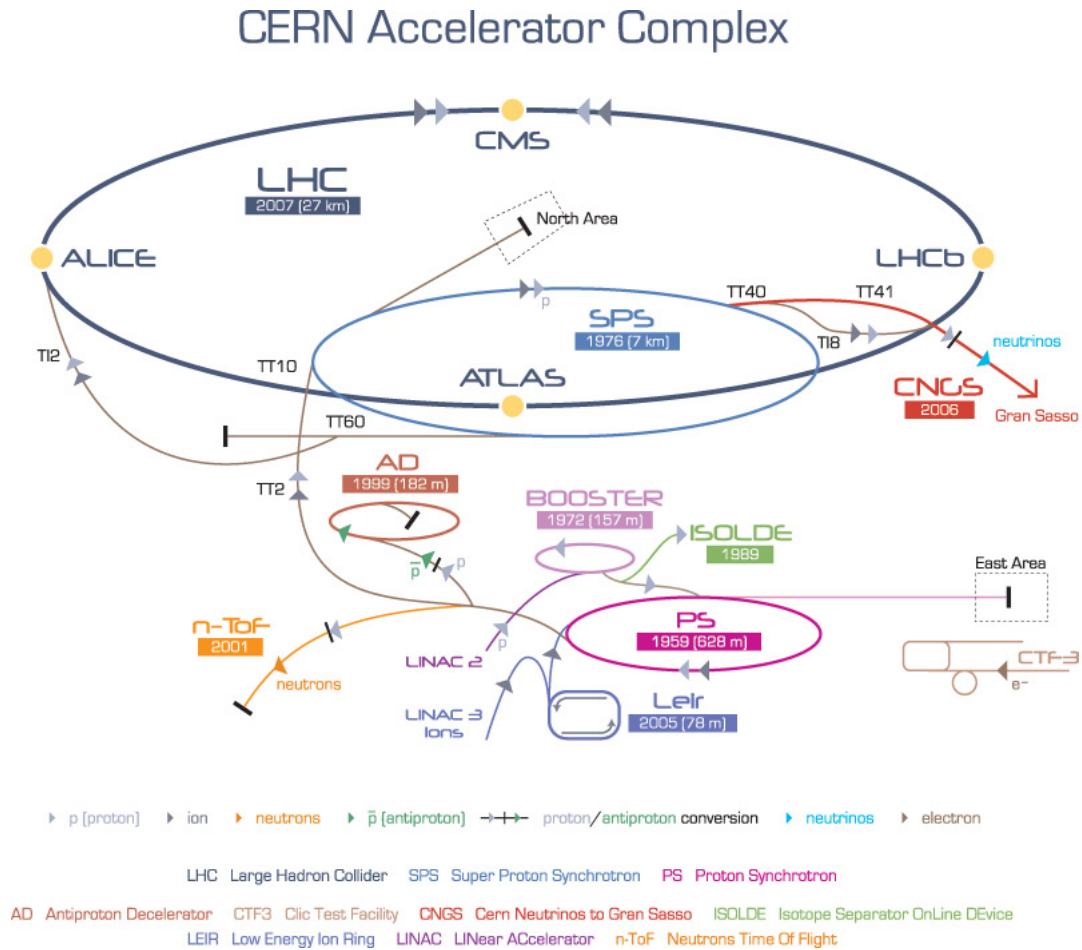


Figure 3.1.: A schematic image of the collider system located at the CERN site near Geneva. The largest accelerator ring is the LHC which is provided pre-accelerated protons by smaller accelerators like the SPS. There are four important locations, where the two proton beams can collide, each of them hosting a particle detector [17].

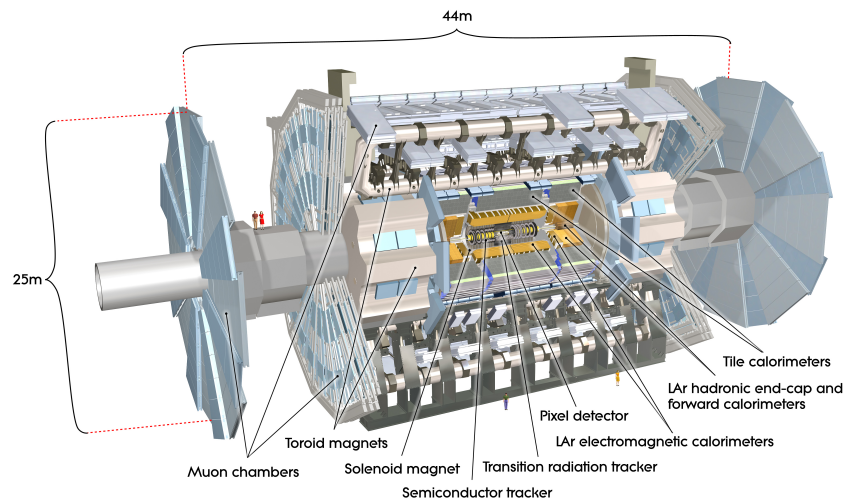


Figure 3.2.: The ATLAS detector is composed of several separate detector systems, which can be roughly divided into the inner detector (tracking of charged particles), the electromagnetic calorimeter (energy measurement of electromagnetically interacting particles), the hadronic calorimeter (energy measurement of strongly interacting particles) and the muon spectrometer (tracking detector for charged minimally ionizing particles) [18].

## 3.2. ATLAS Experiment

The ATLAS detector (A Toroidal LHC Apparatus) is one of two multipurpose detectors at the LHC. In the following, certain conventions and nomenclature used at ATLAS will be explained, as well as its internal structure and its functionality, which is also depicted in figure 3.2.

### 3.2.1. ATLAS Coordinate System

The right-handed coordinate system which is commonly used at ATLAS is defined with the nominal interaction point at its origin and the  $z$ -axis pointing in beam direction. The direction of the  $x$ -axis is towards the LHC ring's center and the  $y$ -axis is defined to point upwards. In the  $x$ - $y$ -plane the azimuthal angle  $\phi$  is measured around the beam line, whereas the polar angle  $\theta$  is determined from the beam line [19].

The longitudinal momentum of the colliding partons can only be determined by means of probability (with the so-called parton density functions or PDFs), making it difficult to fully reconstruct all the (undetected) momenta in  $z$ -direction. Thus, as momentum conservation is practically only applicable in the  $x$ - $y$ -plane, mainly observables in the transverse plane are used, as the transverse momentum  $p_T = \sqrt{p_x^2 + p_y^2}$  and the transverse energy  $E_T$ . The momentum or energy in the transverse plane  $p_T^{\text{miss}}$ ,  $E_T^{\text{miss}}$  respectively, which is carried away by undetected particles is then determined via momentum and energy conservation (see section 5.5). The  $z$ -component of the momentum is used in the Lorentz invariant variable called rapidity  $y$  [19]:

$$y = \frac{1}{2} \cdot \ln \frac{E + p_z}{E - p_z} \quad (3.3)$$

For massless (or nearly massless) particles, the rapidity is equal to the pseudorapidity  $\eta$ :

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

The distance between two particles in the space which is spanned by  $\phi$  and  $\eta$ ,  $\Delta R$ , is calculated as

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}. \quad (3.5)$$

### 3.2.2. Inner Detector (ID)

The inner detector region consists of three different layers of tracking detectors. The detector part which is located closest to the beam tube consists of three layers of silicon pixel detectors. In run 2 an additional layer (called B-layer) has been installed as the innermost layer. These four layers allow for high spatial resolution, which is needed to provide a reliable vertex identification.

Afterwards, particles have to pass the semiconductor tracker (SCT) and the transition radiation tracker (TRT). The SCT consists of eight silicon microstrip layers which are arranged in concentric cylinders around the pixel detector. The TRT is made of straw tubes with a diameter of 4 mm, which are placed parallel to the beam line.

The inner detector is enveloped by a solenoid magnet, generating a magnetic field of 2 T. In this magnetic field the particles' trajectories are bent to allow for momentum and charge measurements. As the number of particles emerging at the collision point is very high (around 1000 particles) and a collision happens every 25 ns the particle tracks occur with a very high density. To handle this, high resolution is needed in the tracking detectors. Taking all three parts of the ID together, they provide sufficient resolution to make high precision measurements of tracks in every direction possible [19].

### 3.2.3. Electromagnetic Calorimeter (ECAL)

The electromagnetic calorimeter is used for precision measurements of the energy deposits of electrons and photons. It is divided into two end-caps and a barrel-part, which itself consists of two identical half-barrels, separated by a gap of 4 mm. The end-caps are made of two components, the inner and the outer wheel, which allows for measurements in a larger pseudorapidity-range. The calorimeter consists of lead absorber plates and kapton electrodes and uses liquid argon as active material [19].

### 3.2.4. Hadronic Calorimeter (HCAL)

The hadronic calorimeter is mainly needed for the measurement of jets. It is located next to the ECAL and can be separated into three main systems. The barrel region and its



extension in  $z$ -direction is covered by the tile calorimeter which is a sampling calorimeter consisting of steel as absorber and scintillating tiles as active material. Like in the ECAL, the hadronic calorimeter part in the end-caps consists of two wheels which are located directly outside the ECAL wheels. They are made of liquid argon as active material and copper plates as absorber. The third part of the HCAL is the so-called forward calorimeter (FCAL) which is embedded in the end-cap wheels of ECAL and HCAL. It is needed to provide an almost complete coverage of the calorimeter and is also crucial for the reduction of radiation before the muon spectrometer. The innermost module of the FCAL is made of copper, the other two consist of tungsten. Again, liquid argon is used as active material [19].

### 3.2.5. Muon Spectrometer (MS)

The muon spectrometer is another tracking detector system, which is meant to measure tracks of particles passing the calorimeter systems. For this purpose another set of three (toroidal) magnets is installed between the MS and the HCAL, each of them consisting of eight coils. In the end-caps the toroid magnets are rotated by  $22.5^\circ$  with respect to the barrel magnets. The bending of a particle track is dependent on the range of pseudorapidity it is located at. In the region where the magnetic fields of the barrel and the end-cap magnets overlap the bending power is decreased.

The muon chambers themselves consist of several separate systems. The Monitored Drift Tubes (MDTs) provide high precision measurements of spatial coordinates of the tracks over the major part of the pseudorapidity-range. For large pseudorapidities this is done by Cathode Strip Chambers (CSCs).

The muon spectrometer is also equipped with a trigger system composed of Resistive Plate Chambers (RPC) in the barrel region and Thin Gap Chambers (TGC) in the end-caps. It is used for the identification of bunch-crossings and the measurement of the muon coordinates in a direction orthogonal to the space measurement in the MDTs and CSCs [19].

### 3.2.6. Trigger System

To reduce the large amount of data that is produced with the collisions in ATLAS, triggers have to be used in order to provide a certain selection of events that are to be stored. The triggers are implemented in a two-layer system. The level 1 triggers (L1) search for high-momentum particles based on a subset of the detector information available to reduce the time needed for the decision. Using data from the calorimeters and the muon system the L1 trigger identifies so-called regions of interest (RoIs) in the detector, given by certain coordinates in the  $\eta$ - $\phi$ -space. This happens within  $2.5 \mu\text{s}$  and reduces the event rate from 40 MHz to 100 kHz [20].

The second trigger step is called high-level trigger (HLT) and in contrast to the L1 trigger system it is purely based on software and either takes the information on the RoI's from L1 as an input or uses the full-event information for offline-like algorithms. The HLT further reduces the event rate to 1 kHz [19, 20].



## 4. Data and Monte Carlo Simulation

### 4.1. Data

The data that is used for this analysis has been recorded by ATLAS in 2015 and 2016 at 13 TeV center-of-mass energy. The luminosity of 2015 and 2016 sums up to approximately  $36.1 \text{ fb}^{-1}$  (figure 4.1).

To neglect events which are not suitable for analyses, only a set of data events is selected in an event cleaning procedure. There it is ensured that events at which the detector was malfunctioning or emitting a large noise burst are omitted. Also, incomplete events are rejected, in cases where parts of the detector were not recording. Furthermore, events are removed if they are recorded twice, which happens very rarely. Finally, only events are recorded for data analysis with at least one primary vertex, reconstructed by two or more tracks. If more than one vertex is identified, the one with the largest sum of its tracks' transverse momenta is chosen.

### 4.2. Monte Carlo Event Generation

The aim of event simulation is not only to provide an estimate for the so-called hard process but also to simulate other features a real event would have, like initial and final state radiation, multiple interactions, hadronization and the detector response and noise. Afterwards, the simulated event can be treated like a real data event and a comparison between data and simulation can be done.

First, the hard process has to be simulated. The cross-section of the process will depend on the corresponding matrix element, which can be derived from the Feynman diagram and the phase space, which is spanned by the degrees of freedom of the process (e.g. decay angles) [22, 23]. A candidate event is defined by the choice of these free parameters. The differential cross-section of the specific candidate event is related to the probability for the event to occur and can be understood as an event weight. When averaging over the event weights of the candidate events, the total measured cross-section is approximated. Typical generators of tree level matrix elements are ALPGEN, MadGraph and POWHEG, where the former two only take into account leading order Feynman diagrams and the latter also next-to-leading order [23].

However, as stated before, this is not sufficient as in reality the initial and final particles can radiate. Additionally, quarks occurring in the process cannot be isolated in the collision and not be detected as such. Instead they form bound, color-neutral states during a process called hadronization [22]. This means, that higher orders have to be taken into account as well as hadronization. This can be done either by calculating higher orders in perturbation theory, or by estimating the dominant effects by (phenomenological) Parton Shower techniques. For this, algorithms like HERWIG++ or PYTHIA8 are used [23].

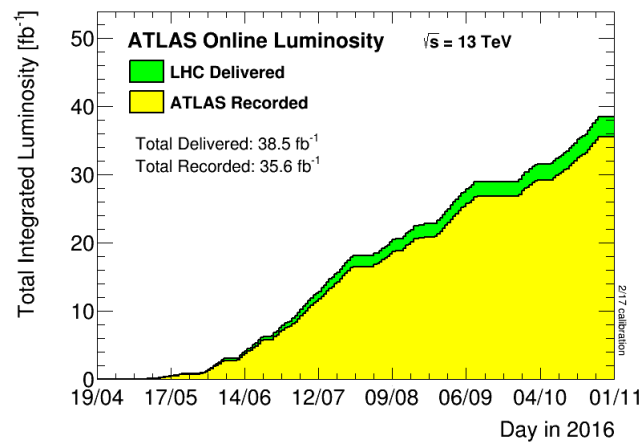


Figure 4.1.: The total luminosity delivered by the LHC and recorded by ATLAS in 2016 at a center-of-mass energy of 13 TeV [21].

There are also generators which provide a combined simulation of both the tree-level matrix element and parton showering, e.g. Sherpa and HERWIG7.

In addition, the hits of the particles in each of the detector components are simulated (with an algorithm like GEANT) and combined with an estimate of the internal detector noise, so that the full detector response of an event can be approximated [24]. The reconstruction and pattern recognition algorithms can then be applied in the same way as for real data.

# 5. Object Reconstruction and Identification

## 5.1. Electrons

Electrons in the central region of the ATLAS detector ( $|\eta| < 2.47$ ) are reconstructed using information of both the inner detector and the electromagnetic calorimeter. The reconstruction process starts with the latter by systematically scanning it for cluster "seeds" with energy deposits of at least 2.5 GeV. The actual clusters are then generated out of the seeds, where duplicates are forbidden. Afterwards, the charged particle track is reconstructed in a two step procedure using the inner detector information. First, a pattern recognition algorithm is performed, modeling the energy loss of the particle in the detector material with that of a charged pion, also considering bremsstrahlung. If the pion hypothesis does not suffice to reconstruct a full track and the track seed is located in a region of interest (determined by the ECAL information), instead an electron hypothesis is used, which allows for higher energy loss. In the second step, the track is fit also using either the pion or the electron hypothesis. Finally, a matching of the fitted track and the cluster in the ECAL is performed. Some selections are applied at this point to the electron candidates. To make sure that they are compatible with the primary interaction vertex, the electrons have to fulfill  $d_0/\sigma_{d_0} < 5$  and  $\Delta z_0 \sin \theta < 0.5$  mm. Here,  $d_0$  is the smallest distance of the track to the beam line (transverse impact parameter),  $\sigma_{d_0}$  is the uncertainty on  $d_0$  and  $\Delta z_0$  is the distance between the nominal interaction point and the point, at which  $d_0$  is measured, projected onto the z-axis (longitudinal impact parameter) [25].

The identification of electrons is done via a likelihood based method, exploiting several properties of the reconstructed objects at once. The method relies on building signal and background probability density functions (PDFs) of those input variables. Based on the PDFs a discriminant  $d_{\mathcal{L}}$  is constructed:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B} \quad (5.1)$$

$$\mathcal{L}_{S(B)} = \prod_{i=1}^n P_{S(B),i}(x_i) \quad (5.2)$$

Here,  $x_i$  are the values of the input parameters and  $P_{S(B),i}$  the PDF for signal (background) for the  $i^{\text{th}}$  variable, evaluated at  $x_i$  [25]. With this, three working points can be defined, depending on the electron selection efficiency of a cut on the likelihood discriminant: Loose ( $\sim 95\%$ ), medium (90%) and tight (80%) [26]. (A "very loose" working point is e.g. used for vetoing electrons in the  $\tau$ -lepton identification.) It has to be noted that, as the shower shapes of the electrons in the calorimeter vary with the pseudorapidity (because a different amount of material has to be passed) and with the energy, the identification working points

are optimized in different  $\eta$ - and  $E_T$ -bins, therefore leading to slightly different efficiencies [25].

For the analysis, the baseline electron selection requires a  $p_T > 10$  GeV and a loose identification. If signal electrons are required, the selection is stricened to a tight identification and the electrons have to pass the GradientLoose isolation, i.e. the sum of the transverse momenta of the tracks and energy deposits in a cone of  $\Delta R < 0.2$  around the electron must not exceed a maximum value, which is dependent on the  $p_T$  of the electron itself.

## 5.2. Muons

Muons are minimally ionizing particles, which therefore do not leave behind large energy deposits in the calorimeter. Thus, the reconstruction is mainly done with information from the inner detector and from the muon spectrometer. The track reconstruction in the inner detector is performed analogously to that of electrons or any other charged particle. In the muon spectrometer a search for hit patterns is done in each muon chamber in order to construct segments. In the MDTs the segments are formed by fitting a straight line to the detected hits which gives the particle's trajectory in the bending plane. The trajectory part in the orthogonal plane is determined in the RPCs and the TGCs, respectively. The building of segments in the CSCs is done by a separate combinatoric algorithm. The hits from the found segments are then fit together resulting in a muon track candidate. Afterwards an overlap removal is applied to prevent that the same segment is used for more than one track candidate. Basing on the separate reconstruction in each detector component, muon candidates are divided into four types. The combined muons (CB) rely on the independent track reconstructions in the inner detector and the muon spectrometer. The tracks are combined by a global refit that is using all the hits from the ID and the MS as inputs. Segment-tagged muons (ST) are built by extrapolating an ID track to the MS, where there must be at least one local segment in the MDTs or the CSCs. One can also attempt to match ID tracks to energy deposits in the calorimeter which are in agreement with those of a minimally ionizing particle. These muons are referred to as calorimeter-tagged muons. Although they have the least purity of all muon types, they can be useful as they compensate for decreased acceptance in regions not covered by the MS due to cabling. The fourth type are the so-called extrapolated muons (ME). There the trajectory reconstruction is mainly based on the track in the MS whereas the matching to an ID track has less priority. In this case, the muons are required to pass at least two layers of the MS chambers (three in the forward region). To ensure that the different muon types do not overlap, a type hierarchy is established, meaning that if a track fulfills the requirements for two of the categories, priority is given first to CB muons, then to ST and CT muons. If overlap with ME muons occur, the reconstruction is selected that provides the better fit quality and larger number of hits [27].

The main background for muon identification stems from decaying pions and kaons. To suppress this background, quality requirements are imposed on the muon candidates. For CB muons the fit quality and the difference of the transverse momenta measured in the ID and the MS are the most important discriminants, as a decay of hadrons in the inner detector would yield a "kink" in the track topology and the measured momenta in the ID and the MS would differ significantly. As measures for these properties, three variables are used in the CB muon identification. The first is the  $q/p$  significance which corresponds

to the ratio of the charge and the momentum measured in the ID and the MS, divided by the uncertainty on these measurements. The second variable is denoted as  $\rho'$ , which corresponds to the absolute value of the difference between the transverse momenta, measured in the ID and in the MS. The third discriminating variable is the normalized  $\chi^2$  of the combined track fitting procedure. Based on cuts on these variables, four muon identification working points are defined, depending on the corresponding selection efficiency: Loose (98.1%), medium (96.1%), tight (91.8%) and high- $p_T$  (80.4%). The reason for the definition of a high- $p_T$  working point is to provide a selection that allows for a good precision of the momentum measurement, at the cost of a smaller selection efficiency. By default, medium muons are used in the analysis, i.e. only CB and ME muons [27].

Furthermore, the baseline selection for muons requires  $p_T > 10$  GeV and  $|\eta| < 2.7$ . The requirements for signal muons are a GradientLoose isolation, like for the electron, and (also similar to the electron)  $d_0/\sigma_{d_0} < 3$  and  $\Delta z_0 \sin \theta < 0.5$  mm. Additionally, cosmic muons (i.e. tracks not originating from the interaction point,  $z_0 > 1$  mm and  $d_0 > 0.2$  mm) are vetoed.

### 5.3. Jets

Quarks cannot be detected individually, as the strong force increases with the distance, so they form bound states during hadronization. (An exception is the top-quark, as it decays even before it can hadronize.) What can therefore be observed, if a quark or a gluon is radiated from a particle, is a bunch of hadrons, called jet. Depending on the boost of the jet, its shape in the detector resembles that of a cone. Several different methods for jet reconstruction have been developed which can be separated roughly into two classes. The first (and maybe more intuitive) approach takes a cone with a fixed radius  $R$  as a starting point, calculating the sum of the constituents' momenta and taking the resulting direction as the center of the next cone in the iteration, until a stable cone is found. However, these methods are generally found to be infrared and/or collinear unsafe, which means that they are not robust against adding a soft parton and/or splitting the momentum of one parton into two collinear partons. The second class of algorithms reconstructs jets by iteratively "recombining" jet constituents until the remaining objects are too far away from each other, thus using the distance between the objects as a defining parameter. The definition of the distance is generally done in the plane spanned by the azimuthal angle  $\phi$  and the rapidity  $y$ :

$$d_{ij}^2 = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \left( \frac{(\Delta y_{ij}^2 + \Delta \phi_{ij}^2)^2}{R^2} \right) \quad (5.3)$$

Here, the transverse momentum is denoted as  $k_t$ . The radius parameter  $R$  has to be given as an input to the algorithm. The parameter  $p$  can be chosen freely, each choice resulting in a different concrete method. ( $p = 0$  is used for the Cambridge-Aachen algorithm and  $p = 1$  for the  $k_t$  algorithm.) In this analysis, jets are reconstructed using the anti- $k_t$  algorithm, for which  $p = -1$ . In this particular case,  $R$  is chosen to be 0.4. The advantage of the anti- $k_t$  algorithm is that it is infrared and collinear safe while it is also very fast. In contrast to the Cambridge-Aachen and  $k_t$  algorithms, the anti- $k_t$  algorithm also provides jets with circular boundaries, as soft particles are recombined with hard ones before being recombined with other soft particles [28]. As inputs, the anti- $k_t$  algorithm uses

the four-momenta of the particles which are measured in the detector by using Topological Clusters (TopoClusters). This means that several calorimeter cells are combined together in a three-dimensional cluster according to the energy deposits in the cells. In addition, the measurement of the energy deposit in the cells is calibrated for each cell (local cell reweighting), to account for the difference in the detector performance for electromagnetic and hadronic particles. Furthermore, the jet-energy-scale calibration (JES) is applied, which includes the correction of the four-momentum of the jet to make it point to the primary vertex, pile-up correction and an absolute correction of the detector response based on Monte Carlo simulations [29].

In the baseline selection, jets are accepted if they have  $p_T > 20 \text{ GeV}$  and are within  $|\eta| < 2.8$ . Also, a b-tagging algorithm is applied to identify jets originating from a b-quark, which is operating at an efficiency of 77%.

## 5.4. Tau Leptons

$\tau$ -leptons have a mass of  $1.777 \text{ GeV}$  and a proper decay length of  $87 \mu\text{m}$ , before they decay either into leptons or into hadrons. Therefore, they cannot be identified directly but via their decay products in the detector. Leptonic decays are not reconstructed as  $\tau$ -leptons but are handled as light signal leptons. Hadronic decays (which make up for 65% of the  $\tau$ -decays) can be further distinguished into cases where one charged hadron is produced (72% of the hadronic decays) and cases, in which three charged hadrons are produced (22% of the hadronic decays). The first are referred to as 1-prong and the latter as 3-prong  $\tau$ -leptons.  $\tau$ -leptons are, compared to other particles, very difficult to reconstruct, as they can be imitated by jets as well as by electrons in the detector [30].

To reconstruct  $\tau$ -leptons, jets with either one or three tracks are taken into account which were reconstructed by the anti- $k_t$  algorithm with a distance parameter of  $R = 0.4$  and which have to be within  $|\eta| < 4.5$ . In addition,  $\tau$ -candidates are demanded to have  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.5$  and should not be located between the barrel and the forward calorimeters ( $1.37 < |\eta| < 1.52$ ). The  $\tau$ -vertex is assumed to be the vertex of the candidate tracks which have the largest momentum fraction within the jet. There are special calibration methods for obtaining the  $\tau$ -candidate's energy, one called baseline calibration, the other using a boosted regression tree (see [30], [31]).

The next step is the  $\tau$ -identification, which is an algorithm designed to distinguish real  $\tau$ -leptons from jets. The identification is based on boosted decision trees, one designed each for 1-prong and 3-prong  $\tau$ -candidates. Using the BDT response, three working points are defined, depending on the efficiency of the  $\tau$ -selection - loose at 60% (50%), medium at 55% (40%) and tight at 45% (30%) for 1-prong (3-prong)  $\tau$ -leptons. This holds for the offline reconstruction of  $\tau$ -leptons. In fact, an analogous scheme is performed online for the trigger selection. There, also a BDT is trained for discriminating  $\tau$ -leptons from background, but with the 1-prong  $\tau$ -leptons being selected with 95% of the offline identification's efficiency and the 3-prong  $\tau$ -leptons with 70% of the offline efficiency [30].

As it has already been stated, not only jets form a background for  $\tau$ -identification but also electrons. For this reason, 1-prong  $\tau$ -candidates are rejected offline if they are within a distance  $\Delta R < 0.4$  of a reconstructed electron, that passes the very loose electron working point. However, there is no discrimination of  $\tau$ -leptons and electrons at trigger level [30].



For this analysis,  $\tau$ -leptons (having either one or three tracks) are used which pass the baseline criteria, i.e.  $p_T > 20$  GeV and  $|\eta| < 2.5$ , with their charge being either +1 or -1 (in units of the elementary electric charge).

## 5.5. Missing Transverse Energy (MET)

As the exact momentum fraction of the colliding partons cannot be determined, it is not possible to use energy and momentum conservation in the longitudinal direction of the detector to measure particles escaping detection. However, this can be done in the transverse detector plane, assuming that the momenta of the initial partons are negligibly small in this direction.

The missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ ) is calculated as the negative vector sum of the transverse momenta of all reconstructed physics objects in an event:

$$\vec{p}_T^{\text{miss}} = - \sum_{i \in \Lambda}^n \vec{p}_{T,i}, \quad (5.4)$$

for  $n$  reconstructed objects. Its magnitude is denoted as the missing transverse energy  $E_T^{\text{miss}}$ . The reconstructed electrons, photons, muons,  $\tau$ -leptons and jets make up the so-called hard terms of the missing transverse energy reconstruction. Additionally, soft terms have to be considered for all the reconstructed momenta that cannot be associated to a hard physics object. The soft terms can be determined in several ways, e.g. using the energy deposits in the calorimeter or the tracks in the inner detector. The latter is used for ATLAS in run 2 by a reconstruction algorithm called Track Soft Term (TST). The advantage of the TST algorithm is its robustness against varying pile-up conditions. Furthermore, the jet-vertex-tagger (JVT) technique associates jet tracks to vertices, enabling the removal of pile-up jets. This also improves the resolution of the missing transverse energy reconstruction [29].

## 5.6. Overlap Removal

The reconstruction methods for different particles share the same input tracks or energy deposits, consequently interpreting the same data in different ways. It can happen that multiple objects are reconstructed from the exactly same detector signature. This kind of double-counting is accounted for in a procedure called overlap-removal. Normally, an object is removed if the distance to another object in  $\eta$ - $\phi$ -space is too small, usually  $\Delta R < 0.2$  or  $\Delta R < 0.4$ , depending on the objects that are compared. For example,  $\tau$ -leptons passing the loose selection are removed if they lie within  $\Delta R < 0.2$  of a reconstructed baseline electron or muon. Contrary, a reconstructed  $\tau$ -lepton is kept if it matches within  $\Delta R < 0.2$  with a jet, while the jet is rejected. Jets are not only rejected if compared to  $\tau$ -leptons but also to muons if they lie in a cone with  $\Delta R < 0.4$ . For the electron and jet comparison, the overlap removal is a little more sophisticated. If the distance in  $R$  between a reconstructed electron and a jet is smaller than 0.2 the electron is preferred. If it is larger than 0.2 but still smaller than 0.4, the jet is kept and the electron is removed.

Additionally, an electron is rejected if it shares the track in the inner detector with a muon.

## 5.7. Variable Definitions

### Number and Charge of Particles

The number of a particular type of particles is essential for the restriction of the events to a specific signature. Therefore, the number of reconstructed  $\tau$ -leptons  $N(\tau)$  is usually already used in the preselection. If not stated differently, the medium identification criterium is used. The leptonic decays of  $\tau$ -leptons are covered by setting requirements on the number of light leptons. For this, the number of light leptons fulfilling "signal" requirements  $N(\text{lep}_{\text{sig}})$  is used. The electric charge of  $\tau$ - and light leptons is not used directly as a variable, but in events with at least two leptons the information whether the two leading leptons have same (SS) or oppositely signed electric charges (OS) is exploited. The number of b-tagged jets  $N(\text{b-jets})$  with transverse momenta above 20 GeV is particularly useful for reducing the background processes originating from processes with top quarks.

### Transverse Momenta

The transverse momenta of the leading and next-to-leading hadronic  $\tau$ -leptons are denoted as  $p_T(\tau_1)$  and  $p_T(\tau_2)$ , respectively. For light signal leptons the notation is  $p_T(\text{lep}_{\text{sig}})$  or  $p_T(e/\mu)$ . The transverse momenta of signal particles are not only used for discriminating signal and background themselves, but are also needed for the definition of other more sophisticated variables.

### Effective Mass

The effective mass  $m_{\text{eff}}$  is calculated for each event according to the formula

$$m_{\text{eff}} = \sum_{\tau} |\vec{p}_T| + E_T^{\text{miss}}. \quad (5.5)$$

Here, the sum of the  $\tau$ -leptons' transverse momenta (or also that of the light signal lepton and the hadronic  $\tau$ -lepton in the case of the LepHad-channel) is added to the missing transverse energy.

### Azimuthal Angle and Pseudorapidity

The azimuthal angle of particles is typically equally distributed over the whole  $\phi$ -range. Nevertheless, the difference in the azimuthal angle of two particles, e.g.  $\Delta\phi(\tau_1, \tau_2)$  can differ between physical processes, making it a suitable discriminating variable. Additionally, the azimuthal angle between leptons and the missing transverse energy, e.g.  $\Delta\phi(\tau_1, E_T^{\text{miss}})$  is used as a variable. Here it is defined, that always the smaller one of the possible two angles is chosen. Similarly, the difference in pseudorapidity between two particles, e.g.

$\Delta\eta(\tau_1, \tau_2)$ , can be used. Both of the mentioned variables are used in combination to calculate the distance in the  $\eta$ - $\phi$ -space between the two leptons in consideration:

$$\Delta R = \sqrt{\Delta\eta^2(\tau_1, \tau_2) + \Delta\phi^2(\tau_1, \tau_2)} \quad (5.6)$$

What distinguishes supersymmetric particles from Standard Model particles are not only their masses but also the spin. Because of this a variable that is sensitive to the spin of the decaying particles would be an interesting addition to the "conventional" set of variables. The angular distribution of such particles has a different form depending on their spin. In the case of sleptons that are pair produced in a Drell-Yan process, it is proportional to  $(1 - \cos^2 \theta^*)$ , with  $\theta^*$  being the angle between one of the incoming protons and one of the produced sleptons. However, if the sleptons decay dominantly into Standard Model leptons and some undetectable particles, this production angle of the sleptons cannot be reconstructed anymore. For this, one would need to perform a boost into the center-of-mass frame of the sleptons, which would require full information in the  $z$ -direction of the detector. A solution for this is to pick a Lorentz invariant quantity instead, like the difference in pseudorapidity of the two Standard Model leptons. As the leptons provide some information on the pseudorapidity of their predecesing sleptons, also this quantity is related to the slepton production angle. To make a better comparison with the production distribution possible, the pseudorapidity difference is converted into an angular variable  $\cos \theta_{ll}$  [32]:

$$\cos \theta_{ll} \equiv \cos \left( 2 \tan^{-1} \exp(\Delta\eta(\tau_1, \tau_2)/2) \right) = \tanh \left( \frac{1}{2} \Delta\eta(\tau_1, \tau_2) \right) \quad (5.7)$$

Here,  $\cos \theta_{ll}$  inherits the Lorentz invariance of  $\Delta\eta_{\tau_1, \tau_2}$  and has the geometrical interpretation as the cosine of the polar angle between each lepton and the beam axis in the frame in which the leptons have equal and opposite pseudorapidities. The definition in the case of one hadronic  $\tau$ -lepton and one light signal lepton is analogous to equation 5.7.

### Invariant Mass

The Lorentz invariant mass  $m_{\text{inv}}$  of a particle is given by the relativistic energy-momentum relation:

$$p^2 = E^2 - \vec{p}^2 = m_{\text{inv}}^2 \quad (5.8)$$

with  $p$  the four-momentum of the considered particle. Suppose, a particle  $C$  is decaying into two particles  $A$  and  $B$ . Then the invariant mass of particle  $C$  can be calculated as:

$$m_C^2 = (p_A + p_B)^2 \quad (5.9)$$

If the particles  $A$  and  $B$  are highly relativistic, i.e.  $E \gg m$ , the squared terms in  $p_A$  and  $p_B$  can be neglected, as they are equal to the invariant masses of  $A$  and  $B$ , respectively. Then one can write equation 5.9 as

$$m_C^2 = 2p_A p_B = 2 \cdot [E_A E_B - p_{xA} p_{xB} - p_{yA} p_{yB} - p_{zA} p_{zB}]. \quad (5.10)$$

In the coordinate system of ATLAS the four-momentum of a relativistic particle is given as

$$p = |p_T| \cdot \begin{pmatrix} \cosh \eta \\ \cos \phi \\ \sin \phi \\ \sinh \eta \end{pmatrix}. \quad (5.11)$$

Inserting this into equation 5.10 a formula for the invariant mass of two particles  $A$  and  $B$  is obtained:

$$m_C^2 = 2 \cdot |p_{TA}| |p_{TB}| \cdot [\cosh(\eta_A - \eta_B) - \cos(\phi_A - \phi_B)] \quad (5.12)$$

In the case of direct stau production, the only particles for which an invariant mass can be calculated are the two  $\tau$ -leptons (or the  $\tau$ -lepton and the light lepton in the LepHad-channel). The invariant mass of the two  $\tau$ -leptons is not equal to the stau mass, but it can e.g. be used for eliminating events in which the  $\tau$ -leptons originate from a  $Z$ -boson.

### Transverse Mass

When searching for new particles, one usually aims to estimate its mass first. This is no problem if the considered particle has fully observable decay products, because the invariant mass of the measured particles can then be calculated. But if one of the decay products cannot be detected, like neutrinos or neutralinos, the mass of the decaying particle must be estimated by other means. One possibility is the so-called transverse mass, which is given here for the example of a  $W$ -boson, decaying into a Standard Model  $\tau$ -lepton and a neutrino, which carries away missing transverse energy [33]:

$$m_T^2(\tau) = 2 \cdot E_T^{\text{miss}} \cdot p_T(\tau) \cdot (1 - \cos(\Delta\phi(\tau, E_T^{\text{miss}}))) \quad (5.13)$$

The transverse mass is not equal to the invariant mass and therefore is not identical to the  $W$ -boson mass (unless the  $\tau$ -lepton and the neutrino are produced at the same pseudorapidity), but has the following property [33]:

$$m_T^2(\tau) \leq m_W^2 \quad (5.14)$$

Additionally, in events with at least two hadronic  $\tau$ -leptons or one  $\tau$ - and one light lepton, the sum of the transverse masses of the two leading leptons can be used and is denoted as  $m_T(\tau_1) + m_T(\tau_2)$ .

Note here, that the missing transverse energy cannot be separated in the case of more than one particle contributing to it. This means that 5.14 does not necessarily hold for  $m_{\tilde{\tau}}$  in stau pair production. Nevertheless, it can be useful for eliminating e.g. background events with one  $W$ -boson decaying into a  $\tau$ -lepton and a neutrino.

### Stransverse Mass

In direct stau production (as well as in many other SUSY searches) two neutralinos are produced which both contribute to the missing transverse energy, so they cannot be distinguished. Therefore, neither the invariant nor the transverse mass is a suitable measure for the stau mass. However, another variable can be constructed analogously to the transverse mass. Suppose the missing transverse momentum could be split up into the separate contributions of the neutralinos  $a$  and  $b$ . Then the lower limit for the stau mass would be the largest of the two transverse masses [33]:

$$\vec{p}_T^{\text{miss}} = \vec{p}_T(\tilde{\chi}_{1,a}^0) + \vec{p}_T(\tilde{\chi}_{1,b}^0) \quad (5.15)$$

$$m_{\tilde{\tau}}^2 \geq \max\{m_T^2(\tau_a, \tilde{\chi}_{1,a}^0), m_T^2(\tau_b, \tilde{\chi}_{1,b}^0)\} \quad (5.16)$$

Because the exact splitting of the missing transverse momentum is not known, it is approximated by minimizing over each possible splitting [33]:

$$m_{\tilde{\tau}}^2 \geq m_{T2}^2 \equiv \min_{\vec{p}_{T,a} + \vec{p}_{T,b} = \vec{p}_T^{\text{miss}}} \left[ \max\{m_T^2(\tau_a, \tilde{\chi}_{1,a}^0), m_T^2(\tau_b, \tilde{\chi}_{1,b}^0)\} \right], \quad (5.17)$$

giving a lower limit on the stau mass and therefore usually providing good discrimination power to Standard Model processes. It has to be noted that the mass of the neutralinos is neglected as well as the mass of the  $\tau$ -lepton.



## 6. Samples

### 6.1. Signal Samples

The direct stau production process is referred to as "signal" in these studies. It is simulated via the Monte Carlo method for different masses of stau and neutralino, also called mass points. The simulation of the tree level process is done with MadGraph, the parton shower is added by PYTHIA8. The signal grid consists of 44 mass points, ranging from 80 GeV to 260 GeV in stau mass with neutralino masses up to 140 GeV. The mass points are also shown in figure 6.1. For the LepHad-channel only mass points up to a stau mass of 160 GeV have been available for these studies.

The cross-sections, corrected for the filter efficiency, are listed in tables A.1 and A.2. For each of the mass points two subsets are generated - one for  $\tilde{\tau}_1$  and one for  $\tilde{\tau}_2$ , which have different cross-sections, but are merged together into one sample, as they are mass-degenerate. The samples are weighted to an integrated luminosity of  $36.1 \text{ fb}^{-1}$ .

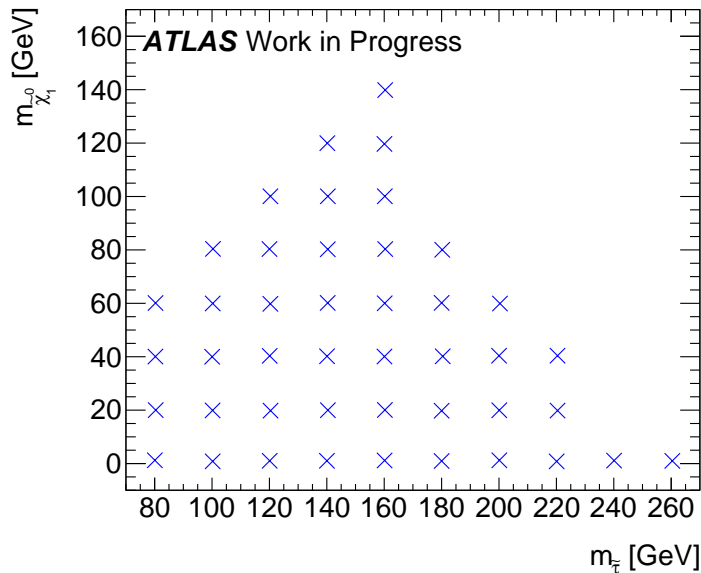


Figure 6.1.: Available signal mass points for the direct stau HadHad-channel in the plane of stau and neutralino masses. For the LepHad-channel only stau masses of up to 160 GeV have been available for these studies.

## 6.2. Background Samples

As "background" one understands in this case (Standard Model) processes, which show the same or a very similar detector signature, either because the processes share the same final state, or because the signal final state is imitated by the background process due to a falsely reconstructed particle. In the following, the background processes considered for these studies are listed.

### 6.2.1. $W$ +jets

" $W$ +jets" is an umbrella term for all events in which a  $W$ -boson is produced in association with one or more jets. Here, only leptonic decay modes of the  $W$ -boson are considered. When investigating the HadHad-channel, the  $W$ -boson decays into a (hadronically decaying)  $\tau$ -lepton, while the required second  $\tau$ -lepton is imitated (or "faked") by a jet. For the LepHad-channel, the  $W$ -boson can either decay directly into a light lepton (electron or muon) or into a leptonically decaying  $\tau$ -lepton, while the hadronic  $\tau$ -lepton is again faked by a jet. This background is estimated via Monte Carlo simulation, using Sherpa 2.2.1 as generator. A list of the samples used for these studies can be found in tables A.3 and A.4.

As the selection of two medium  $\tau$ -leptons (see section 7.1) in this sample implies that at least one of them must be faked, the number of events is related to the efficiency with which a jet is mistaken as a medium  $\tau$ -lepton (fake efficiency). As this efficiency (especially for medium or tight  $\tau$ -leptons) is relatively small (in the order of 1% to 10%, depending on the transverse  $\tau$ -momentum and  $E_T^{\text{miss}}$ ), this results in very few selected  $W$ +jets events, compared to the amount initially provided. Consequently, the statistical uncertainty on any number of selected events (in particular after a rather tight signal region definition) suffers from the small (unweighted) number of selected events. This will turn out to be one of the major issues for obtaining sensitivity for direct stau production. One approach to solve this problem is the container tau promotion method, which is described in chapter 9.

### 6.2.2. $Z$ +jets

Analogously to  $W$ +jets,  $Z$ +jets are processes in which a  $Z$ -boson is produced together with a certain number of jets. The main contribution to the HadHad-channel are  $Z$ -bosons decaying into a pair of hadronic  $\tau$ -leptons. For the LepHad-channel, only one of the  $\tau$ -leptons decays hadronically, the other leptonically. Contrary to  $W$ +jets, the main contribution from  $Z$ +jets is provided by real  $\tau$ -leptons. The  $Z$ +jets background is estimated with Monte Carlo simulation generated by Sherpa 2.2.1., like  $W$ +jets. The list of  $Z$ +jets samples is given in tables A.5 and A.6.

### 6.2.3. Dibosonic Processes

Events, in which two electroweak bosons are produced are called dibosonic events. E.g. for  $ZZ \rightarrow \tau\tau\nu\nu$  and  $WW \rightarrow \tau\nu\tau\nu$ , those processes lead to the same final signature as for direct stau production. Therefore the dibosonic background is a very important



one. Even though the cross-sections for these processes are typically not very large, the discrimination to the signal is rather difficult. The simulation of these processes is also done with the Sherpa 2.2.1 generator. The specific samples are listed in table A.7.

#### 6.2.4. Top Quarks

Another source of background are processes containing top quarks, i.e. top quark pair production ( $t\bar{t}$ ) and single top production. For reasons of completeness, also the top pair production with an associated boson radiation ( $t\bar{t}V$ ) is taken into account, although these processes make up a very small fraction of the total background. However, the associated production of a top quark pair and a Higgs boson is neglected in these studies. Top quarks almost always decay into a bottom quark and a  $W$ -boson, which can further decay into  $\tau$ -leptons, light leptons or quarks. However, these processes can be fairly reduced by requiring only events without b-jets. The simulation is done by the POWHEG event generator, with the Parton Shower added by PYTHIA. ( $t\bar{t}V$  events are generated with MadGraph and a PYTHIA Parton Shower.) The list of samples is provided in tables A.8, A.9 and A.10.

#### 6.2.5. QCD Multijet Processes

Because processes only containing jets in the final state cannot be calculated perturbatively, an accurate simulation is difficult to provide. Instead, these processes are estimated with the help of data from the ATLAS detector in a reweighting technique, commonly called ABCD-method. The idea is to divide the phase space spanned by two (uncorrelated) variables into four regions, one of them containing the signal region (SR) (i.e. a region of phase space, where a large amount of signal events is expected). The other three regions, called control regions (CR) are assumed to be signal free. Therefore, if subtracting the Monte Carlo estimations of all of the above listed backgrounds from data in the CRs, what is left must be the QCD multijet background. From the differences of data and Monte Carlo in the control regions A and B a transfer factor is calculated. Assuming that the region B relates to region A like region D to C, the same transfer factor is then applied to the data to Monte Carlo difference in region C to get an estimate for region D, which is (or contains) the signal region. To validate this procedure and also get an approximation of the systematic uncertainty, two more regions orthogonal to the other regions, E and F (also called validation regions or VRs) are introduced. The transfer factor which is calculated from A and B is then applied to region E to get an estimate of the QCD contribution in region F, analogously to the signal region. This estimate is then compared to the real data to Monte Carlo difference in F. From this, the reliability of the QCD estimation procedure can be shown.

In this analysis the phase space used for the ABCD-estimation is spanned on the y-axis by the  $m_{T2}$ -variable (see section 5.7) and on the x-axis by a combination of the  $\tau$ -identification quality (loose, medium, tight) and the sign of the  $\tau$ -charges in the event (OS for oppositely signed  $\tau$ -leptons and SS for same signed  $\tau$ -charges). The control regions A and B are located below a value of  $m_{T2}$  of 10 GeV, control region C and signal region D above 20 GeV. Inbetween ( $10 \text{ GeV} < m_{T2} < 20 \text{ GeV}$ ) there lie the validation regions E and F. For the regions A, C and E, the  $\tau$ -leptons in the event must either have the same charge sign (SS) or, if they are oppositely signed (OS), they must not fulfill the medium

$\tau$ -ID, but only the loose selection. For the regions B, D and F, the  $\tau$ -leptons must have opposite sign and at least medium  $\tau$ -quality. With these requirements an overlap between the regions is prevented. A schematic image of this setup of the ABCD-method is shown in figure 6.2. The results of the estimation are given in table 6.1. For all regions only the online trigger selection described in section 7.1 is applied and at least two loose  $\tau$ -leptons are required.

A measure for the correlation between the two axes of the defined phase space is given in figure 6.3 by plotting the (normalized)  $m_{T2}$ -distribution for the data to Monte Carlo difference for regions A and C together, and regions B and D respectively, up to a  $m_{T2}$ -value of 100 GeV. One can see that at least for small values of  $m_{T2}$  there is almost no correlation, i.e. the normalized distributions match. For higher values of  $m_{T2}$  this is not visible due to large statistical fluctuations.

The validation of the QCD estimation is depicted in figure 6.4, where the absolute value of the azimuthal angle between the leading  $\tau$ -lepton and the missing transverse energy is plotted for the real data to Monte Carlo difference in region F and the estimated QCD contribution in F. The same is done for the missing transverse energy itself, as well as for the leading  $\tau$ -lepton's transverse momentum. Although there are large statistical uncertainties in some of the bins, the overall distributions match rather well, with the ratio of both fluctuating around 1.

Figure 6.5 shows the distributions of the Monte Carlo modeled background processes for the  $m_{T2}$ -variable, the missing transverse energy and the two leading transverse  $\tau$ -momenta, compared to the recorded ATLAS data in regions A, E and C. The discrepancy between data and the Monte Carlo simulations is assumed to entirely correspond to the QCD multijet contribution in these regions.

Furthermore, in figure 6.6, the QCD estimation for region D, together with the Monte Carlo simulations of the remaining Standard Model backgrounds are compared with data up to a certain threshold, by plotting some kinematic variables. It can be observed that the simulated Standard Model backgrounds together with the semi-data-driven QCD estimate describe the behavior of the data distribution sufficiently well for these studies in the unblinded bins, apart from statistical fluctuations. For a more accurate QCD estimate other methods like the fake-factor method (as described in [34]) or jet smearing [35] could be used.

Finally, it has to be noted that this estimation is only designed for the case of two hadronically decaying  $\tau$ -leptons, which means that it is not applicable in this form for the LepHad-channel, as two hadronic  $\tau$ -leptons are required.

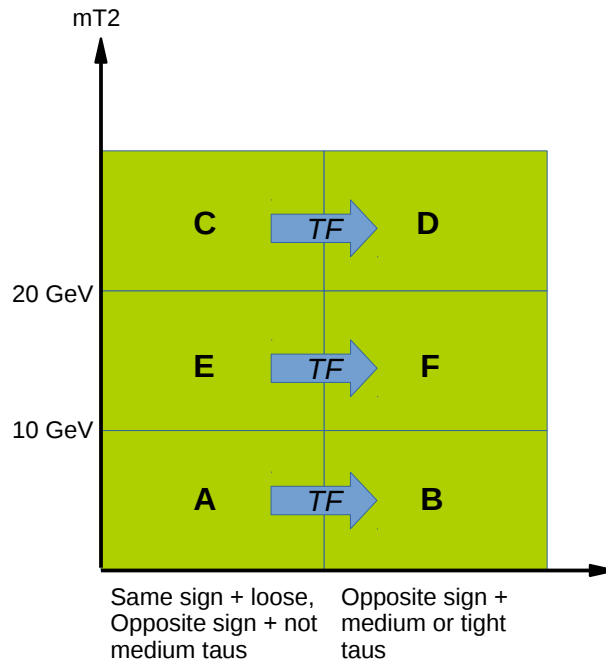


Figure 6.2.: This figure shows the principle of the ABCD estimation technique. A, B and C are the control regions, E and F the validation regions and D is called signal region (although in this case, D is not a final signal region but only containing it). The transfer factor (TF) is calculated from the difference between data and Monte Carlo simulation in the regions A and B. Applying this transfer factor to the same difference in region C (E), one obtains an estimate for the QCD contribution in region D (F).

Region	Data	Monte Carlo	Data – Monte Carlo	QCD estimate ( $\sigma_{\text{stat}}$ )
A	37121	2270.6	34850.4	–
B	21642	10752.7	10889.3	–
C	42514	3322.01	39192.0	–
E	13622	621.308	13000.7	–
F	6491	2025.18	4465.82	<b>4062.18</b> (1.05%)
D	–	–	–	<b>12245.90</b> (0.80%)

Table 6.1.: Listed here are the yields of data and Monte Carlo in the control and validation regions, as well as the resulting expected QCD multijet contribution for validation region F and signal region D (and the statistical uncertainty  $\sigma_{\text{stat}}$ ) for an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The transfer factor calculated from regions A and B is 0.31. The ratio of the data to Monte Carlo difference in region F to the corresponding QCD estimate is 1.10, implying a discrepancy of about 10%.

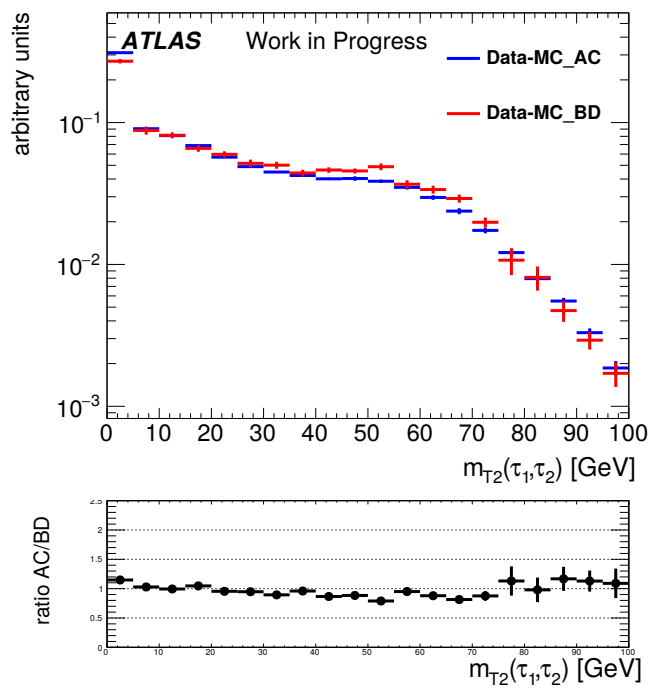


Figure 6.3.: Illustrated here is the correlation between  $m_{T2}$  and the  $\tau$ -ID and charge sign requirements. In blue the  $m_{T2}$  distribution of the data to Monte Carlo difference is plotted for regions A, E and C and in red for regions B, F and D. Looking at the ratio plot below, the two distributions match almost perfectly up to a value of  $m_{T2} = 40$  GeV, meaning that there is only negligible correlation. At higher values of  $m_{T2}$  the fluctuations increase but the ratio is mostly still compatible with 1.

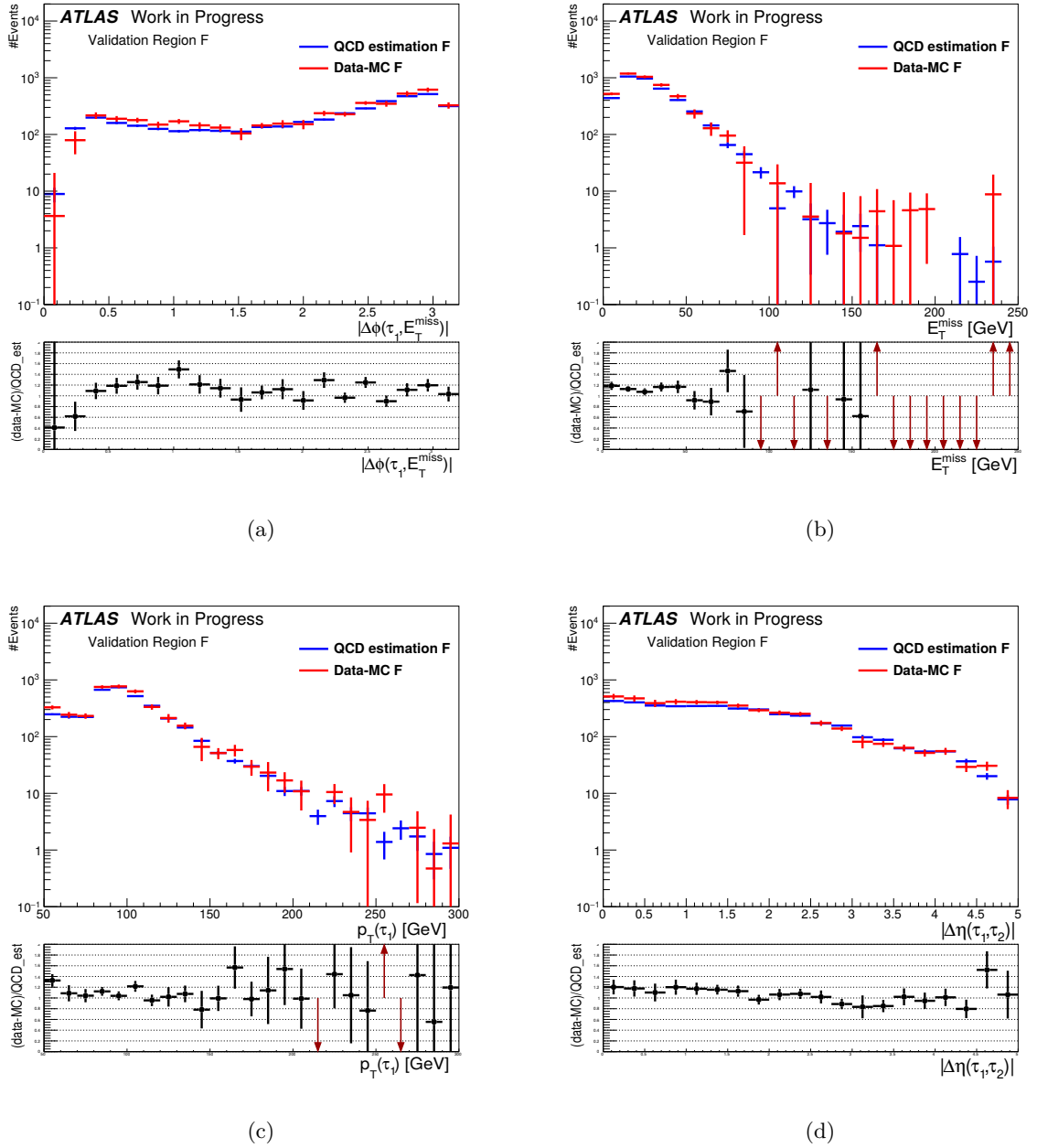


Figure 6.4.: Shown in this figure are comparative plots of the data to Monte Carlo difference and the QCD estimation in region F for the absolute value of the azimuthal angle between the leading  $\tau$ -lepton and the missing transverse energy (a), the missing transverse energy itself (b), the transverse momentum of the leading  $\tau$ -lepton (c) and the pseudorapidity difference between the two  $\tau$ -leptons (d). The plots below each histogram depict the ratio of the QCD estimation to the data to Monte Carlo difference for each bin. It can be observed that this ratio at some point varies a lot but is mostly consistent with 1, with a slight tendency for underestimation.

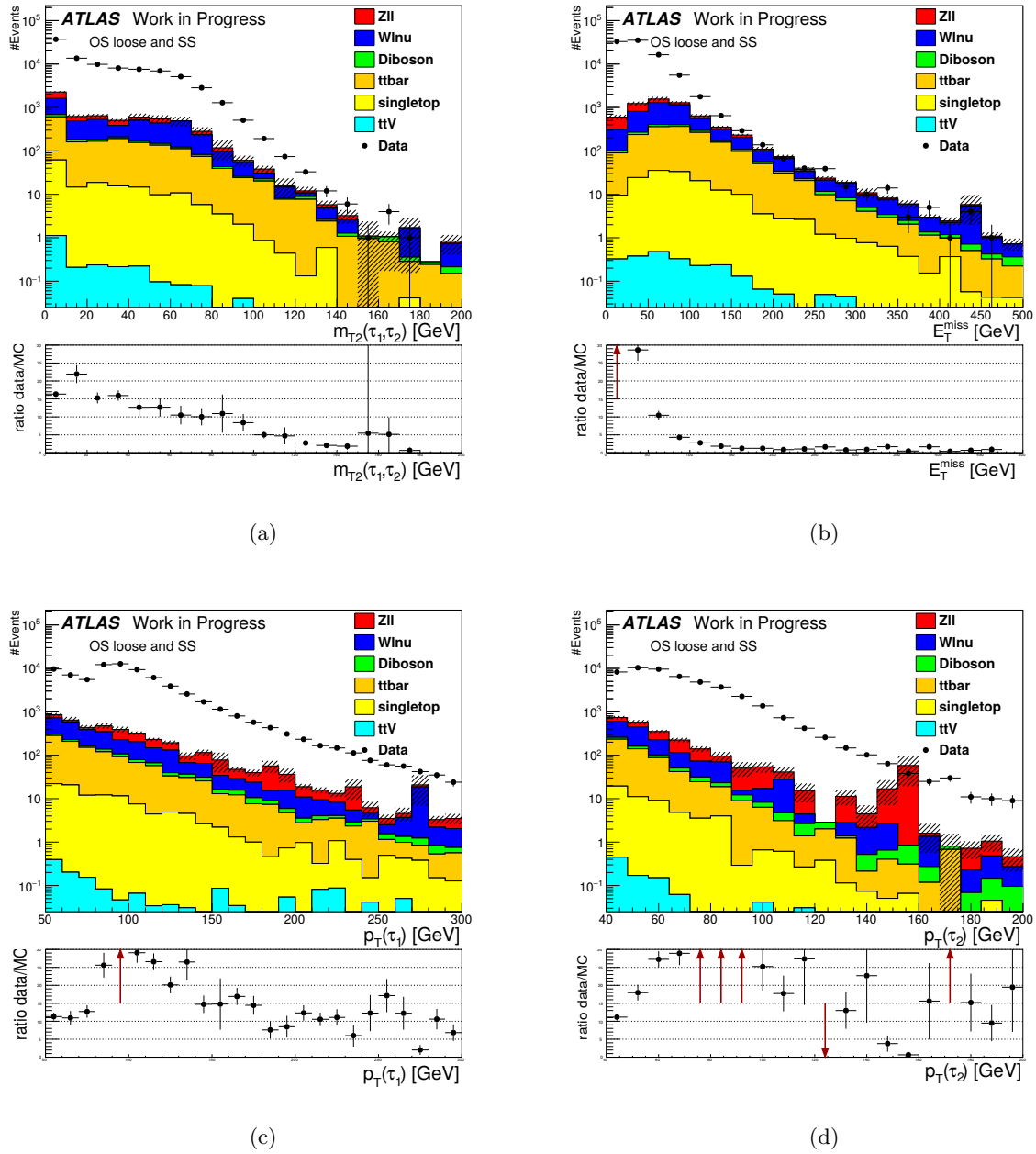


Figure 6.5.: These four histograms show the distributions of  $m_{T2}$  (a), the missing transverse energy (b) and the transverse momenta of the leading (c) and next-to-leading  $\tau$ -lepton (d) in the regions A, E and C. The colored shapes correspond to the Monte Carlo simulations of the Standard Model backgrounds, whereas the black points show the yields of the recorded ATLAS data. The plots below the histograms show the ratio of the simulated to the recorded data. For the ABCD-method it is assumed that the discrepancy of these two corresponds entirely to QCD multijet events. The kink in (c) at about 80 GeV is due to one of the applied triggers (the asymmetric  $\tau$ -trigger, as described in section 7.1), which has its online selection threshold at this value of the transverse momentum of the leading  $\tau$ -lepton.

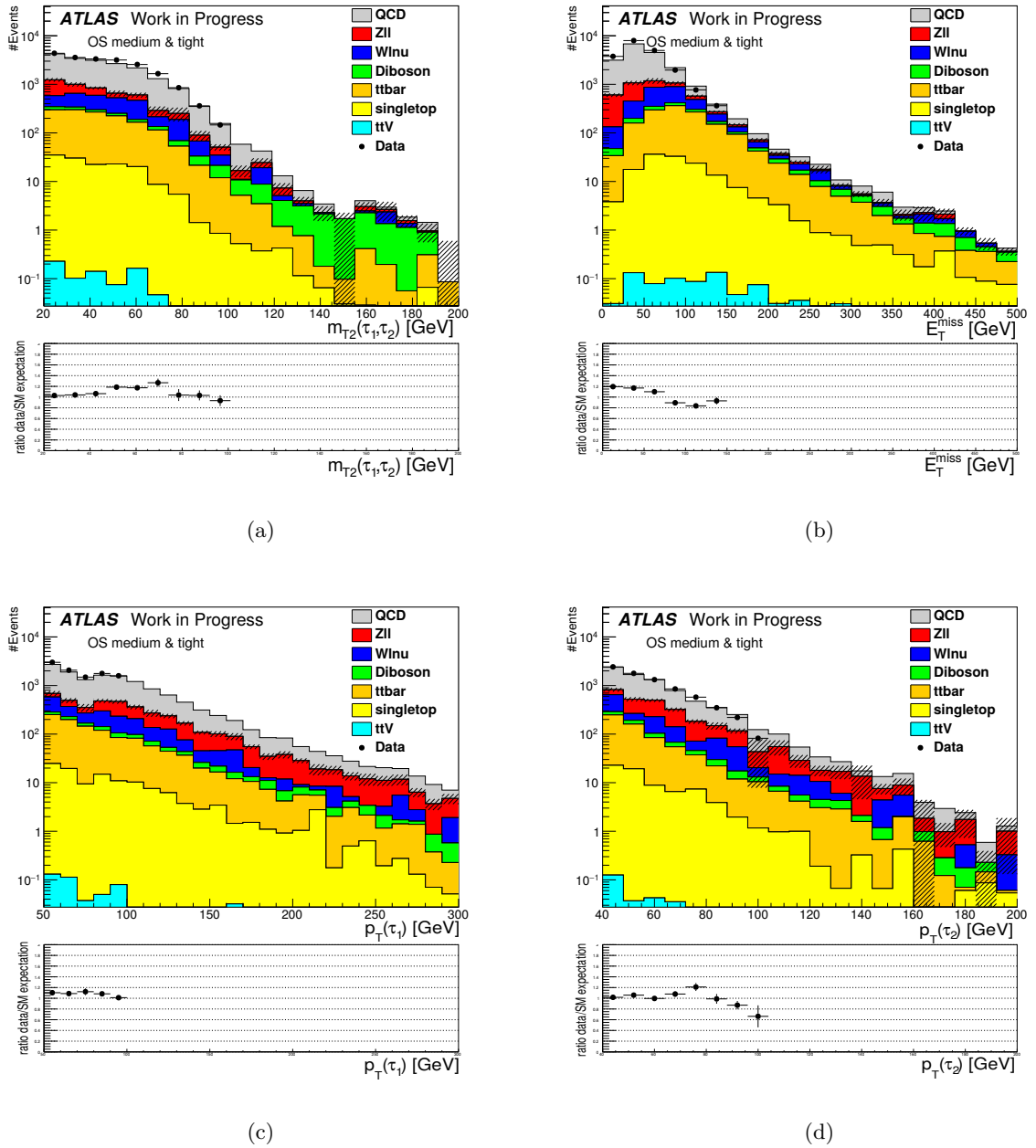


Figure 6.6.: Depicted here are the distributions for the  $m_{T2}$ -variable (a), the missing transverse energy (b) and the transverse momentum of the leading (c) and the next-to-leading  $\tau$ -lepton (d) for all Standard Model backgrounds (including the QCD estimate in grey) in comparison with data in region D. The comparison is only done for some of the lower valued bins in order not to unblind the whole phase space region. In the lower plots, the ratio of data to estimated backgrounds is given for the unblinded bins. It can be observed, that this ratio is mostly fluctuating around 1.





## 7. Preselection and Triggers

Although the discrimination of signal and background is done in a separate optimization procedure, a certain event selection has to be performed beforehand in order to define the final state that is searched for and to reduce the amount of events to a convenient level. As the final states are different, the two considered decay channels have separate preselection criteria and triggers.

### 7.1. Hadronic Decay of both $\tau$ -Leptons (HadHad-Channel)

To select events with two hadronically decaying  $\tau$ -leptons, a set consisting of two triggers is applied, where at least one of the triggers has to fire at an event. The first is called asymmetric  $\tau$ -trigger<sup>1</sup>, as it selects events in which the two leading  $\tau$ -leptons are showing a very high difference in their transverse momenta. The two  $\tau$ -leptons also have to pass the online medium requirement. Online, the trigger fires at transverse  $\tau$ -momenta of 80 GeV and 50 GeV for the leading and next-to-leading  $\tau$ -lepton, respectively. As the trigger efficiency increases with the  $\tau$ -momenta (the so-called turn-on curve) one has to apply certain offline cuts to make sure that the trigger efficiency is in the plateau region, i.e. in the region of stable optimal efficiency. For the asymmetric  $\tau$ -trigger the plateau region is reached at a transverse momentum of 95 GeV and 65 GeV, respectively, for the two  $\tau$ -leptons. The second trigger that is used for the HadHad-channel is the so-called ditau+MET trigger<sup>2</sup>, which does not only select events with two (online) medium  $\tau$ -leptons but also with missing transverse momentum above a certain threshold. In this case the online requirements are 35 GeV and 25 GeV for the transverse  $\tau$ -momenta and 50 GeV for the missing transverse energy. Again, the offline thresholds for the  $\tau$ -momenta lie 15 GeV above the online ones, i.e. at 50 GeV and 40 GeV, respectively. For the missing transverse energy the plateau region is reached at much higher values, resulting in an offline cut at 150 GeV. The two triggers are applied in a logical "or"-decision, i.e. at least one of them has to fire for an event to be selected. It has to be noted that the offline cut-values are rather high compared to those of the ditau-trigger which was used in run 1 ( $p_T(\tau_1) > 40$  GeV and  $p_T(\tau_2) > 25$  GeV [5]) and therefore already pose a great restriction to this analysis.

Apart from the online and offline trigger selection, each selected event is required to yield at least two  $\tau$ -leptons that pass the (offline) medium criteria. From these, the two  $\tau$ -leptons with the largest transverse momenta should have an oppositely signed electric charge.

In addition, for the cut-and-count analysis (section 8.2), events with b-tagged jets are rejected (b-veto) as well as events with an invariant mass of the two leading  $\tau$ -leptons below 80 GeV (Z-veto). The reason for choosing an invariant mass cut at 80 GeV to

<sup>1</sup>Asymmetric  $\tau$ -trigger: HLT\_tau80\_medium1\_tracktwo\_L1TAU60\_tau50\_medium1\_tracktwo\_L1TAU12

<sup>2</sup>Ditau+MET trigger: HLT\_tau35\_medium1\_tracktwo\_tau25\_medium1\_tracktwo\_xe50

account for events with  $Z$ -bosons ( $m_Z \approx 91 \text{ GeV}$ ) is that the hadronically decaying  $\tau$ -leptons are reconstructed based on the "visible" parts of their decay-products, whereas the  $\tau$ -neutrino escapes detection leading to a slightly lower visible  $Z$ -mass.

A summary of the preselection criteria and their respective event yields for the used samples is given in table 7.1.

## 7.2. One Leptonic and One Hadronic $\tau$ -Decay (LepHad-Channel)

In the case of the LepHad-channel, triggers for events with two  $\tau$ -leptons cannot be applied, as a leptonically decaying  $\tau$ -lepton is reconstructed as the respective light lepton, instead of the initial  $\tau$ -lepton. Therefore, a logical "or" of two light-lepton triggers is used, one for muons<sup>3</sup> and one for electrons<sup>4</sup>. As electrons and muons have lower fake rates, the triggers have much lower online and offline thresholds. In this case the online transverse momentum requirement on the light leptons is 24 GeV, with the offline plateau cut being slightly higher at 30 GeV.

Events passing at least one of these two triggers must also yield exactly one reconstructed medium  $\tau$ -lepton and one light lepton, that passes the "signal lepton" criteria. Additionally, the  $\tau$ - and the light lepton's electric charge should have opposite signs. Like in the preselection for the HadHad-channel, a b-veto and a Z-veto (again in the form of a cut on the invariant mass of the two leptons at 80 GeV) are applied.

The cutflow of the preselection for the LepHad channel is shown in table 7.2.

## 7.3. Comparison of HadHad- and LepHad-Channels

From tables 7.1 and 7.2 it can be observed that the preselection seems to be a lot more efficient for the HadHad-channel than for the LepHad-channel, which is also illustrated in figures 7.1 and 7.2. This is due to two main reasons.

First, the LepHad-channel has a higher total background due to the larger branching fractions for some backgrounds in this channel, whereas the branching fraction for the signal is the same in the LepHad and the HadHad case. The largest effect originates from the  $W$ -boson background. Due to the fact that a  $W$ -boson decays two times more often into a light lepton than into a  $\tau$ -lepton, the branching ratio for  $W$ +jets is notably increased. Not taking into account any reconstruction efficiencies, the probability for a  $W$ -boson to decay hadronically is 2/3 and 1/3 for a leptonic decay. This leads to the following probabilities for a  $W$  boson to yield a hadronic  $\tau$ -lepton or a light lepton in the final decay step:

<sup>3</sup>Single muon trigger: HLT\_mu24\_ivarmedium

<sup>4</sup>Single electron trigger: HLT\_e24\_lhtight\_nod0\_ivarlose

$$p(\tau_{\text{had}}) = p(W_{\text{leptonic}}) \cdot p(\text{lepton} = \tau) \cdot p(\tau \rightarrow q\bar{q}'\nu_\tau) \quad (7.1)$$

$$= \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{27} \quad (7.2)$$

$$p(e/\mu) = p(W_{\text{leptonic}}) \cdot \left( p(\text{lepton} = e/\mu) + p(\text{lepton} = \tau) \cdot p(\tau \rightarrow l\nu_l\nu_\tau) \right) \quad (7.3)$$

$$= \frac{1}{3} \cdot \left( \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} \right) = \frac{7}{27}. \quad (7.4)$$

Taking all of these probabilities together one arrives at a branching ratio for  $W$ +jets in the LepHad-channel which is by a factor of 3.5 higher compared to the HadHad-channel.

Second, the trigger thresholds are a lot lower for the single lepton triggers than for the ditau triggers. That has not only the effect of more background events passing the trigger selection. From figure 7.3 it can be seen that light leptons from a direct stau signal accumulate at low values of the transverse momentum. This is due to the fact, that an additional neutrino is carrying away momentum (which is not the case in a hadronic  $\tau$ -decay), resulting in a light lepton that can be much softer than the original  $\tau$ -lepton. Therefore, any cut on the transverse momentum will lead to a stronger decrease in the amount of selected signal events compared to a corresponding cut on a hadronically decaying  $\tau$ -lepton.

Cut	Signal	$W$ +jets	$Z$ +jets	Diboson	Top	$S/B$
online triggers	179.26	87572.90	47254.11	6012.87	55834.58	$9.1 \cdot 10^{-4}$
$N(\tau_{\text{medium}}) \geq 2$	64.31	4753.66	12104.42	698.38	3963.75	$3.0 \cdot 10^{-3}$
OS	63.46	3300.16	11625.28	599.55	3300.72	$3.4 \cdot 10^{-3}$
offline cuts	36.12	461.55	3163.66	127.98	495.43	$8.5 \cdot 10^{-3}$
$N(\text{b-jets}) = 0$	34.87	435.99	2947.48	114.20	82.23	$9.7 \cdot 10^{-3}$
$m_{\text{inv}}(\tau_1, \tau_2) > 80 \text{ GeV}$	34.00	430.35	1961.25	71.60	77.68	$1.3 \cdot 10^{-2}$

Table 7.1.: Cutflow table with the number of events for each process after each of the preselection cuts at an integrated luminosity of  $35 \text{ fb}^{-1}$ . As an example for a signal, the mass point with  $m_\tau = 200 \text{ GeV}$  and  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$  is given. The last column shows the ratio of the event yields of the same mass point and the sum of all background events, as a measure of the selection efficiency.

Cut	Signal	$W$ +jets	$Z$ +jets	Diboson	Top	$S/B$
online triggers	306.34	$1.634 \cdot 10^6$	$1.071 \cdot 10^6$	50669.85	227091.30	$1.0 \cdot 10^{-4}$
$N(\tau_{\text{medium}}) = 1$	304.38	$1.628 \cdot 10^6$	$1.068 \cdot 10^6$	50111.60	225378.23	$1.0 \cdot 10^{-4}$
$N(e_{\text{signal}}/\mu_{\text{signal}}) = 1$	224.12	$1.27 \cdot 10^6$	824449.14	43873.23	198529.62	$9.6 \cdot 10^{-5}$
OS	220.53	871529.37	700700.08	36755.52	165915.12	$1.2 \cdot 10^{-4}$
offline cuts	198.51	753468.16	536958.58	33309.39	152634.49	$1.3 \cdot 10^{-4}$
$N(\text{b-jets}) = 0$	191.25	736440.82	523666.34	32077.61	24457.44	$1.5 \cdot 10^{-4}$
$m_{\text{inv}}(\tau, e/\mu) > 80 \text{ GeV}$	158.34	396486.79	295673.71	21226.50	16384.54	$2.2 \cdot 10^{-4}$

Table 7.2.: Cutflow of the LepHad-preselection with the event yields for each sample at an integrated luminosity of  $35 \text{ fb}^{-1}$ . In analogy to table 7.1 the last column gives the efficiency of the signal selection compared to the sum of all backgrounds. The signal sample used for this comparison is that for the mass point with  $m_{\tilde{\tau}} = 120 \text{ GeV}$  and  $m_{\tilde{\chi}_1^0} = 0 \text{ GeV}$ . It can be noted that the signal efficiency is in general much lower than in the HadHad-channel (see table 7.1) which is mostly due to tighter triggers and offline cuts in the HadHad-channel. The first three cuts, which are necessary to define the final state, even decrease the ratio of signal and background events. However, about  $\frac{1}{3}$  of the  $W$ +jets events do not pass the opposite sign criterium, which again raises that ratio.

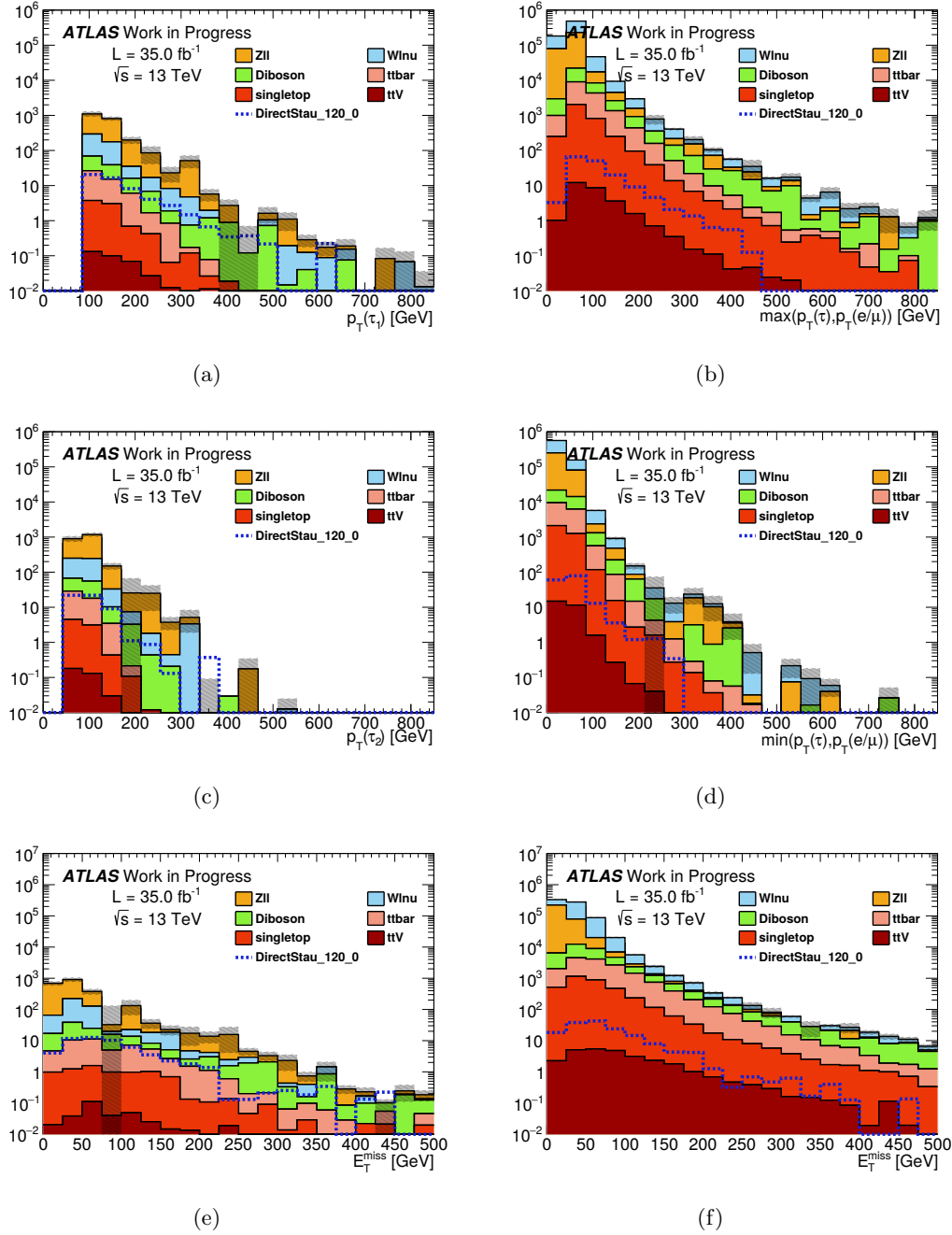


Figure 7.1.: Comparison of the transverse momenta of the leading ((a),(b)) and next-to-leading lepton ((c),(d)) as well as the missing transverse energy ((e),(f)) for the HadHad-channel (left) and LepHad-channel (right) with the respective preselection applied. The dotted blue line corresponds to the expected distributions for the signal mass point ( $m_{\tilde{\tau}} = 120$  GeV,  $m_{\tilde{\chi}_1^0} = 0$  GeV). For this comparison only the asymmetric  $\tau$ -trigger decision was requested for HadHad-events.

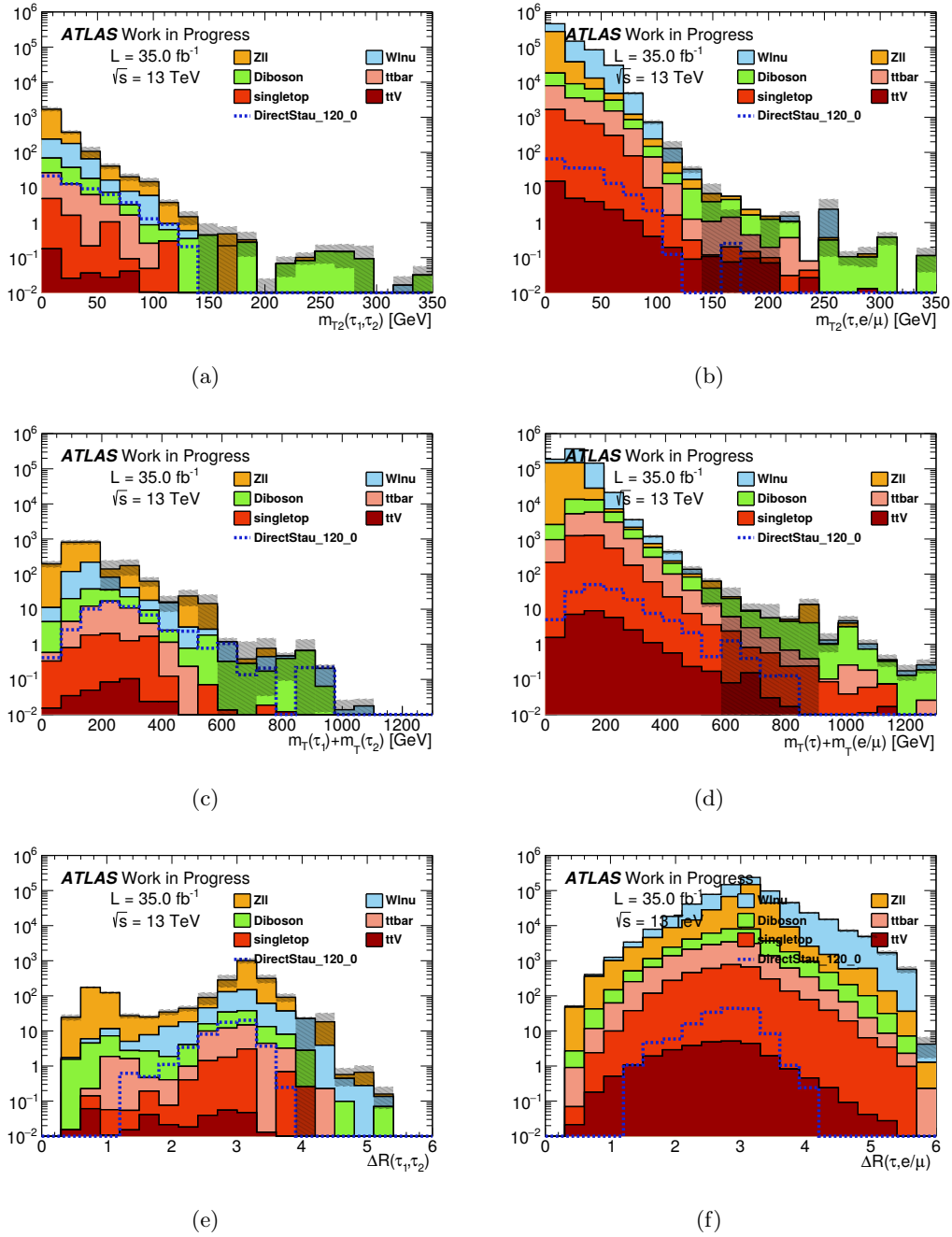


Figure 7.2.: Comparison of  $m_{T2}$  ((a),(b)), the sum of the transverse lepton masses ((c),(d)) and the distance in the  $\eta$ - $\phi$ -plane between the leptons ((e),(f)) for the HadHad-channel (left) and the LepHad-channel (right). The distributions for the signal mass point ( $m_{\tilde{\tau}} = 120 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 0 \text{ GeV}$ ) are represented by the dotted blue lines. For both channels the respective preselection is applied. For this comparison only the asymmetric  $\tau$ -trigger decision was requested for HadHad-events. It can be noted that the amount of background events gets far more reduced by the preselection in the HadHad- than in the LepHad-channel.

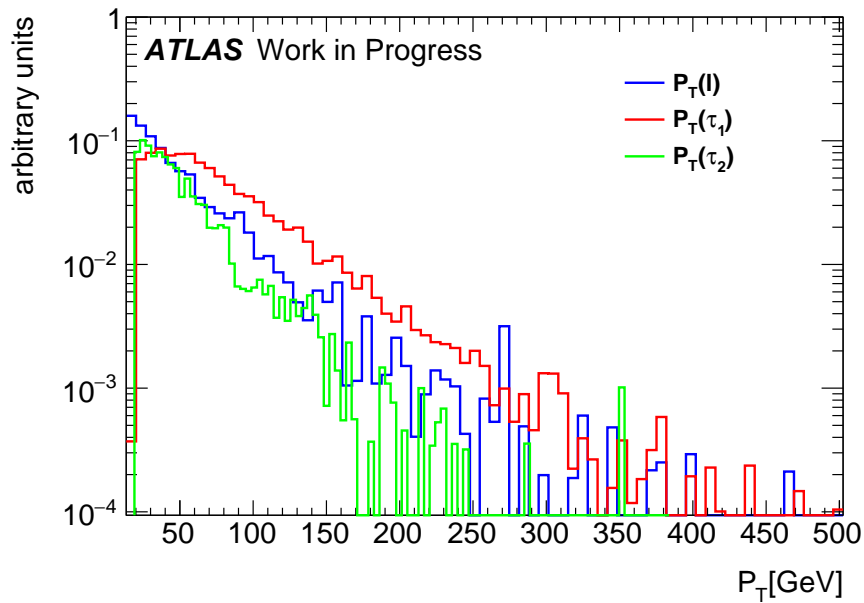


Figure 7.3.: In this histogram the normalized distribution of the transverse momenta of a light signal lepton (electron or muon) is compared to that of hadronically decaying  $\tau$ -leptons. The distribution for the light lepton is depicted in blue and is taken from the LepHad-channel selection of the mass point with  $m_{\tilde{\tau}} = 120$  GeV and  $m_{\tilde{\chi}_1^0} = 0$  GeV. The red and the green line correspond to the leading and next-to-leading  $\tau$ -leptons of the HadHad selection of the same mass point. Here, no preselection except the online trigger decisions (for the HadHad-channel only the asymmetric  $\tau$ -trigger is used in this particular case) and the respective criteria on the number of leptons are applied. The fraction of events at a low transverse momentum range is higher for light leptons compared to  $\tau$ -leptons. This means that a lower cut on the transverse momentum eliminates more events for the LepHad-channel than a corresponding cut would do in the HadHad case.





## 8. Signal Region Optimization

In the following sections it is studied if a selection can be defined with which a signal could be identified over the estimated Standard Model background. To decide whether a certain set of cuts provides a reliable discrimination, a measure for statistical significance is needed.

### 8.1. Statistical Significance

Given there is a signal region SR in which after certain cuts a number of  $b$  background events is expected as well as  $s$  signal events. When applying the same cuts to data events one obtains an amount of  $N_{\text{obs}}$  events. How can one tell if  $N_{\text{obs}}$  is more compatible with  $s+b$  than with  $b$ ? In the following,  $b$  will be referred to as the "background-only" hypothesis and  $s+b$  as the "signal+background" hypothesis. A common statistical measure for the probability of one of these hypotheses being true is the so-called p-value  $p$ . The p-value is defined as the probability that, given a hypothesis is true, an experiment will give a certain result or even a more definite one. As it is a counting experiment the probability density will be generally given by a Poisson distribution  $P(x|\mu)$  with a mean value of  $\mu$ :

$$P(x|\mu) = \frac{\mu^x e^{-\mu}}{x!} \quad (8.1)$$

For the p-value this means:

$$p(N_{\text{obs}}|\mu) = \int_{N_{\text{obs}}}^{\infty} P(N|\mu) dN, \quad (8.2)$$

where  $\mu$  can be either  $b$  or  $s+b$  and  $P(N|\mu)$  the probability for an outcome of  $N$  events given the respective hypothesis is true. By convention, a hypothesis is rejected if the p-value is 0.05 or smaller. This is also called exclusion by 95% confidence level (CL). Typically, the confidence level is expressed in terms of Gaussian standard deviations. A 95% confidence level is translated into an exclusion by  $2\sigma$ . The number of standard deviations by which a hypothesis can be rejected is denoted as  $z$  or  $z_N$ . This z-value is also referred to as significance and related to the p-value by the following equation:

$$z = \sqrt{2} \cdot \text{erf}^{-1}(1 - 2p) \quad (8.3)$$

However, as the negative part of a Gaussian distribution does not have any physical meaning only a one-sided Gaussian is used in an analysis. This changes the value of  $z$  for a 95% confidence level from 2 to approximately 1.64. As data should not be shown in a candidate signal region at this stage to avoid biasing the analysis,  $N_{\text{obs}}$  is assumed to be the number of expected background events.

In these studies a selection should be found which leads to a number of signal and background events such that the  $z$ -value is at least 1.64 in order to exclude the hypothesis of a signal model being true. For the calculation of the  $z$ -value a predefined function in the RooStats-package [36] is used which takes the expected number of signal and background events as an input, as well as the statistical and systematic uncertainty on the latter. As the systematic uncertainties are not determined in these studies, a flat value of 30% is assumed, which is in accordance with the run 1 total systematic uncertainty [5].

## 8.2. Cut-and-Count Method

One way to discriminate signal and background is to find more or less optimal, independent cuts on different variables additional to the preselection. In the cut-and-count method, a set of variables with a certain number of cuts is defined. Then an algorithm calculates the significance for each of the possible cut combinations for one of the signal samples. The combination with the maximum significance is selected and applied to all of the other signal samples. However, there are some requirements that the signal region must fulfill, apart from maximum significance, in order to avoid cutting on statistical fluctuations. Therefore, a minimum number of 2 background events is demanded with a maximum total statistical background uncertainty of 50%. Large statistical fluctuations can be observed in particular for  $W$ +jets and  $Z$ +jets, originating from the small number of unweighted events surviving the cuts and on some events with (large) negative event weights. Negative event weights are e.g. generated by Sherpa. Therefore, the number of events is required to be larger than 0 for all processes involved.

The cut combination is afterwards applied to all other signal points and the significance for each of them is calculated. The respective significances can then be plotted in the plane of the stau and neutralino mass. Another way to illustrate the obtained cut combination are so-called N-1 plots. Each of them shows the distribution of a particular variable with all of the signal region cuts applied except the cut performed on the plotted variable itself, which is indicated by a red line.

### 8.2.1. HadHad-Channel

In the HadHad-channel the optimization has been performed with cuts on the missing transverse energy, the  $m_{T2}$ -variable, the distance in the  $\eta$ - $\phi$ -plane between the two leading  $\tau$ -leptons, the transverse momenta of the leading and next-to-leading  $\tau$ -leptons, their invariant mass, the sum of their transverse masses and the absolute value of  $\cos\theta_{ll}$ . The exact cut values that have been set for the algorithm to choose from are listed in table 8.1.

#### Low Stau Mass

It is reasonable to start with the optimization at low stau masses, which have the largest cross-section and where the sensitivity was best in run 1. The first signal region candidate is optimized for a stau mass of 100 GeV and a lightest neutralino mass of 1 GeV. However, the cut combination obtained by the exhaustive search of cut combinations (see table 8.2) shows significances of about 1, as can be seen in figure 8.1, which is only insufficient

Variable	Possible cuts
$E_T^{\text{miss}} >$	0, 20, 50, 70, 100, 120, 150, 200 GeV
$m_{T2}(\tau_1, \tau_2) >$	20, 25, 30, 35, 40, 50, 75, 100 GeV
$\Delta R(\tau_1, \tau_2) <$	1.0, 1.5, 2.0, 2.3, 2.4, 2.5, 2.6, 4.0, 6.0
$p_T(\tau_1) >$	50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 GeV
$p_T(\tau_2) >$	40, 50, 65, 70, 75, 85, 95, 100 GeV
$m_{\text{inv}}(\tau_1, \tau_2) >$	80, 90, 100, 110, 120, 140 GeV
$m_T(\tau_1) + m_T(\tau_2) >$	0, 50, 100, 200, 250, 270, 300 GeV
$ \cos \theta_U  <$	0.2, 0.4, 0.6, 0.7, 0.8, 1.0

Table 8.1.: List of cuts for each variable that can be permuted in the optimization algorithm for the HadHad-channel. It has to be noted that the lowest cuts for the transverse momenta are set to the offline thresholds of the ditau+MET trigger, but the trigger decisions together with the corresponding offline cuts are performed beforehand. So, if the cut on the leading (next-to-leading) transverse  $\tau$ -momentum is set between 50 and 90 GeV (40 and 65 GeV) this will only have an impact on the events selected by the ditau+MET trigger. Similarly, a cut on the missing transverse energy below a value of 150 GeV would only affect events selected by the asymmetric tau trigger.

sensitivity. Therefore the obtained cut combination would not be suitable as a signal region. Moreover, the number of expected background events is larger than the expected event yield for the signal mass point which the cuts are optimized for, which is also illustrated in the N-1 plots in figures B.1 and B.2.

SR DS HadHad Low Mass
$m_{T2}(\tau_1, \tau_2) > 30 \text{ GeV}$
$p_T(\tau_1) > 140 \text{ GeV}$
$m_{\text{inv}}(\tau_1, \tau_2) > 120 \text{ GeV}$
$m_T(\tau_1) + m_T(\tau_2) > 270 \text{ GeV}$
$ \cos \theta_U  < 0.6$

Table 8.2.: Cut combination optimized for the signal mass point ( $m_{\tilde{\tau}} = 100 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ). Redundant cuts are not shown in this list.

### Intermediate Stau Mass

The following optimization was done for a mass point at an "intermediate" scale, namely at a stau mass of 140 GeV and neutralino mass at 1 GeV. The cut combination selected by the algorithm is shown in table 8.4 with the event yields for each process listed in table 8.5. Even though the significances obtained with it (see figure 8.2) are higher than those for a low stau mass optimization, the highest z-values are still below 1.64, making an exclusion by 95% CL unlikely. This is also illustrated in the N-1 plots in figures B.3 and B.4.

Process	Events (statistical uncertainty)
$W$ +jets	8.36 (79.31%)
$Z$ +jets	2.01 (48.76%)
Diboson	1.75 (49.71%)
Top	1.41 (74.47%)
QCD	4.64 (45.04%)
Total background	18.17 (39.00%)
Signal	16.76 (10.08%)

Table 8.3.: Expected event yields in the "SR DS HadHad Low Mass" region for each background process and for the signal mass point ( $m_{\tilde{\tau}} = 100$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The total number of expected background events exceeds the number of signal events, although they are about equal within the statistical uncertainties. The dominant backgrounds in this region are  $W$ +jets and the QCD multijet background.

#### SR DS HadHad Intermediate Mass

$$\begin{aligned}
& E_{\text{T}}^{\text{miss}} > 70 \text{ GeV} \\
& m_{\text{T}2}(\tau_1, \tau_2) > 50 \text{ GeV} \\
& p_{\text{T}}(\tau_1) > 90 \text{ GeV} \\
& p_{\text{T}}(\tau_2) > 70 \text{ GeV} \\
& m_{\text{inv}}(\tau_1, \tau_2) > 110 \text{ GeV} \\
& m_{\text{T}}(\tau_1) + m_{\text{T}}(\tau_2) > 270 \text{ GeV} \\
& |\cos \theta_{ll}| < 0.8
\end{aligned}$$

Table 8.4.: Cut combination optimized for the signal mass point ( $m_{\tilde{\tau}} = 140$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV).

### High Stau Mass

The best signal region is found when optimizing for high stau masses with the signal mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV). The obtained cut combination is shown in table 8.6, with the separate event yields for each process in table 8.7. The corresponding values for the significance for each signal mass point are depicted in the plane of the stau and lightest neutralino mass in figure 8.3. Two of the mass points can be observed to yield significances slightly above 1.64, but with large statistical uncertainties, especially on  $W$ +jets to be considered. The N-1 plots are shown in figures B.5 and B.6.

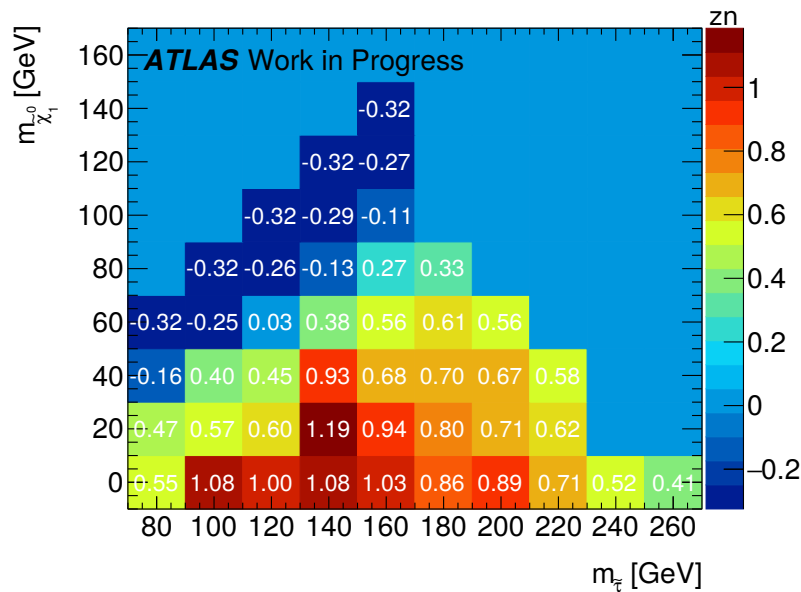


Figure 8.1.: Significance plot in the plane of stau and neutralino mass optimized for ( $m_{\tilde{\tau}} = 100$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The highest z-values are reached for a stau mass of 140 GeV at a neutralino mass of 20 GeV and for stau masses of 100 GeV and 140 GeV at a neutralino mass of 1 GeV. However, as these values are far below 1.64, probably no exclusion would be possible if the cuts from table 8.2 were applied to data.

Process	Events (statistical uncertainty)
$W$ +jets	3.63 (95.32%)
$Z$ +jets	1.01 (60.40%)
Diboson	1.16 (65.52%)
Top	2.44 (36.07%)
QCD	4.10 (36.83%)
Total background	12.35 (32.41%)
Signal	13.41 (10.44%)

Table 8.5.: Expected event yields in region "SR DS HadHad Intermediate Mass" for each background process and for the signal mass point ( $m_{\tilde{\tau}} = 140$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . In this region the number of signal events is slightly higher than the amount of background events, although they are about equal within the statistical uncertainties. The dominant backgrounds in this region are  $W$ +jets and the QCD multijet background. It has to be noticed that the statistical uncertainty on the  $W$ +jets background is especially large. This is due to the small number of unweighted simulated events that are left after the cuts. Also, some of these events show negative event weights which additionally increases the uncertainty.

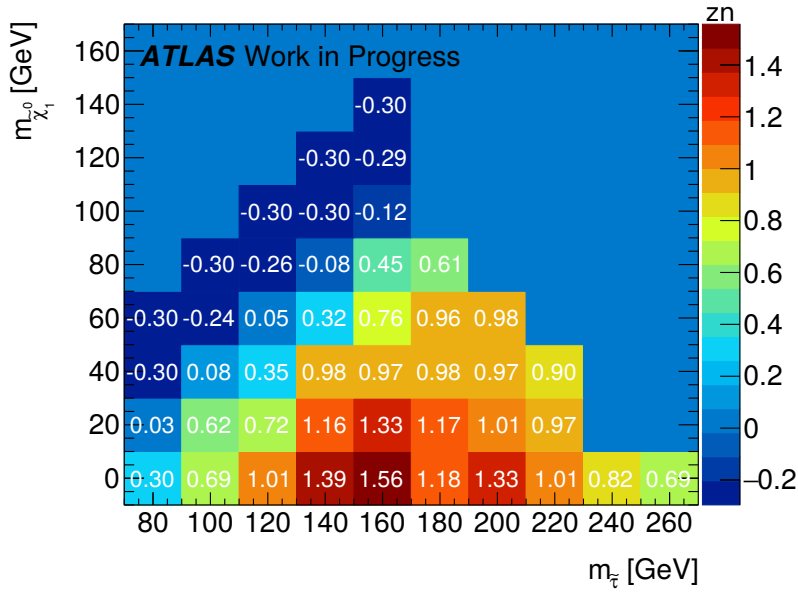


Figure 8.2.: Significance plot in the plane of stau and neutralino mass for a region optimized on  $(m_{\tilde{\tau}} = 140 \text{ GeV}, m_{\tilde{\chi}_1^0} = 1 \text{ GeV})$  at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The highest  $z$ -values are reached for stau masses of 160 GeV and 140 GeV at a neutralino mass of 1 GeV. The  $z$ -values are still below 1.64, which is probably not sufficient for an exclusion if the cuts from table 8.4 are applied to data.

SR DS HadHad High Mass

$$\begin{aligned}
 & E_{\text{T}}^{\text{miss}} > 70 \text{ GeV} \\
 & m_{\text{T}2}(\tau_1, \tau_2) > 100 \text{ GeV} \\
 & p_{\text{T}}(\tau_1) > 110 \text{ GeV} \\
 & m_{\text{inv}}(\tau_1, \tau_2) > 90 \text{ GeV} \\
 & |\cos \theta_{ll}| < 0.7
 \end{aligned}$$

Table 8.6.: Signal region definition optimized for the mass point  $(m_{\tilde{\tau}} = 200 \text{ GeV}, m_{\tilde{\chi}_1^0} = 1 \text{ GeV})$ . For the variables listed in table 8.1 which do not appear in this list the selected cut values do not have an impact on the event yields.

Process	Events (statistical uncertainty)
$W$ +jets	0.27 (88.89%)
$Z$ +jets	1.13 (67.26%)
Diboson	0.94 (68.42%)
Top	0.37 (83.78%)
QCD	0.93 (78.49%)
Total background	3.63 (35.60%)
Signal	7.10 (9.01%)

Table 8.7.: Expected event yields in the region "SR DS HadHad High Mass" for each background process and for the signal mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The total number of expected background events exceeds the number of signal events, even if the statistical uncertainties are taken into account. The dominant background is  $Z$ +jets, although the event yields do not differ much among the background processes. Again, the largest uncertainty is due to the  $W$ +jets background, followed by processes involving top quarks and multijet processes.

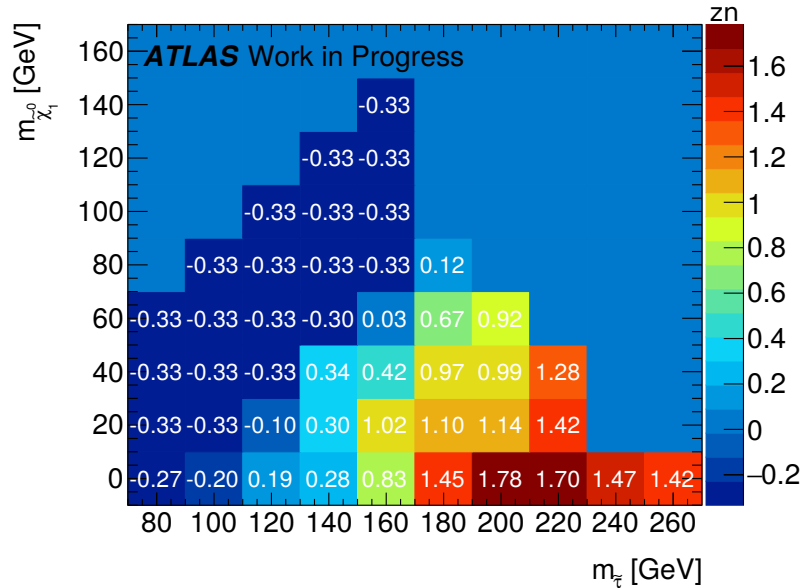


Figure 8.3.: Significance for each mass point in signal region DS High Mass at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . Two mass points, namely ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV), for which the region was optimized, and ( $m_{\tilde{\tau}} = 220$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) show z-values slightly above 1.64.

### 8.2.2. LepHad-Channel

For the signal region optimization in the LepHad-channel the  $m_{T2}$ -variable has been used, as well as the missing transverse momentum, the transverse momenta of the  $\tau$ - and the light leptons, the sum of their transverse masses, the distance in the  $\eta$ - $\phi$ -plane and the difference in the azimuthal angle. Additionally, a cut on the number of jets can be placed. The cut values that are permuted in the algorithm are listed in table 8.8. It has to be noted, that only the Monte Carlo based estimations of Standard Model backgrounds are considered because the estimation of the QCD background as it has been shown in section 6.2.5 is constructed only for the HadHad-channel. However, the presence of one signal lepton already should strongly suppress multijet processes.

The best result has been obtained with the signal point  $m_{\tilde{\tau}} = 140$  GeV and  $m_{\tilde{\chi}_1^0} = 0$  GeV. The resulting cut combination is shown in table 8.9.

As can be seen from the significance plot, figure 8.4, not a single signal mass point shows a significance larger than 0. The occurrence of negative z-values is a feature of the RooStats function [36] used for calculation. They have the same meaning as a significance of exactly 0, namely that there is no sensitivity at all for a direct stau signal in the LepHad-channel. The cuts cannot isolate a hypothetical signal from the overwhelming amount of background events, which can also be observed in the N-1 plots in figures B.7 and B.8.

Variable	Possible cuts
$E_T^{\text{miss}} >$	0, 20, 60, 100, 125, 135, 145 GeV
$m_{T2}(\tau, e/\mu) >$	0, 5, 10, 20, 30 GeV
$m_T(\tau) + m_T(e/\mu) >$	0, 100, 200, 250, 300 GeV
$\Delta R(\tau, e/\mu) <$	2.0, 2.5, 3.0, 3.2, 2.5, 4.0, 6.0
$p_T(\tau) >$	30, 50, 70, 90, 110 GeV
$p_T(e/\mu) >$	30, 50, 70, 90, 110 GeV
$ \Delta\phi(\tau, e/\mu)  >$	0.0, 1.0, 1.5, 2.0, 2.5, 2.7
$N(\text{jets}) \leq$	0, 1, 2, 3, 4

Table 8.8.: Possible cuts for the permutation algorithm to select from for the LepHad-channel.



SR DS LepHad
$m_{T2}(\tau, e/\mu) > 20 \text{ GeV}$
$m_T(\tau) + m_T(e/\mu) > 300 \text{ GeV}$
$\Delta R(\tau, e/\mu) < 3.2$
$p_T(\tau) > 110 \text{ GeV}$
$p_T(e/\mu) > 110 \text{ GeV}$
$ \Delta\phi(\tau, e/\mu)  > 1.5$
$N(\text{jets}) \leq 2$

Table 8.9.: Cut combination with the highest significance for the signal mass point  $m_{\tilde{\tau}} = 140 \text{ GeV}$  and  $m_{\tilde{\chi}_1^0} = 0 \text{ GeV}$  as selected by the cut permutation algorithm for the LepHad-channel. There is no cut listed on the missing transverse energy as the permutation resulted in a "cut"  $E_T^{\text{miss}} > 0 \text{ GeV}$ .

Process	Events (statistical uncertainty)
$W$ +jets	43.08 (17.55%)
$Z$ +jets	2.02 (69.80%)
Diboson	28.21 (18.50%)
Top	18.80 (74.47%)
Total background	92.11 (10.52%)
Signal	4.44 (23.65%)

Table 8.10.: Expected event yields for each background process and for the signal mass point ( $m_{\tilde{\tau}} = 140 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 0 \text{ GeV}$ ) in the LepHad-channel scaled to an integrated luminosity of  $35 \text{ fb}^{-1}$ . The total number of expected background events exceeds the number of signal events considerably. The dominant background is  $W$ +jets. QCD multijet processes are not considered for the LepHad-channel.

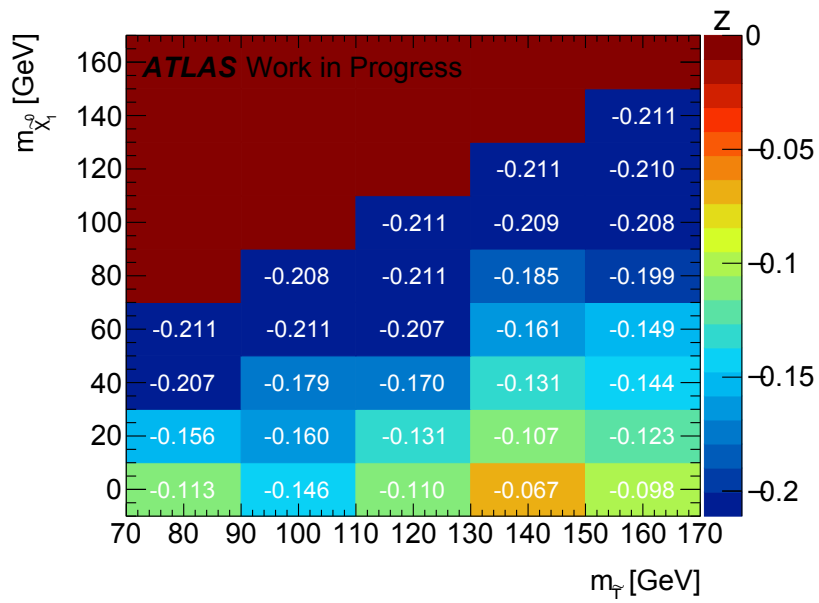


Figure 8.4.: Distribution of the resulting significance for each signal mass point with the cut combination defined in table 8.9 for the LepHad-channel in the plane of the stau and neutralino mass at an integrated luminosity of  $35 \text{ fb}^{-1}$ . As all of the significances are below 0, no sensitivity is achieved for the LepHad direct stau signal with this cut combination.

### 8.3. Multivariate Analysis

As the cut-and-count method seems to deliver mostly insufficient results concerning the discrimination of the direct stau signal and the corresponding Standard Model background, another method has to be tested. A possibility for this are so-called multivariate methods, like neural networks or boosted decision trees (BDTs) [37, 38].

In the following, the BDT method is explained and some studies on the discrimination power for the direct stau HadHad-channel are presented. As the LepHad-channel shows no sensitivity at all, the focus will be set only on the HadHad-channel.

#### 8.3.1. Boosted Decision Trees

The difference between the cut-and-count method and a single decision tree is very small. In both methods a number of successive cuts is performed in order to obtain the best possible signal efficiency and background rejection. However, in a decision tree the cut chain is split after every cut leading to a branch-like structure (see figure 8.5). There are two possibilities to optimize such a decision tree. Either each cut in one decision tree is optimized resulting in one rather powerful discriminating tree or multiple decision trees, called forest (which do not have to be optimized) are created and combined afterwards. The latter is referred to as a boosted decision tree (BDT) which will be explained in the following.

As input, dedicated datasets for signal and background are needed as well as information on each event contained in these samples in the form of (kinematic) variables. Each event  $i$  can be described as a tuple  $(\vec{x}_i, y_i)$ , with  $\vec{x}_i$  being a collection of certain input variables and  $y_i$  the corresponding label for event  $i$  (e.g. +1 for "signal" and -1 for "background"). The algorithm now compares the distributions of the given variables and decides on an optimal cut on the best discriminating variable. The number of events above and below this cut are then counted for signal and background. If the number of signal events exceeds the background events by a certain threshold value, the corresponding region in phase space is declared as "signal-like" and no further cuts are performed. The same holds for "background-like" regions of phase-space. In case that neither signal nor background dominate the region, the next best variable is selected and a cut on it is set. The same procedure is repeated until either a region is found to be signal- or background-like or a maximum number of decisions is reached. Hence, the so-called maximum depth of a tree serves as an important parameter for the construction of a decision tree.

For a decision tree as described previously, the following issue can occur: Assuming two variables have approximately the same discrimination power, then one of them might be preferred by the decision tree merely due to a statistical fluctuation in its distribution. This can lead to an entirely different tree structure compared to the one designed in absence of this fluctuation. To prevent such cases it can be useful to design a whole forest of decision trees and combining them by some means. After one decision tree is constructed as described above, the algorithm checks which events have been classified correctly as signal- or background-like and which events have been misidentified by comparing the hypothesis  $h(x_i)$  of the tree with the label  $y_i$ . Then, the events are assigned a certain weight before the samples are reused in the next decision tree design, which also slightly changes the variables' distributions. In this way, events that are wrongly identified get a

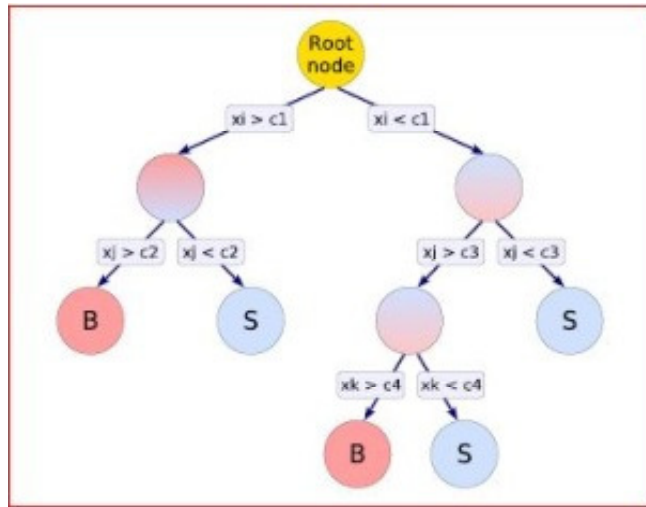


Figure 8.5.: Schematic principle of a decision tree. Starting at the root node, the best discriminating variable is selected to perform a cut. The regions above and below the cut value are checked for its fraction of signal and background events and declared signal- or background-like if the respective fraction exceeds a certain threshold. Otherwise, an additional cut is performed on the second best discriminating variable, repeating the steps until either signal or background dominate the region, or the number of cuts reaches a maximum value, which is a fixed parameter. Ultimately, the phase space spanned by the input variables is separated into dedicated regions in which mainly signal or background are expected [38].

higher priority in the training of the next tree. Repeating this, a certain number of trees (which is a predefined parameter for the BDT) is formed. Boosting, as this procedure is called, improves the statistical stability of the final BDT and can also enhance the separation of signal and background. A total output, referred to as "BDT response", of the decision tree forest is generated by taking the weighted average of every single classification, resulting in one final distribution with enhanced separation power.

In the standard boosting method (called adaptive boosting or AdaBoost algorithm), the original event weights  $w_{t,i}$  used in decision tree  $t$  are multiplied by an additional weight depending on a parameter  $\alpha_t$ , which is calculated from the misclassification rate  $\epsilon_t$  of the decision tree:

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (8.4)$$

Afterwards, the event weights are modified for the next decision tree  $t + 1$ :

$$w_{t+1,i} = \frac{w_{t,i}}{Z_t} \cdot e^{-\alpha_t h_t(\vec{x}_i) y_i} \quad (8.5)$$

This results into a decreased weight for correctly classified events and a larger weight if the hypothesis does not match the label  $y_i$ . In order to conserve the sum of all event weights, the whole sample is renormalized by a factor  $Z_t$  before the next training. The total boosted classification hypothesis  $H(\vec{x})$  for an event defined by the input variable values  $\vec{x}$  can be written as the weighted average of all hypotheses  $h_t$ :

$$H(\vec{x}_i) = \frac{1}{N_{\text{trees}}} \cdot \sum_{t=1}^{N_{\text{trees}}} \alpha_t \cdot h_t(\vec{x}_i). \quad (8.6)$$

$N_{\text{trees}}$  is the number of decision trees that are combined and  $h_t(\vec{x}_i)$  the classification of an event by each individual tree, respectively.

It can be shown that the misidentification rate of the combined hypothesis  $H(\vec{x}_i)$  decreases very rapidly with the number of iterations, i.e. the number of decision trees, if the individual misidentification rates  $\epsilon_t$  are only slightly better than random guessing ("weak learners") [39]. In the case that the decision trees only have to separate between two possible labels  $+1$  and  $-1$ , this means that  $\epsilon_t$  has to be only slightly smaller than  $\frac{1}{2}$  for most decision trees. Moreover, if the number of decision trees is of the order of  $\ln k$ , with  $k$  being the number of events used to train the BDT, the combined boosted hypothesis will almost perfectly match the events' labels. Nevertheless, this is not to be seen as an advantage because also the complexity of the final hypothesis increases with the number of trees at the cost of lacking generality. For a different sample of events, referred to as "testing" sample, this can result in a strong increase of the misidentification rate, making the BDT, in spite of its perfect match, useless (see figure 8.6). This phenomenon is referred to as overtraining, meaning that the algorithm is trained on each particular event instead of general properties of the signal- and background-class, making it again sensitive to statistical fluctuations. Therefore the BDT parameters like maximum depth and the number of trees have to be optimized to yield a misidentification rate as low as possible in the training, keeping the error on the testing sample small at the same time.

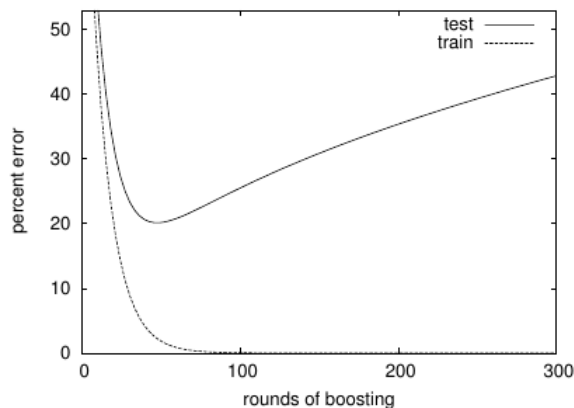


Figure 8.6.: Development of the misidentification rate for the training and the testing sample with increasing number of boosting rounds, i.e. number of decision trees for a dummy example. It can be observed that inspite of the continuously dropping error rate for the training sample, the misidentification rate in the testing sample is increasing again at some point. In this regime the BDT is overtrained [39].

Previously, the AdaBoost algorithm has been described approximately in the way it was developed by R. Schapire and Y. Freund [37]. However, the whole procedure can also be understood as a minimization of a dedicated loss function, which is the so-called exponential loss

$$L(H, y) = \frac{1}{k} \sum_{i=1}^k e^{-H(\vec{x}_i) \cdot y_i} \quad (8.7)$$

for a sample of  $k$  training events, as first indicated by L. Breiman [40]. This loss function, although it performs well in general, has some shortcomings with respect to outliers. As the loss function can be chosen freely, it is possible to design different boosting algorithms basing on the minimization of another loss function. In these studies, the GradientBoost algorithm has been used, which minimizes the following loss function:

$$L(H, y) = \ln \left( 1 + e^{-2H(\vec{x}_i) \cdot y_i} \right) \quad (8.8)$$

This minimization has to be performed numerically by a steepest-descent approach. Each decision tree  $h_{t+1}(\vec{x}_i)$  is grown such that it iteratively decreases the above loss function of the latest combined classifier  $H_t$ . After each iteration, a weight  $\gamma_{t+1}$  is computed which minimizes the updated loss function:

$$L(H_t, y) \rightarrow L(H_t + \gamma_{t+1} \cdot h_{t+1}, y) \quad (8.9)$$

With this, a new combined classifier is created:

$$H_{t+1} = H_t + \nu \cdot \gamma_{t+1} \cdot h_{t+1} \quad (8.10)$$

The factor  $\nu$  is called learning rate or shrinkage and is usually included to decrease the weight of the added tree in each iteration step in order to avoid a too rapid loss of generality of the combined classifier. The previous steps are iterated until a maximum number of decision trees is reached [38, 41, 42]. As it is the case with AdaBoost, GradientBoost also shows very good performance for weak learners. Additionally, a resampling procedure, called bagging, is performed, by randomly selecting subsamples of the training sample for the decision tree development.

### 8.3.2. General Settings for BDTs

In these studies, the BDT implementation provided by the TMVA package [38] is used.

All of the variables described in section 5.7 can be used as an input for BDT training. However, it has to be empirically tested, with which combination of variables the BDT performs best with respect to separation power and overtraining. Usually, correlations between the variables are taken into account as well. To prevent redundancies originating from these correlations in the BDT training, the variables can be decorrelated by e.g. the so-called Principal Component Analysis (PCA) [43] algorithm. In general, PCA performs a rotation of the coordinate system spanned by two (correlated) variables until the maximum variability is reached. PCA can also be performed for multiple variables simultaneously. In the optimization process of BDTs the most powerful combination turned out to be a simultaneous decorrelation of the invariant mass, the effective mass, the transverse momenta of the two leading  $\tau$ -leptons and the missing transverse energy. This can also be illustrated in the correlation plots before and after the transformation for each BDT that is presented.

Apart from variable correlations also the respective discrimination power can be useful for the choice of variables. As a starting point serves the so-called method independent ranking provided by TMVA. This ranking is computed by comparing the shapes of the signal and background distributions for each variable and calculating the overlap as a measure of discrimination power. A variable for which the signal and background distributions only have very small overlap will consequently be ranked higher than a variable with large overlap. After a BDT is trained, also a method dependent ranking is provided. Here, the rank is determined by the relative number of cuts that were set on a particular variable in the decision trees, which is in some sense a "generalized" form of the method independent procedure, as the cut variables are specified by the shape overlaps after reweighting the events. It has to be noted, that both rankings will not necessarily be identical or even similar as the boosting procedure changes the variable distributions. As the method dependent ranking in general varies strongly with every new BDT setup, only the method independent ranking will be shown.

The training and testing events are provided by the samples described in chapter 6. In this analysis, 50% of the events are taken for training and testing each, chosen at random. For training the GradientBoost algorithm is used, in combination with a bagging procedure, which selects 50% of the training events at random as a subsample.

The number of decision trees trained has been varied in these studies between 200 and 600, the maximum depth (i.e. the maximum number of cuts along one branch of a decision tree) of each tree is chosen between 2 and 4. The number of possible cuts on each variable is set to 20 in an equidistant manner. Another parameter for the BDT, the shrinkage,

controls the "learning rate" of the BDT and is set to 0.1. This rather low value for the shrinkage requires a higher number of trees to be trained but it can improve the accuracy especially in the presence of statistical fluctuations. The minimum node size specifies the minimum fraction of events that is allowed for a final node in a decision tree and is set to 2.5% of the number of training events.

All the parameters for BDT training with their respective values are listed in table 8.11.

As already described, event weights are modified in each iteration of the boosting procedure. In some cases this might cause problems if the original Monte Carlo weight features a negative sign which is preserved in the boosting. Therefore, in absence of a better solution, events with negative weights can be simply ignored in training but are nevertheless used in testing.

As a measure for the separation power of the BDT response, TMVA computes a so-called ROC (Receiver Operating Characteristic) curve. Usually, this curve gives the signal efficiency for a possible cut at each point of the BDT response with the corresponding background rejection (which is 1 minus the background efficiency). A perfectly discriminating BDT response would generate a rectangular ROC curve. As a quantitative measure for the discrimination power the area below the ROC curve is calculated.

Generally, the BDT is trained for a signal sample and all Standard Model backgrounds taken together, making it impossible to discriminate between the separate backgrounds afterwards. In order to obtain the BDT distribution for each considered process, the BDT has to be evaluated again for each background (and of course also signal) sample separately. This is done in a dedicated step of the BDT analysis. With the information of the BDT evaluation of each signal and background sample it is then possible to determine and optimize the significance of a cut on the BDT response.

In order to increase the statistics available for the BDT training, the preselection described in section 7.1 is loosened for this step to a minimum. At least two medium  $\tau$ -leptons are required, where the two leading ones should have opposite signs of their electric charge. Furthermore, the events have to be selected either by the asymmetric  $\tau$ -trigger or the ditau+MET-trigger. However, the corresponding offline cuts are not applied anymore in the training step in order to increase the number of training events, but are kept in the BDT evaluation. The  $b$ -jet and the  $Z$ -veto are not applied anymore, neither in the training nor in the evaluation. Instead, the number of  $b$ -jets and the invariant mass of the two  $\tau$ -leptons are added as input variables for BDT training. Another cut is performed due to technical reasons. The absolute value of the initial event weight is required to be no larger than 25, because such high event yields (positive or negative) lead to issues in the training process. This cut is of course not applied in the evaluation of the BDT, it just serves to ensure a correctly performed training. In the case that the QCD background is included in the evaluation of the BDT,  $m_{T2} > 20$  GeV is demanded as the estimate for the QCD multijet background is only defined above this value by construction (see section 6.2.5).

### 8.3.3. Overtraining

In TMVA, the testing of a BDT is performed directly after its training. The comparison of the final BDT response function for training and testing gives a direct hint at the presence of overtraining. Apart from the "optical" comparison also a Kolmogorov-Smirnov-test [44] is performed for signal and background separately, providing a probability for the



Parameter	Value or range
Boost Type	GradientBoost
$N(\text{trees})$	200-600
min(node size)	2.5%
Shrinkage	0.1
Ignore Negative weights	True
max(cuts)	20
max(depth)	2-4
Bagged Sample Fraction	50%

Table 8.11.: A summary of the chosen parameters for BDT training. The number of trees and the maximum depth of the trees is varied in order to optimize the discrimination of signal and background with as little overtraining as possible.

training and the testing distributions being the same. Hypothetically, a value of 1 would correspond to a perfect match between training and testing distributions. However, this kind of test is not totally reliable when comparing the shapes of binned plots as it originally was designed for continuous functions. It is therefore reasonable to accept probabilities larger than 0.1 and do an additional overtraining check. Typically, three points in the BDT response are chosen, such that a lower cut at these points would result in a background selection efficiency of 1%, 10% and 30%, respectively. For these cuts the efficiency of the signal selection is checked for the signal distribution both for training and testing. By convention, if they match within  $\pm 5$  percentage points the overtraining is assumed to be negligible. With this, there are three different ways for checking overtraining in TMVA, making sure that at least one of them yields reliable results.

The behaviour concerning overtraining can vary strongly between two BDTs, even though the settings might be similar. E.g. if the same parameters, input variables and background samples are used for two BDTs but the signal events in the training differ, this might not only have an impact on the overtraining behaviour of the signal. Changing a certain amount of the training events can result in an entirely different BDT, which also uses different regions of phase space for the discrimination of signal and background. Thus, training on a fluctuation in the background distribution might not happen in one case while it does in the other – causing overtraining for the background samples despite the same datasets have been used in both cases. A similar effect arises by adding or removing an input variable (regardless of its discrimination power, as this might change after a cut on a different variable has been performed). Therefore, the overtraining behaviour does not only vary with the number of decision trees trained or their maximum depth, but also depends strongly on the training events themselves and the distributions of the used variables.

#### 8.3.4. BDT trained on ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ , $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ )

As the signal mass point with 200 GeV stau mass and a neutralino mass of 1 GeV performed best in the cut-and-count analysis, the BDT output shown in figure 8.7 was trained also with this dataset. In total 234989 (unweighted) background events and 1574 signal events

are available, half of which is used for training. The input variables are listed and ranked in table 8.12. A principle component algorithm is applied to the invariant mass, the effective mass, the two leading transverse  $\tau$ -momenta and the missing transverse energy. Their correlations before and after the transformation can be observed in figures C.2 and C.3 for signal and figures C.4 and C.5 for background, respectively. Matrices showing the linear correlation factors between each pair of input variables are shown in figure C.1. The number of decision trees to be trained is set to 200 with a maximum depth of 2. Both of these parameters have been set to the lowest possible values in order to avoid overtraining. Nevertheless, strong overtraining is observed, which is mainly due to the limited amount of training events. This can also be observed in the overtraining check in table 8.13. As the number of decision trees and the maximum depth cannot be decreased anymore without a significant loss in the boosting performance [38], instead training statistics have to be increased.

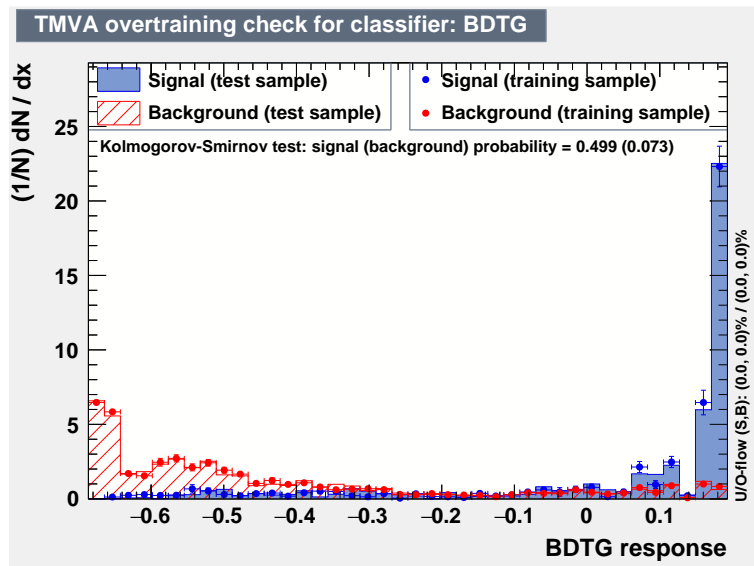


Figure 8.7.: Response distribution of a BDT trained on the signal mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV). The points denote the distribution for the training sample whereas the filled areas show the BDT response of the testing sample. The background samples are shown together in red, the signal samples used in training in blue. The Kolmogorov-Smirnov test yields an especially low probability for the background distribution. Additionally, the distribution does not allow for a good discrimination of signal and background as the signal events are all located at response values below 0.2 (perfect discrimination would correspond to a peak at +1 for signal and at  $-1$  for background).

### 8.3.5. BDT trained with High Statistics Sample

A dedicated signal sample for the mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) has been produced with increased statistics, i.e. with a larger number of simulated (unweighted) events. With this dataset the number of available events for training and testing is increased to 6454 unweighted events. Analogously to the previously presented attempt, the BDT training parameters are optimized, the best results concerning overtraining being ob-

Rank	Input variable
1	$m_{\text{T}}(\tau_1) + m_{\text{T}}(\tau_2)$
2	$m_{\text{T}}(\tau_1)$
3	$m_{\text{eff}}$
4	$m_{\text{T}2}(\tau_1, \tau_2)$
5	$p_{\text{T}}(\tau_1)$
6	$m_{\text{T}}(\tau_2)$
7	$m_{\text{inv}}(\tau_1, \tau_2)$
8	$E_{\text{T}}^{\text{miss}}$
9	$\Delta R(\tau_1, \tau_2)$
10	$\Delta\phi(\tau_1, \tau_2)$
11	$p_{\text{T}}(\tau_2)$
12	$\cos\theta_{ll}$
13	$\Delta\phi(\tau_2, E_{\text{T}}^{\text{miss}})$
14	$N(\text{b-jets})$

Table 8.12.: Input variables and their respective ranks before the training of the BDT on the signal samples with a stau mass of 200 GeV and a neutralino mass of 1 GeV. A PCA transformation is applied to the invariant mass, the effective mass, the two leading transverse  $\tau$ -momenta and the missing transverse energy.

tained with 300 decision trees with a maximum depth of 3. The other parameters as well as the PCA transformations are the same as before. The Kolmogorov-Smirnov probabilities (see figure 8.8) are above 0.1 for both signal and background. The signal efficiencies in the overtraining check in table 8.14 match within  $\pm 5$  percentage points, together with the Kolmogorov-Smirnov probabilities indicating the absence of overtraining. This already is a notable improvement compared to the BDT training in the previous section, as the larger number of training events not only reduces the proneness for overtraining but also allows for more and larger decision trees. However, it can be observed from figure 8.8 that the peak of the BDT response distribution for the signal is located at a BDT score of 0.6, which means that no optimal separation of signal and background has been achieved.

Background efficiency	Signal efficiency in training (testing)
1%	44.2% (35.6%)
10%	76.4% (77.1%)
30%	92.8% (93.6%)

Table 8.13.: Overtraining check for a BDT trained on the signal mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV). The largest discrepancy in the efficiencies of training and testing is observed for a rather tight cut at a background acceptance of 1%. There, the difference in the signal efficiency is almost 10 percentage points, which is a strong indication for overtraining.

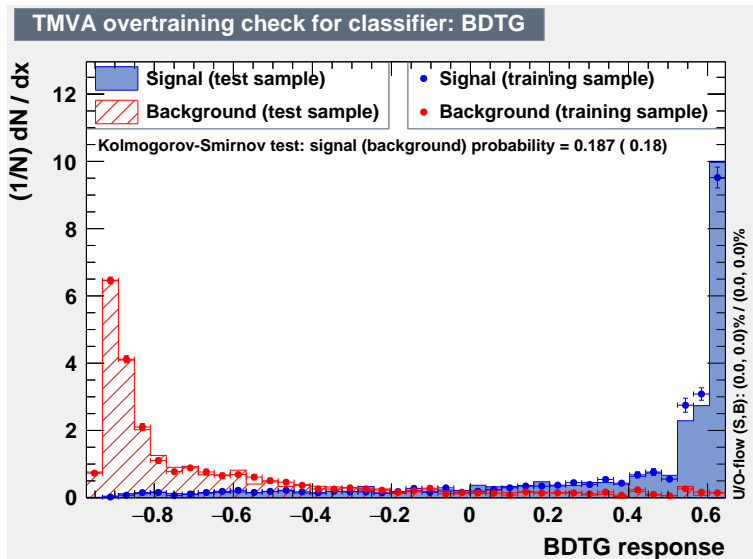


Figure 8.8.: Response of BDT trained on single high statistics mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV). The Kolmogorov-Smirnov probabilities are larger than 0.1 for both signal and background, making overtraining unlikely. Even though there can be observed a better separation of signal and background than in figure 8.7, the optimal BDT score of +1 is still not reached for the signal distribution, which has its peak at a value of 0.6.

Background efficiency	Signal efficiency in training (testing)
1%	52.8% (49.2%)
10%	86.1% (85.8%)
30%	96.5% (96.4%)

Table 8.14.: Overtraining check for a BDT trained on the signal mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) with a high statistics dataset. The signal efficiencies for training and testing do not differ by more than  $\pm 5$  percentage points, indicating that no overtraining has occurred.

The BDT has been evaluated for all background samples and for the signal mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV), which can be seen in the upper plot in figure 8.9. The middle plot shows the development of the significance with increasing lower cuts on the BDT response. The  $z$ -values for all signal mass points at an optimized cut on the BDT score are shown in figure 8.10. The requirements for the optimization were that each background sample must yield more than 0 events with a total background yield of at least 2 and a total maximum background uncertainty of 50%. The event yields for the optimized cut at a BDT score of 0.58 are listed in table 8.15. As the significance curve in figure 8.9 indicates, a tighter cut value on the BDT response than 0.58 would result in a  $z$ -value of approximately 1. However, this cut is not chosen in the optimization as the uncertainty of the  $W$ +jets background grows too high ( $\sim 130\%$ ), leading to an increased total background uncertainty of 54.44%.

Finally it can be stated, that in this case there were enough (unweighted) events available

to prevent overtraining, but the amount of training events was insufficient for achieving a powerful separation of signal and background. Therefore, an even larger amount of simulated signal events would be advantageous to obtain a BDT with a more powerful separation.

Process	Events (statistical uncertainty)
$W$ +jets	18.20 (41.87%)
$Z$ +jets	17.22 (69.79%)
Diboson	13.76 (10.88%)
Top	4.72 (26.10%)
QCD	5.55 (48.95%)
Total background	59.45 (24.58%)
Signal	16.82 (2.70%)

Table 8.15.: Event yields per background and for the ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) signal mass point for a lower cut at a BDT score of 0.58, scaled to an integrated luminosity of  $36.1 \text{ fb}^{-1}$ .

### 8.3.6. Parameterized BDTs

Previously, the BDT training was done for a single signal sample. However, as the number of raw signal events is very limited for just one signal sample, the BDT will always suffer from low signal statistics and consequently will be very prone to overtraining. A first attempt to cure this shortcoming is a method called "Parameterized Machine Learning" and was first presented by P. Baldi et al. [45]. Although the method was developed for neural networks, it can in principle also be applied to boosted decision trees. In the following the performance of this method for a BDT trained for the discrimination of a direct stau signal from background is studied, also with respect to the overtraining behaviour.

The idea of parameterized classifiers is based on the fact that many analyses use multiple signal samples with some parameter  $\theta_i$  which yields a different value for each of these samples. In this particular case this would be the masses of the stau and the neutralino. For simplicity only the stau mass shall be used in these studies. The parameterization is then achieved by employing more than one signal sample for training and at the same time including the parameter  $\theta_i$  as an input variable which is treated exactly equally to the event defining variables  $\vec{x}_i$ . However, as the parameter  $\theta_i$  cannot be associated to background events, to each background training sample a variable is added with the same name and the available values are distributed at random to the background events. It has to be noted, that with this no further discriminating variable should be constructed "artificially". To make sure that the discrimination between the datasets is not distorted, the assignment frequency of a particular stau mass to background events reflects the fraction of signal events with this stau mass. The goal of this method, as stated in [45], is to create a replacement for a classifier which is trained for many mass points separately by providing a smooth interpolation between these mass points. However, the method may also lead to less overtraining, which will be checked in the following.

In this attempt 400 decision trees are trained with a maximum depth of 3. The used signal samples have stau masses ranging from 180 GeV to 260 GeV. For a stau mass of 200 GeV the datasets with neutralino masses of 1 GeV, 20 GeV and 40 GeV are included as well as the datasets with a stau mass of 220 GeV and neutralino masses of 1 GeV and 20 GeV. For all other stau mass points only the respective sample with a neutralino mass of 1 GeV is used. The number of unweighted signal events for training and testing is increased to 13499. A PCA transformation is applied to the invariant mass, the effective mass, the transverse momenta of the two leading  $\tau$ -leptons and the missing transverse energy. The effect of including the stau mass as a training variable should be checked by comparing the resulting BDT distribution with a BDT obtained when only using multiple signal datasets without the stau mass as an input variable while keeping all other parameters the same. The input variables and their method independent ranks are listed in table 8.16.

In figure 8.11 the output distributions of a BDT trained with and without  $m_{\tilde{\tau}}$  as an input variable are compared. From the two BDT distributions it can be concluded that, although overtraining is negligible for both, the BDT response obtained without the stau mass variable seems less prone to it. As an additional check, the signal and background efficiencies for cuts at three points of the BDT response are compared (as described in section 8.3.2) in table 8.17.

Although the main goal of the direct stau search is to improve the significance for the considered signal mass points, at first the reliability of the classifier should be guaranteed as the analysis also suffers from low statistics. In the following, the training of a boosted decision tree will be performed with a multiplicity of signal samples to increase the available number of signal events. However, the stau mass will not be used as training variable as this (empirically) seems to amplify overtraining, while there is only a small gain in separation, as can also be seen from the selection efficiencies given in table 8.17.

Rank	Input variable
1	$m_{\Gamma}(\tau_1) + m_{\Gamma}(\tau_2)$
2	$m_{\Gamma}(\tau_1)$
3	$m_{\text{eff}}$
4	$p_{\Gamma}(\tau_1)$
5	$m_{\Gamma 2}(\tau_1, \tau_2)$
6	$m_{\text{inv}}(\tau_1, \tau_2)$
7	$m_{\Gamma}(\tau_2)$
8	$\Delta R(\tau_1, \tau_2)$
9	$E_{\text{T}}^{\text{miss}}$
10	$\Delta\phi(\tau_1, \tau_2)$
11	$p_{\Gamma}(\tau_2)$
12	$N(\text{b-jets})$
13	$\cos\theta_{ll}$
14	$\Delta\phi(\tau_2, E_{\text{T}}^{\text{miss}})$
15	$m_{\tilde{\tau}}$

Table 8.16.: Input variables and their respective method independent ranks for a parameterized BDT. A PCA transformation is applied to the invariant mass, the effective mass, the two leading transverse  $\tau$ -momenta and the missing transverse energy. The stau mass is listed as the least discriminating variable, as expected by construction. It can take values of 180 GeV, 200 GeV, 220 GeV, 240 GeV and 260 GeV.

Background efficiency:	1%	10%	30%
with $m_{\tilde{\tau}}$ in training	53.4% (49.0%)	84.1% (84.1%)	96.2% (95.2%)
without $m_{\tilde{\tau}}$ in training	51.1% (49.1%)	83.0% (83.1%)	95.4% (95.3%)

Table 8.17.: Signal efficiencies for the testing (training) samples when cutting at those values of the BDT response at which the background efficiency of the testing sample is 1%, 10% and 30%. At a high background acceptance most of the signal statistics is included which leads to high signal efficiencies and low fluctuations. However, at low background efficiencies the difference between the training and the testing sample is larger, especially in case the stau mass was one of the training variables.

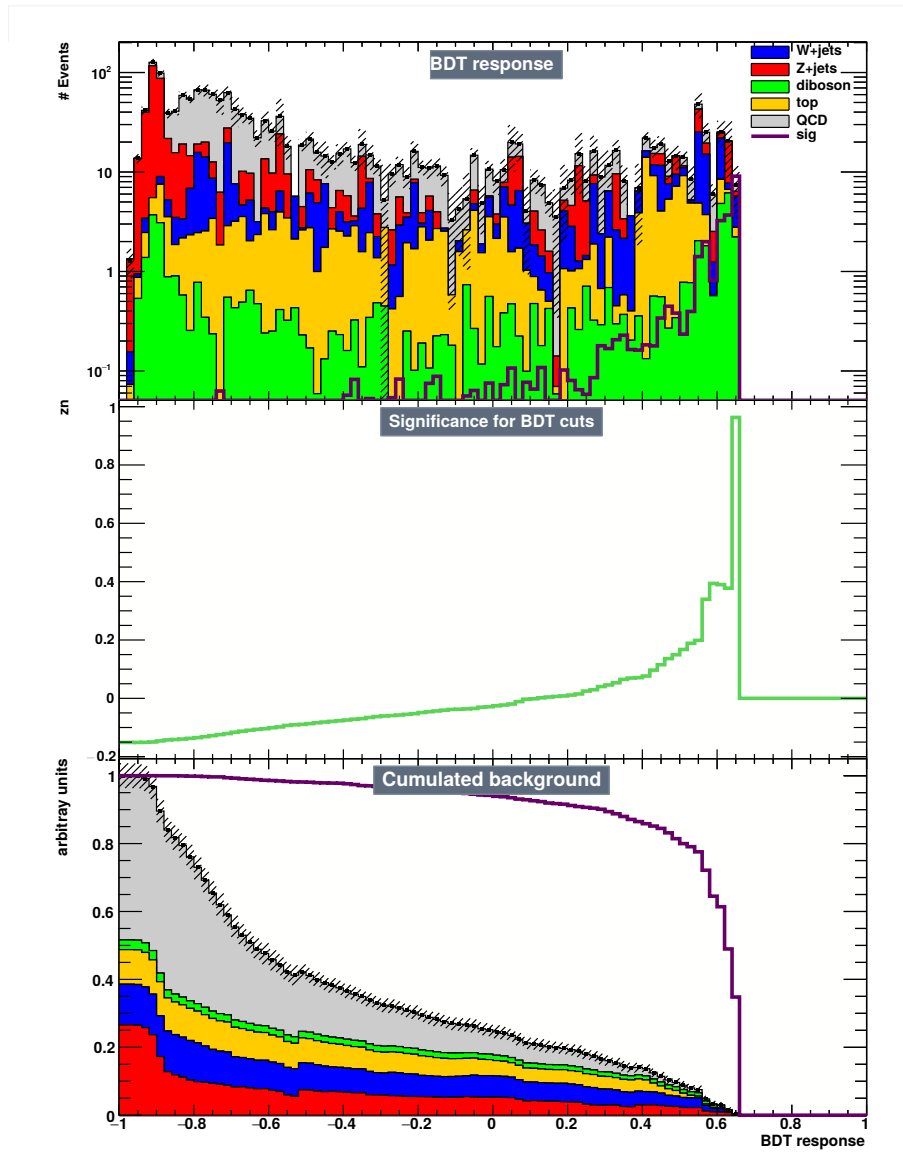


Figure 8.9.: Evaluation of a BDT trained with a high statistics signal sample of the mass point ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ). The upper plot shows the BDT response distributions for each background separately at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ :  $W$ +jets (blue),  $Z$ +jets (red), diboson (green), top (orange) and QCD (grey). The distribution for the signal mass point ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) is depicted by the violet line. In the middle plot the significance for this mass point is shown in dependence on a lower cut on the BDT response. The lower plot illustrates the cumulative behaviour of the backgrounds and the signal.



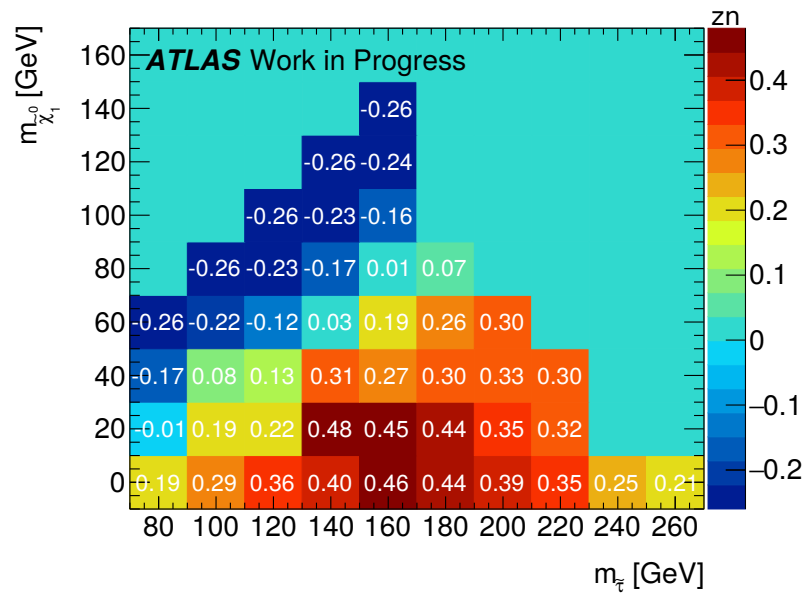
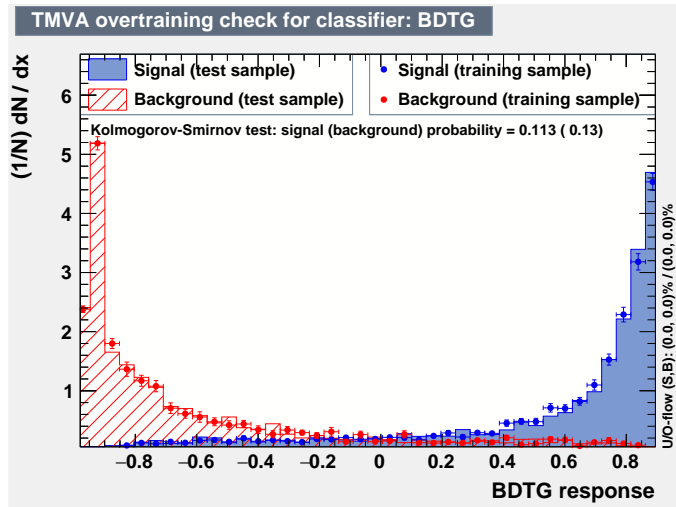
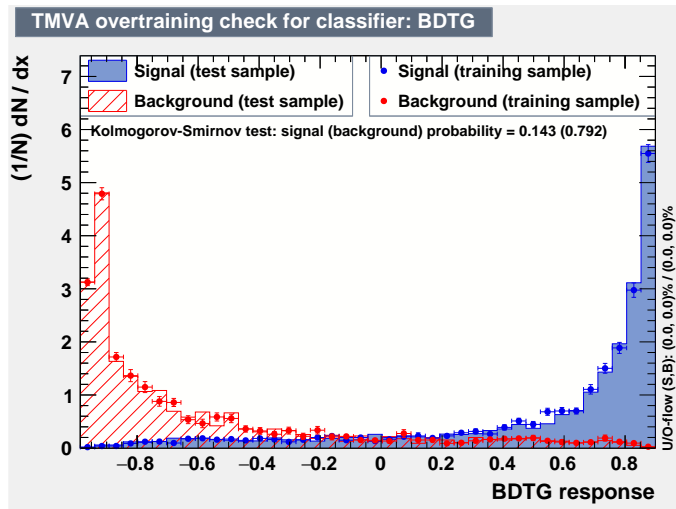


Figure 8.10.: Significances for a cut on the BDT response at 0.58 for all signal mass points, with a BDT trained on a high statistics dataset of the mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV). The largest  $z$ -values are about 0.45. A tighter cut on the BDT response, which would increase the  $z$ -values cannot be selected as the statistical uncertainties on  $W$ +jets and therefore also the total background yield would be notably increased. All significances are calculated for an integrated luminosity of  $36.1 \text{ fb}^{-1}$ .



(a)



(b)

Figure 8.11.: Response distribution for a BDT training including the stau mass (a) and without the stau mass (b). The points denote the distribution for the training sample whereas the filled areas show the BDT response of the testing sample. The background samples are shown together in red, the signal samples used in training in blue. It can be noted, that the distributions of (a) and (b) are quite similar in shape, with a slightly higher peak for the signal when excluding the stau mass from the input variables in training. Although the training and testing distributions match rather well in both cases, the values for the Kolmogorov-Smirnov probabilities are higher for (b), especially for the background samples. This can be explained by the fact that the inclusion (or exclusion) of any variable changes the chosen regions of phase space with which the BDT discriminates signal and background and thus can alter the overtraining behaviour. Consequently, the BDT without the stau mass variable seems to show less overtraining.

### 8.3.7. Results

With the knowledge from the previous section, BDTs for multiple signal datasets and all of the background samples listed in chapter 6 have been trained. The two best results for a "low stau mass"- and a "high stau mass"-BDT are presented in the following.

#### Low Stau Masses

The BDT with the largest significance for lower stau masses has been obtained by using four signal mass points for training:  $(m_{\tilde{\tau}} = 100 \text{ GeV}, m_{\tilde{\chi}_1^0} = 1 \text{ GeV})$ ,  $(m_{\tilde{\tau}} = 100 \text{ GeV}, m_{\tilde{\chi}_1^0} = 20 \text{ GeV})$ ,  $(m_{\tilde{\tau}} = 120 \text{ GeV}, m_{\tilde{\chi}_1^0} = 1 \text{ GeV})$  and  $(m_{\tilde{\tau}} = 140 \text{ GeV}, m_{\tilde{\chi}_1^0} = 1 \text{ GeV})$ . In total, these datasets sum up to 5798 unweighted events available for training and testing. The number of trees is set to 500, the maximum depth to 3.

As for the parameterized BDTs, a principal component transformation applied to the invariant mass, the effective mass, the two largest transverse  $\tau$ -momenta and the missing transverse energy turned out to be of advantage to increase the separation power of the BDT and also to prevent overtraining. The corresponding correlation plots for these variables are shown in figure C.7 for signal and in figure C.9 for background. Their behaviour after applying the PCA transformation to all of them simultaneously is shown in figures C.8 and C.10 for signal and background, respectively. It can be observed, that the PCA transformation helped minimizing their correlation notably. The linear correlation factors for each variable pair are depicted in figure C.6 for signal and background.

As can be seen in figure 8.13, there is no overtraining indicated by the Kolmogorov-Smirnov probabilities for signal and background which yield values of 0.165 and 0.159, respectively. This is confirmed by the overtraining check in table 8.19. The input variables are ranked in table 8.18. It can be noted that the  $m_{T2}$ -variable, which usually shows very good discrimination power in SUSY analyses, is ranked rather low. This is due to the low stau masses used in training, as the cut-off of  $m_{T2}$  is in the same range as for many background processes (see figure 8.12).

After making sure that the BDT is not overtrained, it is evaluated for each signal and background process separately. In figure 8.14 the BDT output distribution for each background and the signal mass point  $(m_{\tilde{\tau}} = 100 \text{ GeV}, m_{\tilde{\chi}_1^0} = 1 \text{ GeV})$  are plotted, as well as the significance for this mass point as a function of a lower cut on the BDT response. The cut optimization is performed with the same restrictions as described in section 8.2. The best result is obtained for the signal mass point  $(m_{\tilde{\tau}} = 100 \text{ GeV}, m_{\tilde{\chi}_1^0} = 1 \text{ GeV})$ , which yields a significance of 0.35 when cutting at a BDT score of 0.64. The significance for each signal point is shown in figure 8.15. The yields for each background and the best performing signal mass point are given in table 8.20.

It can be noted that the resulting z-values are smaller than those obtained for low stau masses in the cut-and-count approach. As stated in section 8.3.1, boosted classifiers consisting of weak learners, i.e. small decision trees, tend to show a better discrimination performance than the usual cut-and-count signal region. However, it is required that the weak learners at least perform better than random guessing, which might not be the case if too little statistics are available for training and the provided variables lack discrimination power, as it is the case e.g. for  $m_{T2}$ . Moreover, only half of the events can be used for training the BDT, whereas in the cut-and-count approach the total available statistics is

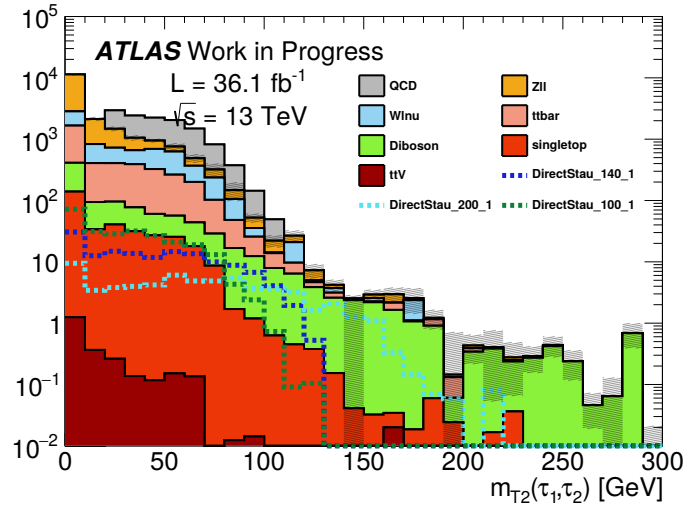


Figure 8.12.: Comparison of the  $m_{T2}$  distribution for the stau masses of 100 GeV (dark green), 140 GeV (dark blue) and 200 GeV (very light blue) and all background samples. For this plot at least two medium  $\tau$ -leptons with oppositely signed charges are required. The cut-off of  $m_{T2}$  increases with the stau mass, leading to a better discrimination power of the variable at high stau masses.

employed. Furthermore, in figure 8.14 it can be observed, that the maximum  $z$ -value is in fact about 1.1 which is in the same range as it was for the "SR DS HadHad Low Mass" region. The reason that the optimal cut was found to be at a lower significance than this is due to the additional requirements on the number of events in this region.

### High Stau Masses

For high stau masses, the best significances are obtained with a boosted decision tree which is trained on the signal mass points ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV), ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 20$  GeV), ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 40$  GeV), ( $m_{\tilde{\tau}} = 220$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV), ( $m_{\tilde{\tau}} = 220$  GeV,  $m_{\tilde{\chi}_1^0} = 20$  GeV) and ( $m_{\tilde{\tau}} = 240$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV), resulting in 10099 unweighted signal events for training and testing the BDT. 500 decision trees are created with a maximum depth of 3.

Again, the principal component analysis is used for the decorrelation of the invariant mass, the effective mass, the two leading transverse  $\tau$ -momenta and the missing transverse energy. Plots illustrating the correlation between the effective mass and the other four variables before the transformation are given in figure C.12 for signal and in figure C.14 for background. The decorrelation due to the application of PCA can be observed in figures C.13 and C.15 for signal and background, respectively.

The BDT output score is shown in figure 8.16. The Kolmogorov-Smirnov test yields a probability of 0.382 for signal and 0.781 for background and therefore indicates that no overtraining has occurred, which can also be observed from the overtraining check in table 8.22. The input variables with their corresponding ranks are listed in table 8.21. The linear correlation factors between each pair of variables is given for signal and background

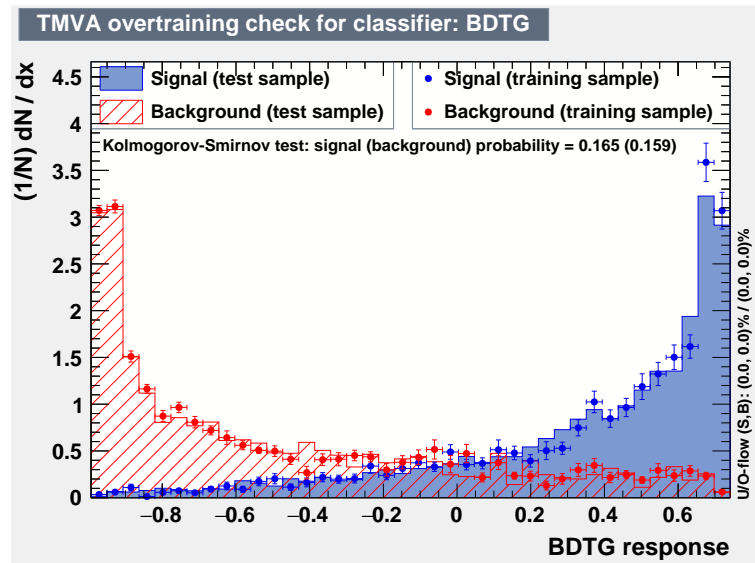


Figure 8.13.: BDT response for a training on signal samples with low stau masses. The background samples are summarized in red, the signal in blue. The training sample is represented by dots, the testing sample by the colored areas. In most bins, both distributions match very well, which together with the Kolmogorov-Smirnov probabilities and the overtraining check in table 8.19 is an indication that no overtraining has occurred.

in figure C.11.

The result of the evaluation for each background process together with the signal dataset for the mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) is plotted in figure 8.17. The significances for each signal mass point are shown in figure 8.18. The highest significance is obtained for a stau mass of 200 GeV and a neutralino mass of 1 GeV, with a z-value of 2.16 at a cut on the BDT response at 0.9. However, significances larger than the critical value of 1.64 are also reached for three other mass points, namely for ( $m_{\tilde{\tau}} = 220$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV), ( $m_{\tilde{\tau}} = 160$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) and ( $m_{\tilde{\tau}} = 160$  GeV,  $m_{\tilde{\chi}_1^0} = 20$  GeV). In table 8.23 the number of expected events for each background process and the signal mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) for an optimized cut on the BDT score at 0.9 are listed. The statistical uncertainty of the total background yield is larger than 50% which can be explained mainly by the 129.44% statistical uncertainty of the dominant  $W$ +jets background. Although there are four mass points with sufficient significance for exclusion, the only obstacle is given by the large statistical uncertainties which originate from the limited statistics, especially for  $W$ +jets. This means, that the number of available un-weighted  $W$ +jets events has to be increased in order to lower the statistical uncertainty without changing the behaviour of the weighted number of events.

Rank	Variable
1	$m_T(\tau_1) + m_T(\tau_2)$
2	$m_T(\tau_2)$
3	$\Delta\phi(\tau_1, \tau_2)$
4	$\Delta R(\tau_1, \tau_2)$
5	$m_{\text{inv}}(\tau_1, \tau_2)$
6	$m_{\text{eff}}$
7	$p_T(\tau_1)$
8	$N(\text{b-jets})$
9	$\cos\theta_{ll}$
10	$p_T(\tau_2)$
11	$\Delta\phi(\tau_2, E_T^{\text{miss}})$
12	$m_{T2}(\tau_1, \tau_2)$
13	$E_T^{\text{miss}}$
14	$m_T(\tau_2)$

Table 8.18.: Ranking of the input variables for a BDT trained on low stau mass samples. The  $m_{T2}$ -variable is not ranked as high as in the previous examples which is due to the low cut-off at small stau masses, making its distribution less distinguishable from the background distributions.

Background efficiency	Signal efficiency in training (testing)
1%	22.8% (22.5%)
10%	69.4% (66.6%)
30%	92.0% (91.3%)

Table 8.19.: Overtraining check for a BDT training on low stau mass samples. The signal efficiencies agree for the training and testing distributions within  $\pm 5$  percentage points.

Process	Events (statistical uncertainty)
$W$ +jets	35.35 (28.70%)
$Z$ +jets	21.16 (53.93%)
Diboson	4.60 (21.81%)
Top	4.91 (24.93%)
QCD	9.79 (32.66%)
Total background	75.82 (20.68%)
Signal	27.47 (8.44%)

Table 8.20.: Event yields per background and for the ( $m_{\tilde{\tau}} = 100 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) signal mass point for a cut on the low mass BDT score at 0.64, scaled to an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The total number of background events is about a factor 4 larger than the expected number of signal events, causing a low significance in this region of phase space.

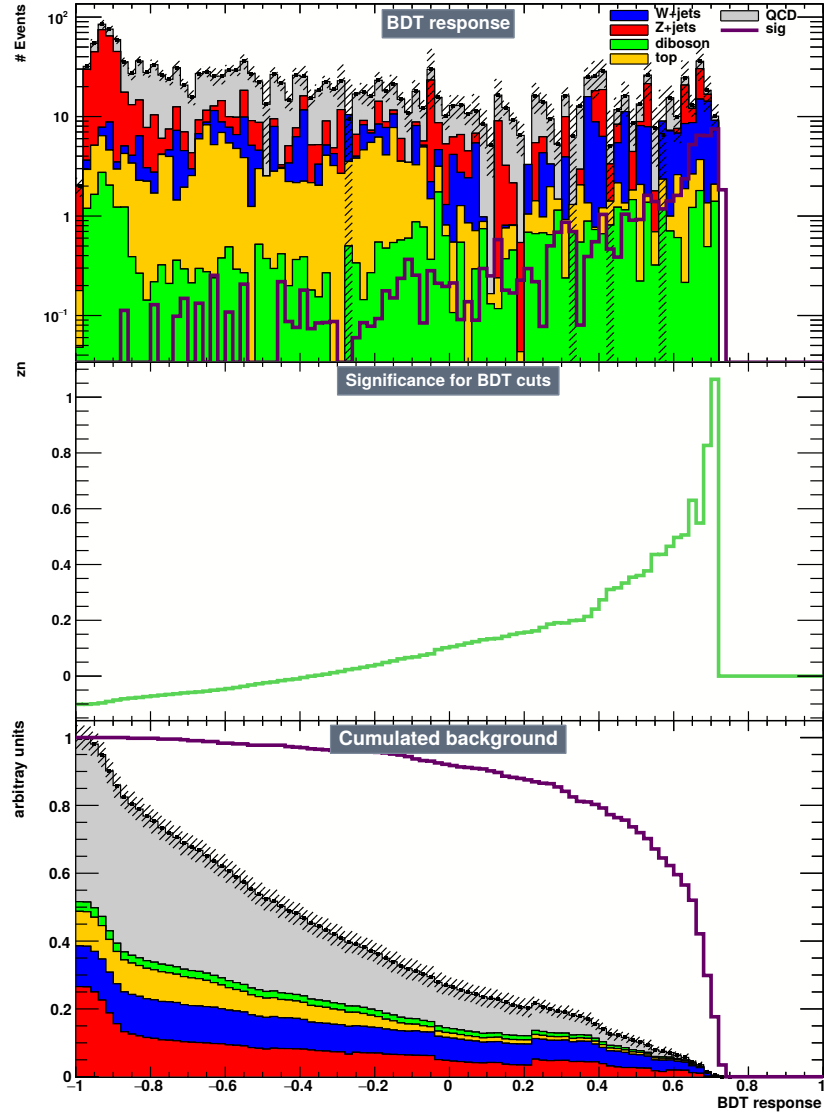


Figure 8.14.: The upper plot shows the distribution of the BDT response for  $W$ +jets (blue),  $Z$ +jets (red), diboson (green), top (orange), QCD (grey) and the signal mass point ( $m_{\tilde{\tau}} = 100 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) (violet) at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The middle plot shows the significance in dependence on a lower cut on the BDT response. The maximum significance of 0.35, at which the requirements described in section 8.2 are fulfilled, is reached at a BDT score of 0.64. The lower plot shows the cumulative distribution for signal and all backgrounds.

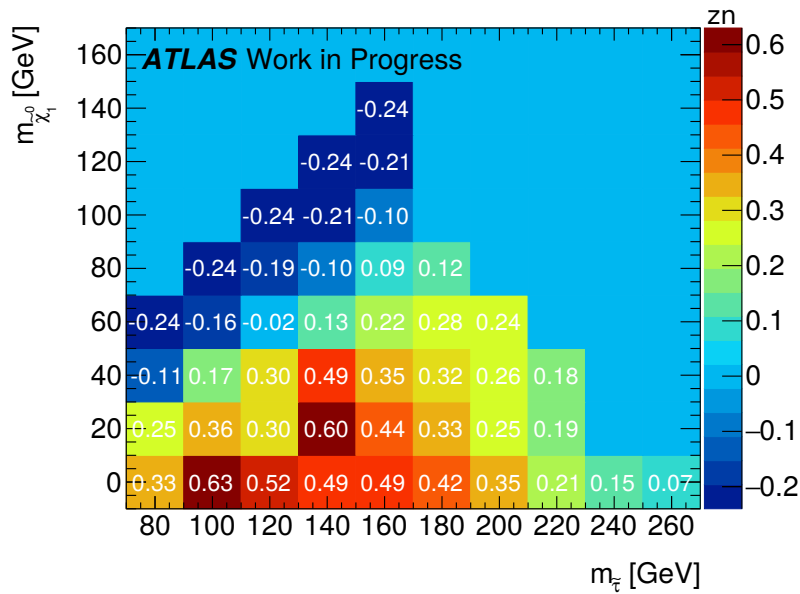


Figure 8.15.: Significance per signal mass point in the plane of the stau and the neutralino mass for an optimized cut on the low-mass-BDT response at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The largest  $z$ -values are reached for a stau mass of 100 GeV (140 GeV) and a neutralino mass of 1 GeV (20 GeV).

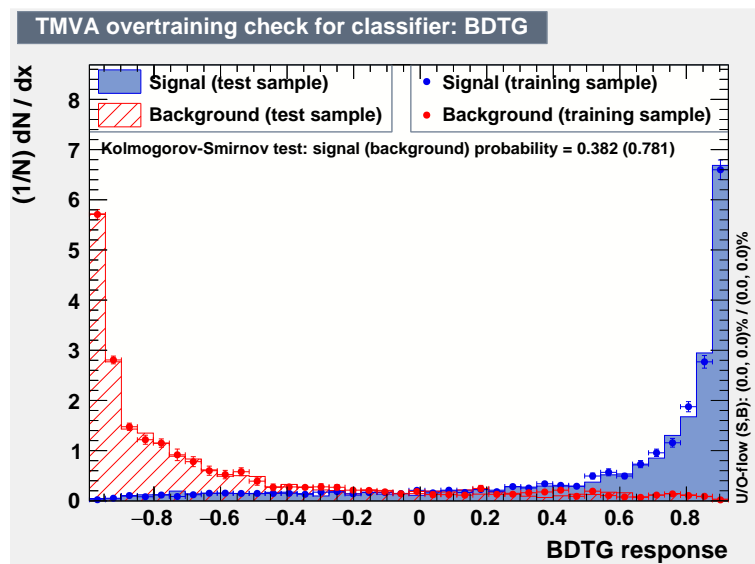


Figure 8.16.: BDT response distributions for signal (blue) and background (red), where the training was performed for high stau masses. The dots corresponding to the training sample and the filled areas for the test sample match sufficiently well, which together with the Kolmogorov-Smirnov probabilities indicates that no overtraining has occurred.



Rank	Variable
1	$m_{\text{T}}(\tau_1) + m_{\text{T}}(\tau_2)$
2	$m_{\text{T}}(\tau_1)$
3	$m_{\text{eff}}$
4	$p_{\text{T}}(\tau_1)$
5	$m_{\text{T}2}(\tau_1, \tau_2)$
6	$m_{\text{T}}(\tau_2)$
7	$m_{\text{inv}}(\tau_1, \tau_2)$
8	$E_{\text{T}}^{\text{miss}}$
9	$\Delta R(\tau_1, \tau_2)$
10	$p_{\text{T}}(\tau_2)$
11	$\Delta\phi(\tau_1, \tau_2)$
12	$N(\text{b-jets})$
13	$\cos\theta_{ll}$
14	$\Delta\phi(\tau_2, E_{\text{T}}^{\text{miss}})$

Table 8.21.: Ranking of the input variables used for a BDT training with high stau mass signal datasets.

Background efficiency	Signal efficiency in training (testing)
1%	52.9% (52.5%)
10%	84.8% (84.3%)
30%	95.8% (95.7%)

Table 8.22.: Overtraining check for a high stau mass BDT. The signal efficiency is checked for cuts on the BDT score which lead to a background acceptance of 1%, 10% and 30%, respectively. The signal efficiencies for training match within  $\pm 5$  percentage points to the testing sample's signal efficiency, given in brackets.

Process	Events (statistical uncertainty)
$W$ +jets	1.05 (129.44%)
$Z$ +jets	0.84 (51.10%)
Diboson	0.80 (81.56%)
Top	0.19 (45.67%)
QCD	0.46 (125.03%)
Total Background	3.34 (50.14%)
Signal	10.14 (8.23%)

Table 8.23.: Event yields for each background process and the signal mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV), scaled to an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . Although the expected number of events for the signal is notably larger than for the total background, the proposed signal region suffers from large statistical uncertainties, due to very limited statistics, especially for  $W$ +jets.

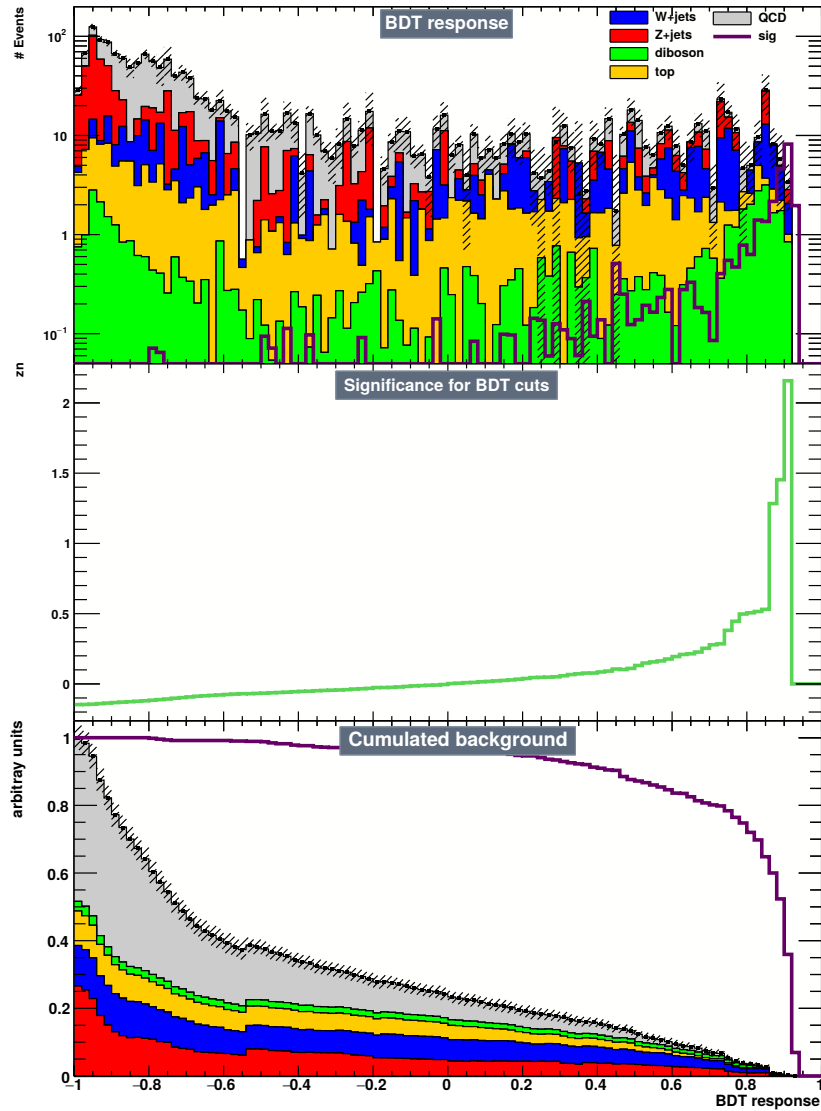


Figure 8.17.: The BDT response evaluated for each background sample and the signal mass point ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) at an integrated luminosity of  $36.1 \text{ fb}^{-1}$  is shown in the upper plot. The green line in the middle plot depicts the development of the significance in dependence on a lower cut on the BDT response. In the lower plot the cumulative background and signal distributions are illustrated.

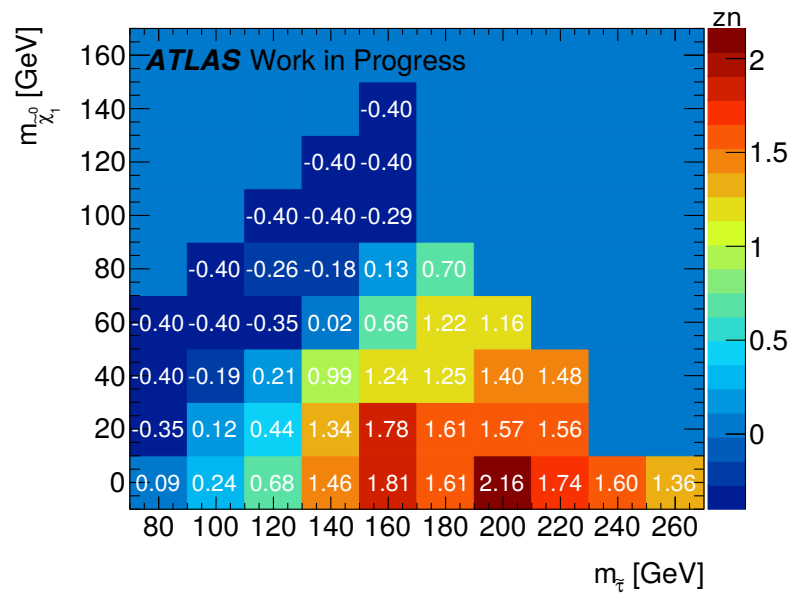


Figure 8.18.: Significances for each considered signal mass point in the stau and neutralino mass plane at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . In total, four mass points show z-values exceeding the critical value of 1.64 when applying a cut on the BDT score at 0.9.



## 9. Improvement of the $W$ +jets Statistics

As it has been observed in the previous chapter, the sensitivity for the HadHad-channel of the direct stau production process is mainly limited due to lacking statistics in some of the background simulations, especially for the  $W$ +jets samples. This can be explained by the fact that the  $W$ +jets background consists in its major part of processes in which the  $W$ -boson decays into a hadronic  $\tau$ -lepton whereas a second  $\tau$ -lepton is faked by a jet. As the fake rate is not very high, the amount of simulated events is largely decreased already by requiring two  $\tau$ -leptons at a medium identification working point. This might be an advantage in general, but such a limited amount of statistics also results in an increase of the statistical uncertainty on the  $W$ +jets background and consequently also of the total background uncertainty, making a meaningful calculation of the significance in a potential signal region very difficult. Thus, a way has to be found to raise the amount of usable simulated events. For this purpose the container  $\tau$ -promotion method has been recently developed at the LMU Munich by Alexander Mann, Adam Samara, Pascal Schmolz and Marco Dembecki.

### 9.1. Container Tau Promotion Method

As explained in section 5.4, the reconstructed  $\tau$ -leptons are identified at three working points (loose, medium and tight), each with different selection efficiencies and fake rates. The pool the  $\tau$ -candidates are selected from is called container  $\tau$ -leptons. In every simulated and selected event for  $W \rightarrow \tau\nu$  the faked  $\tau$ -leptons are also reconstructed as medium  $\tau$ -leptons. The number of simulated events would be much increased, however, if also fake  $\tau$ -leptons of no specific working point, i.e. container  $\tau$ -leptons, were allowed. It is reasonable to use container  $\tau$ -leptons instead of just adding loose  $\tau$ -leptons as the gain of events is larger in the former procedure and because loose  $\tau$ -leptons are already used in QCD estimation, possibly leading to redundancies. First, the possibly faked container  $\tau$ -leptons have to be identified, which is done via truth matching. For this a so-called truth-sample is used that simulates the same process as the  $W \rightarrow \tau\nu$ -datasets listed in table A.4, the only difference being that the information on the particle decay chains and their true identities is still available in those samples. Each of the container  $\tau$ -leptons is then matched to a true  $\tau$ -lepton in the truth sample. The former is called "real"  $\tau$ -lepton if it corresponds to a true  $\tau$ -lepton within  $\Delta R = 0.4$ . It is called "fake"  $\tau$ -lepton if no such match can be found. In each event with only one real  $\tau$ -lepton which lacks a medium fake  $\tau$ -lepton, one randomly chosen container fake  $\tau$ -lepton is selected and artificially "promoted" to medium ID. Afterwards the modified total number of events is reweighted to match the initial number of weighted events. This procedure is schematically depicted in figure 9.1. The reweighting factor is calculated via the fake efficiency  $\epsilon$ , i.e. the fake rate. As the fake efficiency shows a (small) dependence on the transverse  $\tau$ -momentum and the missing transverse energy, it is measured in distinct bins in the  $p_T(\tau)$ - $E_T^{\text{miss}}$ -plane from Monte Carlo simulation. Also, the measurement of the fake efficiency is done sepa-

rately for 1-prong and 3-prong  $\tau$ -leptons. The samples used for this method are listed in table A.4. Assume there are  $n$  non-real container  $\tau$ -leptons of which  $k$  pass the medium ID requirements. After promoting randomly one of these container  $\tau$ -leptons to medium quality, this number is changed to  $k + 1$ . The probability for  $k$  fake medium and  $n - k$  fake container  $\tau$ -leptons is given by the binomial probability distribution:

$$P(k) = \epsilon^k (1 - \epsilon)^{n-k} \binom{n}{k} \quad (9.1)$$

With one additional faked medium  $\tau$ -lepton this changes to

$$P(k + 1) = \epsilon^{k+1} (1 - \epsilon)^{n-k-1} \binom{n}{k + 1} \quad (9.2)$$

As the total sum of weights must not be affected, a reweighting factor  $w$  is introduced to obtain the probability of equation 9.2, after promoting one non-medium fake container  $\tau$ -lepton from a configuration with the probability of equation 9.1:

$$P(k + 1) \stackrel{!}{=} P(k) \cdot w \quad (9.3)$$

$$\Rightarrow w = \frac{P(k + 1)}{P(k)} \quad (9.4)$$

$$= \frac{\epsilon}{1 - \epsilon} \cdot \frac{n - k}{k + 1}, \quad (9.5)$$

which can be interpreted as

$$w = \frac{\epsilon}{1 - \epsilon} \cdot \frac{\text{number of non-medium fake } \tau\text{-leptons before promotion}}{\text{number of medium fake } \tau\text{-leptons after promotion}}. \quad (9.6)$$

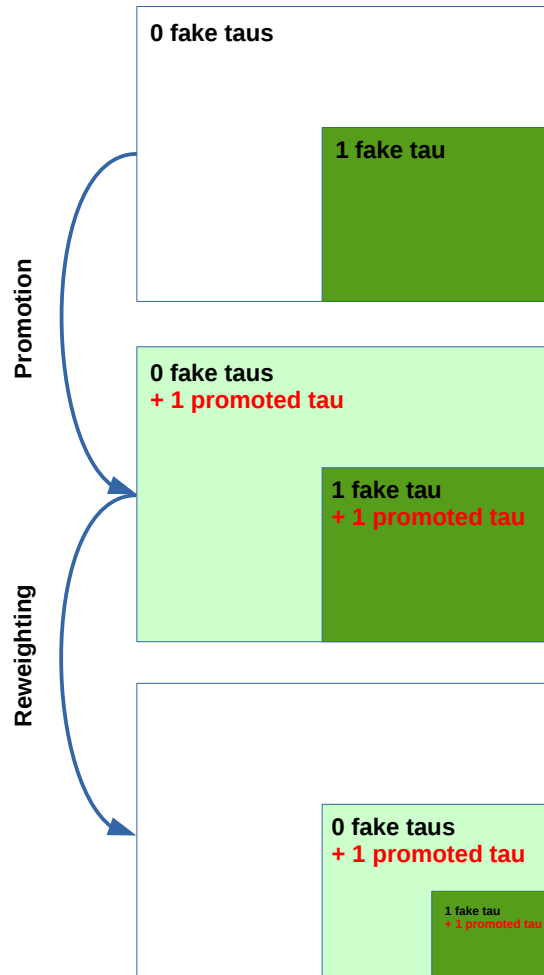


Figure 9.1.: Schematic view of the  $\tau$ -promotion method. The dark green area represents the amount of events with already one fake  $\tau$ -lepton passing medium requirements and therefore the initially available number of  $W \rightarrow \tau\nu$ -events. The lighter green area also is selected after promoting an arbitrary container  $\tau$ -lepton. In the final step the whole green area is reweighted such that the weighted number of events does not change (or changes only slightly within the uncertainties). At the same time, the unweighted number of events is increased which reduces the statistical uncertainty.

## 9.2. Performance of the Tau Promotion Method

In the following the influence of the  $\tau$ -promotion method (TP) is shown on the statistics of the  $W \rightarrow \tau\nu$ -events. First, the promotion of container  $\tau$ -leptons to medium classification is shown, as it is needed for the  $W$ +jets sample. Furthermore, the performance of a promotion to the loose identification working point will be investigated with respect to the estimation of the QCD multijet background. It has to be noted that, differently to the nominal samples, the trigger decisions cannot be applied to the  $\tau$ -promoted samples. This is due to the fact that the events that are gained by  $\tau$ -promotion include one or more (fake)  $\tau$ -lepton that was chosen out of the amount of  $\tau$ -leptons that did not pass the  $\tau$ -identification requirements and did not fire a trigger.

### 9.2.1. $W$ +jets Sample

The container  $\tau$ -promotion is applied to the samples listed in table A.4. Afterwards the weighted and unweighted number of events is compared before and after  $\tau$ -promotion for each sample as shown in table 9.1. The comparison of the distributions for the transverse momenta of the two leading  $\tau$ -leptons, the missing transverse energy, the  $m_{T2}$ -variable, the sum of the transverse masses of the two leading  $\tau$ -leptons and their distance in the  $\eta$ - $\phi$ -plane for the combined  $W$ +jets samples is depicted in figure 9.2. Although the reweighting is working within an acceptable range ( $\sim 5\%$ ) around the nominal event yields, it can be observed that the number of weighted events is systematically underestimated compared to the nominal ones in almost every used sample. As the reweighting factors are proportional to  $\epsilon/(1 - \epsilon)$ , an underestimation of the weights corresponds to an underestimation of the fake efficiency  $\epsilon$ . This can be explained by the fact that the bins in the  $p_T(\tau)$ - $E_T^{\text{miss}}$ -plane in which the fake efficiencies are measured cannot be chosen arbitrarily small. The binning concerning the  $p_T(\tau)$ -axis has originally been chosen as given in figure 9.3. If the bin width is reduced to resemble that in figure 9.4, the number of reweighted events is increased, which can be observed in table 9.2 and is also illustrated in figure 9.5. Despite the enhanced number of reweighted events after the promotion procedure, the statistical uncertainty is increased, limiting potential improvements by a further reduction of the bin widths.

### 9.2.2. QCD Sample

The promotion of container  $\tau$ -leptons can in principle also be performed to loose quality, as it is required for the QCD estimation (see section 6.2.5). Similarly to the previous chapter, the method is validated for each  $W \rightarrow \tau\nu$ -sample separately as shown in table 9.3 and in a combined manner for the final QCD sample as can be seen in figure 9.6. Although the same underestimation is observed as in the medium promotion, the effect on the total QCD estimation is very small, leading to almost identical distributions. The number of events in the dedicated control, validation and signal regions for the QCD estimation procedure as described in section 6.2.5 is shown in table 9.4. In some of the control and validation regions notable discrepancies from the nominal event yields are observed, although the resulting QCD estimation only differs by about 11%. It remains to be shown whether there is an improvement to be gained from  $\tau$ -promotion for the QCD estimation procedure, as only the number of unweighted events with two loose  $\tau$ -leptons



---

is increased in this case, but the amount of events with two medium  $\tau$ -leptons stays the same.

(a) Number of weighted events

Sample ID	$p_T(W)$ -slice [GeV]	filter	before TP ( $\sigma_{\text{stat}}$ )	after TP ( $\sigma_{\text{stat}}$ )	$\frac{\text{after TP}}{\text{before TP}}$
364184	0-70	light-jet	106306 (3.44%)	102287 (0.75%)	0.96
364185	0-70	c-jet	31188.0 (4.29%)	28180.0 (1.09%)	0.90
364186	0-70	b-jet	7321.92 (5.67%)	7114.43 (0.89%)	0.97
364187	70-140	light-jet	24398.1 (3.52%)	24403.8 (0.78%)	1.00
364188	70-140	c-jet	11482.7 (2.56%)	10372.7 (0.76%)	0.90
364189	70-140	b-jet	4099.02 (3.07%)	3771.02 (0.77%)	0.99
364190	140-280	light-jet	7984.53 (1.96%)	7937.88 (0.62%)	0.96
364191	140-280	c-jet	5709.08 (1.66%)	5056.30 (0.41%)	0.89
364192	140-280	b-jet	1807.57 (1.02%)	1631.78 (0.25%)	0.90
364193	280-500	light-jet	1302.03 (2.13%)	1291.90 (0.50%)	0.99
364194	280-500	c-jet	1088.39 (2.18%)	994.595 (0.50%)	0.91
361495	280-500	b-jet	403.869 (3.60%)	370.429 (0.54%)	0.92
364196	500-1000	-	511.796 (4.25%)	508.097 (0.74%)	0.99
364197	> 1000	-	39.1568 (2.39%)	36.7391 (0.45%)	0.94
<b>Total</b>			<b>203642 (1.98%)</b>	<b>193957 (0.46%)</b>	0.95

(b) Number of unweighted events

Sample ID	$p_T(W)$ -slice [GeV]	filter	before TP	after TP	$\frac{\text{after TP}}{\text{before TP}}$
364184	0-70	light-jet	7227	234825	32.49
364185	0-70	c-jet	4917	140480	28.57
364186	0-70	b-jet	6096	196382	32.21
364187	70-140	light-jet	15733	558803	35.52
364188	70-140	c-jet	13853	423309	30.56
364189	70-140	b-jet	10649	354158	33.26
364190	140-280	light-jet	10108	458777	45.39
364191	140-280	c-jet	10505	395712	37.67
364192	140-280	b-jet	28439	1129577	39.72
364193	280-500	light-jet	4733	257124	54.33
364194	280-500	c-jet	3752	176074	46.93
361495	280-500	b-jet	3386	155971	46.06
364196	500-1000	-	5793	325555	56.20
364197	> 1000	-	3457	208356	60.27
<b>Total</b>			<b>128648</b>	<b>5015103</b>	38.98

Table 9.1.: Number of weighted events (a) and unweighted events (b) before and after promoting one container  $\tau$ -lepton to medium quality for each  $W \rightarrow \tau\nu$ -sample at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . To obtain these numbers at least two medium  $\tau$ -leptons are required. It can already be observed from the weighted events that the statistical uncertainty  $\sigma_{\text{stat}}$  is reduced notably. This is due to the large increase of selected unweighted events after applying the  $\tau$ -promotion method. However, the number of weighted events is not exactly equal to the nominal yields. In fact, a slight underestimation can be observed, which is caused by the choice of the size of the bins in which the fake efficiencies are measured.

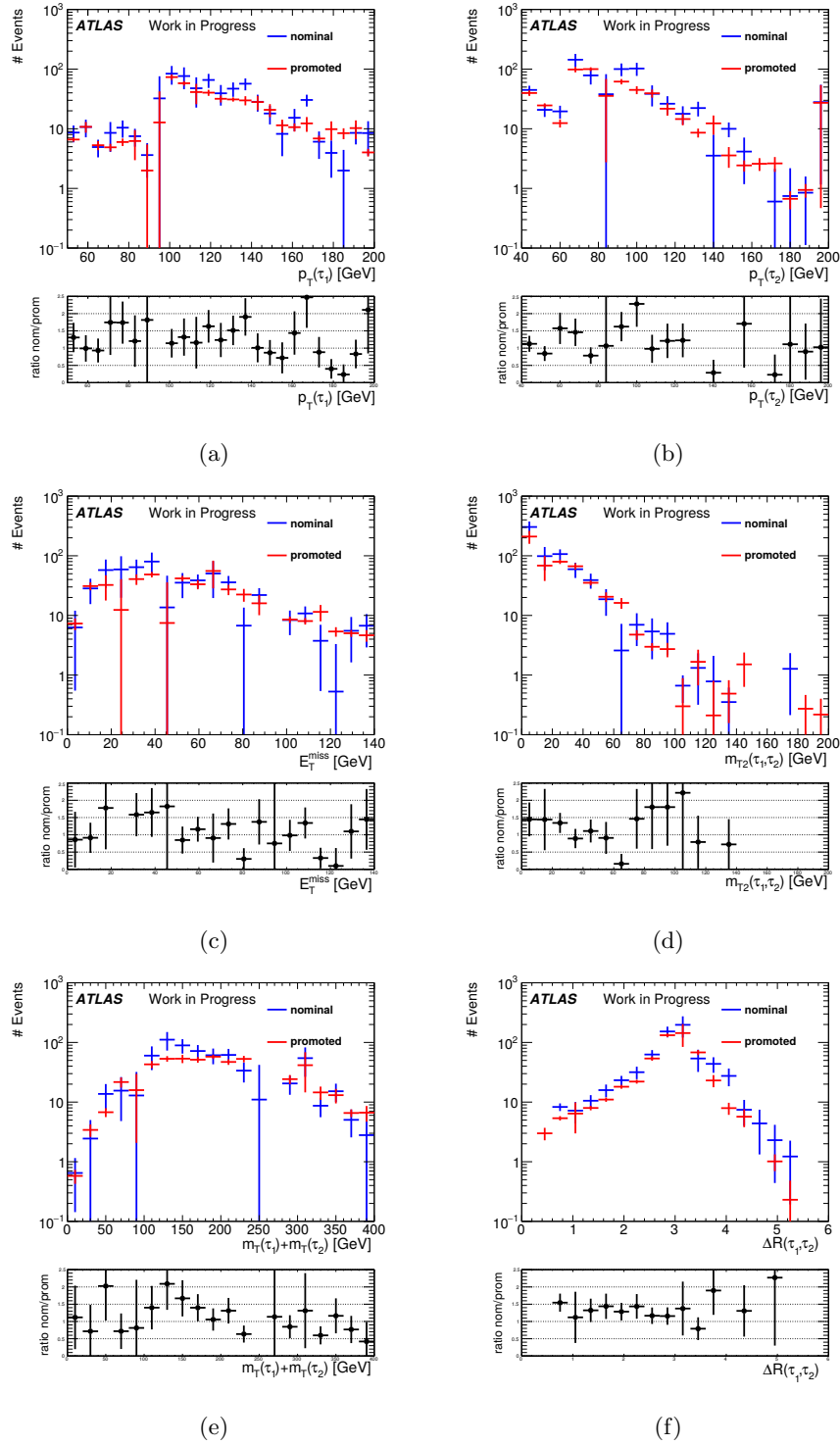


Figure 9.2.: Comparison of the nominal (blue) and the promoted  $W$ +jets distributions (red) for  $p_T(\tau_1)$  (a),  $p_T(\tau_1)$  (b),  $E_T^{\text{miss}}$  (c),  $m_{T2}(\tau_1, \tau_1)$  (d),  $m_T(\tau_1) + m_T(\tau_2)$  (e) and  $\Delta R(\tau_1, \tau_2)$  (f) for an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . In general the error bars for the promoted distributions are smaller apart from a few bins in which some negatively weighted events dominate. As can be observed from the ratio plots, the ratio of the nominal and promoted event yields for each bin fluctuate mostly around 1 with a visible tendency of underestimation in the promoted samples. For these plots at least two oppositely charged medium  $\tau$ -leptons have been required as well as the offline thresholds for the asymmetric  $\tau$ - and the ditau+MET-trigger as described in section 7.1.

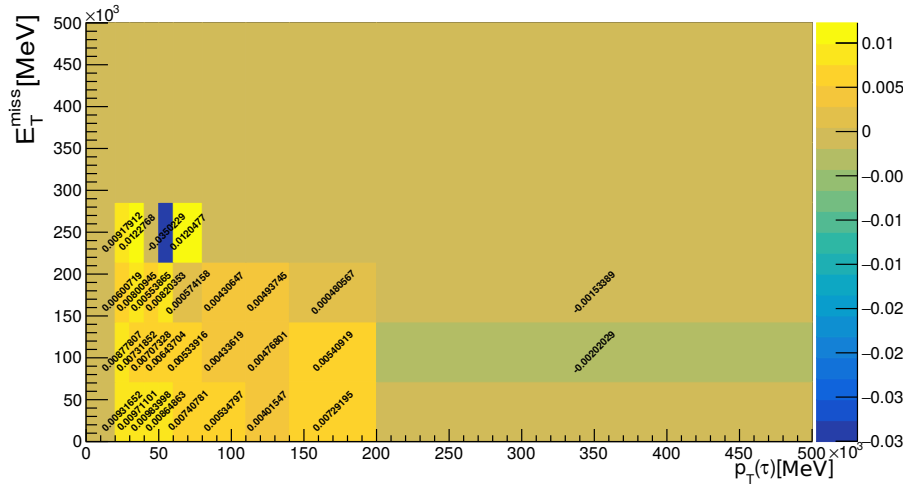


Figure 9.3.: Original binning for the fake efficiency measurement in the  $p_T(\tau)$ - $E_T^{\text{miss}}$ -plane. This example is taken from the measurement for 3-prong  $\tau$ -leptons for the  $p_T(W)$ -slice 140-280 GeV with a b-jet filter (Sample ID 364192).

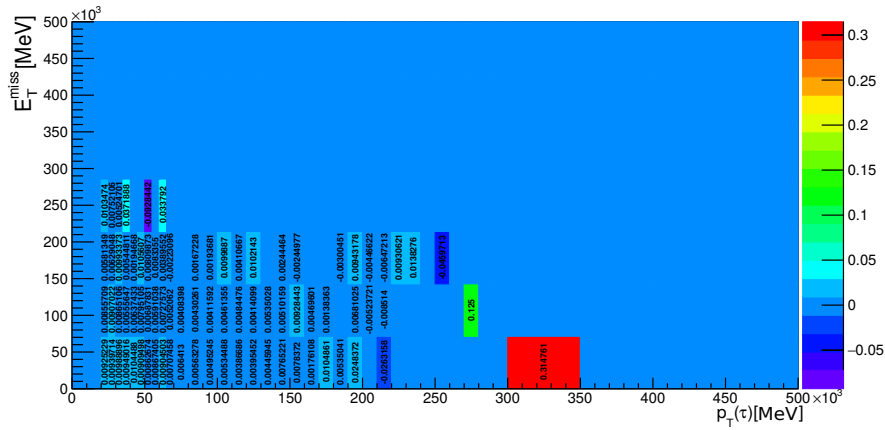


Figure 9.4.: Modified binning for the fake efficiency measurement in the  $p_T(\tau)$ - $E_T^{\text{miss}}$ -plane. This example is taken from the measurement for 3-prong  $\tau$ -leptons for the  $p_T(W)$ -slice 140-280 GeV with a b-jet filter (Sample ID 364192). Compared to figure 9.3, the  $p_T$ -bins are chosen smaller, especially at low transverse momenta.

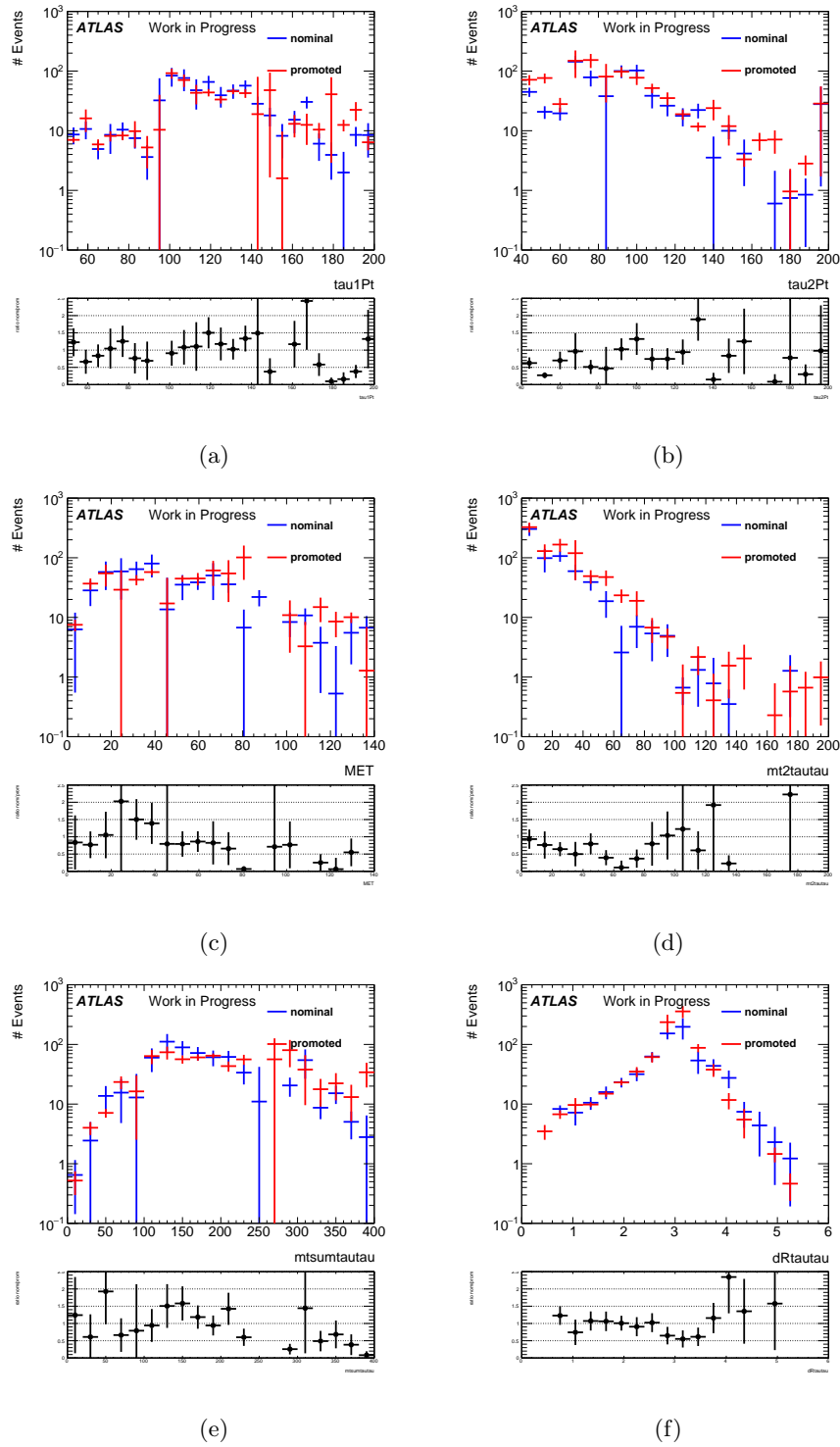


Figure 9.5.: Comparison of the distributions of  $p_T(\tau_1)$  (a),  $p_T(\tau_1)$  (b),  $E_T^{\text{miss}}$  (c),  $m_{T2}(\tau_1, \tau_1)$  (d),  $m_T(\tau_1) + m_T(\tau_2)$  (e) and  $\Delta R(\tau_1, \tau_2)$  (f) for  $W$ +jets before (blue) and after promoting one container  $\tau$ -lepton to medium ID (red) with modified  $p_T(\tau)$ -bins for the fake efficiency measurement at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . Compared to figure 9.2 the matching of the nominal and promoted distributions has notably improved. For these plots at least two oppositely charged medium  $\tau$ -leptons have been required as well as the offline thresholds for the asymmetric  $\tau$ - and the ditau+MET-trigger as described in section 7.1.

Sample ID	$p_T(W)$ -slice [GeV]	filter	before TP ( $\sigma_{\text{stat}}$ )	after TP ( $\sigma_{\text{stat}}$ )	$\frac{\text{after TP}}{\text{before TP}}$
364184	0-70	light-jet	106306 (3.44%)	102286 (0.77%)	0.96
364185	0-70	c-jet	31188.0 (4.29%)	28316.9 (1.27%)	0.91
364186	0-70	b-jet	7321.92 (5.67%)	7270.21 (1.04%)	0.99
364187	70-140	light-jet	24398.1 (3.52%)	27458.1 (3.67%)	1.13
364188	70-140	c-jet	11482.7 (2.56%)	10413.4 (0.86%)	0.91
364189	70-140	b-jet	4099.02 (3.07%)	3792.54 (0.81%)	0.93
364190	140-280	light-jet	7984.53 (1.96%)	8276.25 (0.96%)	0.96
364191	140-280	c-jet	5709.08 (1.66%)	5129.82 (0.53%)	0.90
364192	140-280	b-jet	1807.57 (1.02%)	1640.27 (0.26%)	0.91
364193	280-500	light-jet	1302.03 (2.13%)	1333.67 (0.53%)	0.99
364194	280-500	c-jet	1088.39 (2.18%)	1043.95 (0.60%)	1.02
361495	280-500	b-jet	403.869 (3.60%)	385.612 (0.57%)	0.95
364196	500-1000	-	511.796 (4.25%)	522.881 (0.51%)	1.02
364197	> 1000	-	39.1568 (2.39%)	40.4083 (0.57%)	1.03
<b>Total</b>			<b>203642 (1.98%)</b>	<b>198510 (0.67%)</b>	0.97

Table 9.2.: Event yields before and after  $\tau$ -promotion per  $W \rightarrow \tau\nu$ -dataset with a modified binning for the fake efficiency measurement. Compared to table 9.1 the number of weighted events for the promoted samples has in general increased. However, the statistical uncertainty  $\sigma_{\text{stat}}$  is also enhanced, making a further reduction of the bin widths unreasonable.

(a) Number of weighted events

Sample ID	$p_T(W)$ -slice [GeV]	filter	before TP	after TP	$\frac{\text{after TP}}{\text{before TP}}$
364184	0-70	light-jet	245230 (2.31%)	240703 (0.64%)	0.98
364185	0-70	c-jet	71241.5 (2.84%)	63725.8 (0.90%)	0.89
364186	0-70	b-jet	16948.7 (2.96%)	16379.8 (0.70%)	0.97
364187	70-140	light-jet	56160.4 (2.02%)	55729.5 (0.64%)	0.99
364188	70-140	c-jet	25996.4 (1.71%)	23627.0 (0.65%)	0.91
364189	70-140	b-jet	9236.19 (2.02%)	8599.10 (0.65%)	0.93
364190	140-280	light-jet	18744.7 (1.77%)	18346.6 (0.49%)	0.98
364191	140-280	c-jet	12802.2 (1.07%)	11537.2 (0.36%)	0.90
364192	140-280	b-jet	4141.87 (0.66%)	3767.55 (0.21%)	0.91
364193	280-500	light-jet	3052.41 (1.34%)	3005.92 (0.39%)	0.98
364194	280-500	c-jet	2509.69 (1.42%)	2303.96 (0.44%)	0.92
361495	280-500	b-jet	916.623 (1.93%)	831.150 (0.46%)	0.91
364196	500-1000	-	1197.32 (1.96%)	1150.21 (0.58%)	0.96
364197	> 1000	-	88.8970 (1.45%)	84.7783 (0.39%)	0.95
<b>Total</b>			<b>468267 (1.32%)</b>	<b>449792 (0.38%)</b>	0.96

(b) Number of unweighted events

Sample ID	$p_T(W)$ -slice [GeV]	filter	before TP	after TP	$\frac{\text{after TP}}{\text{before TP}}$
364184	0-70	light-jet	16616	281533	16.94
364185	0-70	c-jet	11384	167825	14.74
364186	0-70	b-jet	14070	238430	16.95
364187	70-140	light-jet	36565	658237	18.00
364188	70-140	c-jet	32126	499859	15.56
364189	70-140	b-jet	24825	421597	16.98
364190	140-280	light-jet	23759	539595	22.71
364191	140-280	c-jet	24180	467842	19.35
364192	140-280	b-jet	66338	1345528	20.28
364193	280-500	light-jet	11065	302857	27.37
364194	280-500	c-jet	8904	208728	23.44
361495	280-500	b-jet	7839	186124	23.74
364196	500-1000	-	13617	385165	28.29
364197	> 1000	-	8121	245921	30.28
<b>Total</b>			<b>299400</b>	<b>5949241</b>	19.87

Table 9.3.: Number of weighted events (a) and unweighted events (b) before and after promoting one container  $\tau$ -lepton to loose quality for each  $W \rightarrow \tau\nu$ -sample at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . To obtain these numbers at least two loose  $\tau$ -leptons are required. It can be already observed from the weighted events that the statistical uncertainty is notably reduced. This is due to the large increase of selected events after applying the  $\tau$ -promotion method. However, the number of weighted events is not exactly equal to the nominal yields. In fact, a slight systematic underestimation can be observed, as is the case for a promotion to medium quality.

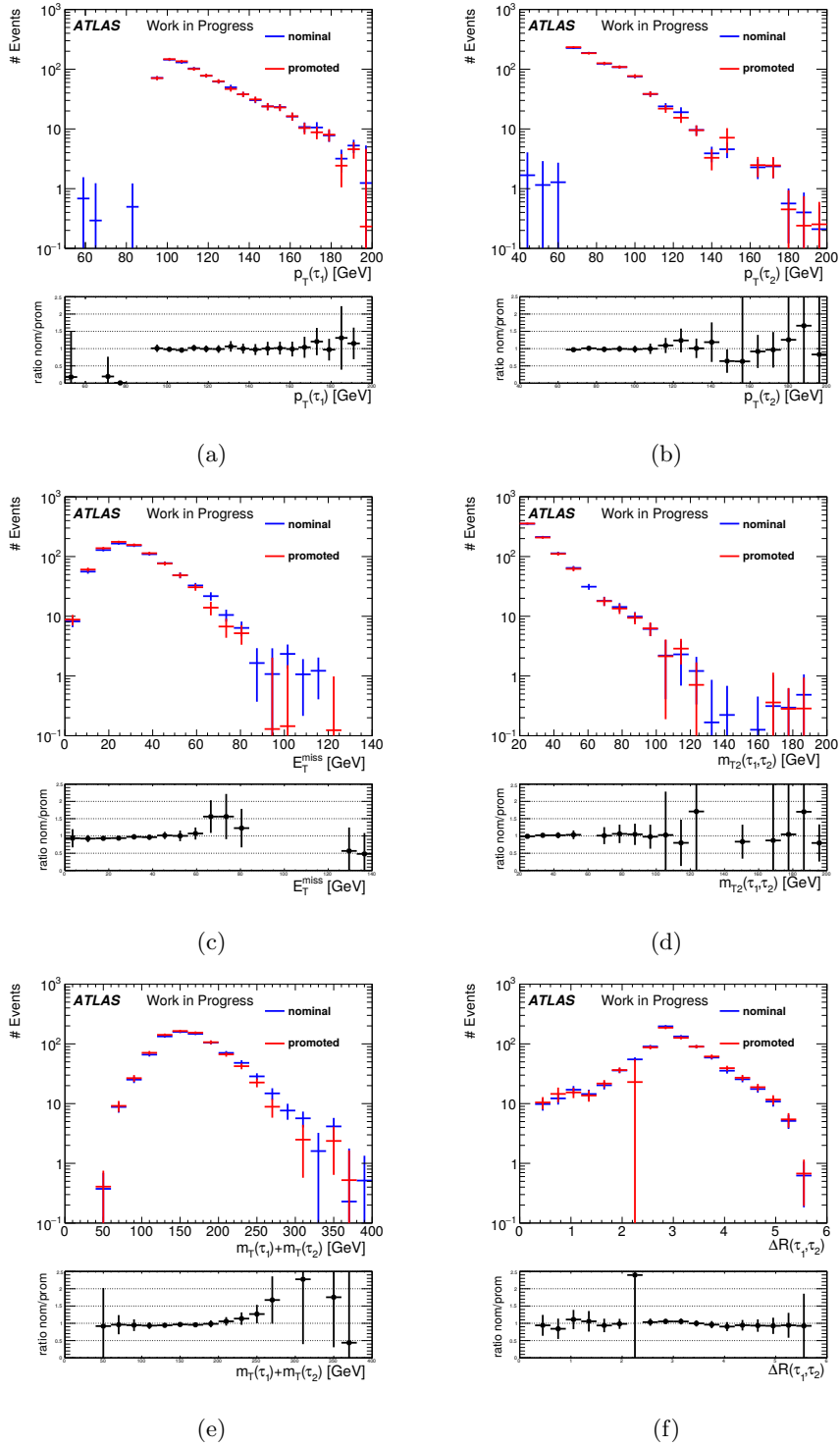


Figure 9.6.: Comparison of the nominal (blue) and the promoted QCD distributions (red) for  $p_T(\tau_1)$  (a),  $p_T(\tau_1)$  (b),  $E_T^{\text{miss}}$  (c),  $m_{T2}(\tau_1, \tau_1)$  (d),  $m_T(\tau_1) + m_T(\tau_2)$  (e) and  $\Delta R(\tau_1, \tau_2)$  (f) for an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . As can be observed from the ratio plots the ratio of the nominal and promoted event yields for each bin fluctuate mostly around 1, meaning that there is no large discrepancy between the QCD estimation with  $W$ +jets samples which includes  $\tau$ -promotion and with the nominal  $W$ +jets simulation. For these plots only the defining cuts for the QCD estimate (see section 6.2.5) have been required as well as the offline thresholds for the asymmetric  $\tau$ - and the ditau+MET-trigger as described in section 7.1.



Region	Data	Monte Carlo	Data – MC	QCD estimate ( $\sigma_{\text{stat}}$ )	nominal ( $\sigma_{\text{stat}}$ )
A	37121	1676.33	35444.67	-	-
B	21642	9582.59	12059.41	-	-
C	42514	1925.31	40588.69	-	-
E	13622	446.153	13175.85	-	-
<b>F</b>	6491	1593.24	4897.76	<b>4482.85</b> (0.79%)	4062.18 (1.05%)
<b>D</b>	-	-	-	<b>13809.6</b> (0.58%)	12245.90 (0.80%)

Table 9.4.: Event yields for data and Monte Carlo simulated samples in the control and validation regions used for QCD estimation (see section 6.2.5) and the expected amount of QCD events in signal region D (with the statistical uncertainty  $\sigma_{\text{stat}}$ ) for an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The computed transfer factor is 0.34. The ratio of the data to Monte Carlo difference and the expected QCD contribution is 1.09. Although the QCD estimations for validation region F and signal region D seem compatible with the results from table 6.1 the same underestimation can be observed as for the promotion to medium  $\tau$ -leptons in each of the control regions, making it unclear whether the QCD estimation is improved by a  $\tau$ -promoted loose  $W$ +jets sample or not.

### 9.3. Results with Tau Promotion

In the following, the  $W$ +jets dataset containing events with promoted  $\tau$ -leptons is tested in both of the signal region optimization methods presented in chapter 8. For each of them, the optimization is performed once with the nominal QCD estimation and once with the QCD dataset including a  $\tau$ -promoted  $W$ +jets sample. It will be checked which of them performs best and if better results can be achieved than in chapter 8 with  $\tau$ -promotion in general. The following optimizations will only be shown for high stau masses, as the previous results indicate a better sensitivity at larger values of  $m_{\tilde{\tau}}$ .

#### 9.3.1. Cut-and-Count Method

The same setting of variables and cuts is used as in section 8.2.1 with the preselection described in section 7.1. The chosen mass point for the optimization is ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV).

##### With Tau Promotion in QCD Estimation

The resulting cut combination that is obtained when not only applying  $\tau$ -promotion to the  $W$ +jets dataset but also to the  $W$ +jets tree contained in the QCD estimation is listed in table 9.5. The respective event yields for each background and for the signal mass point which the cuts are optimized for are shown in table 9.6 and are also visualized in the N-1 plots in figures D.1 and D.2. The significances are decreased compared to the high mass region in section 8.2.1 although the statistical uncertainty is improved for the  $W$ +jets background. However, this effect is not observed for QCD, indicating that at least for the kind of QCD estimation that is used,  $\tau$ -promotion cannot lower the statistical uncertainties.

##### SR DS HadHad High Mass with TP in $W$ +jets and QCD

$$\begin{aligned}
 m_{T2}(\tau_1, \tau_2) &> 50 \text{ GeV} \\
 p_T(\tau_1) &> 120 \text{ GeV} \\
 m_{\text{inv}}(\tau_1, \tau_2) &> 100 \text{ GeV} \\
 m_T(\tau_1) + m_T(\tau_2) &> 300 \text{ GeV} \\
 |\cos \theta_H| &< 0.6
 \end{aligned}$$

Table 9.5.: Resulting cut combination for an optimization on the mass point ( $m_{\tilde{\tau}} = 200$  GeV,  $m_{\tilde{\chi}_1^0} = 1$  GeV) with  $\tau$ -promotion applied to the  $W$ +jets and QCD datasets.

##### Without Tau Promotion in QCD Estimation

As the inclusion of a  $\tau$ -promoted  $W$ +jets dataset in the estimation of the QCD multijet background does not seem to deliver the desired performance, the nominal QCD sample is reused in the following cut optimization. In the resulting collection of cuts (see table 9.7), the significance in fact is improved for many signal mass points, resulting in seven

Process	Events (statistical uncertainty)
$W$ +jets	7.93 (14.25%)
$Z$ +jets	1.47 (99.31%)
Diboson	0.54 (125.92%)
Top	0.59 (61.86%)
QCD	3.18 (97.87%)
total background	13.82 (27.38%)
DS signal	14.32 (6.70%)

Table 9.6.: Event yields for each background process when applying the cuts listed in table 9.5 for an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . For  $W$ +jets and QCD, the  $\tau$ -promotion method was used to increase the number of weighted  $W \rightarrow \tau\nu$ -events. Compared to the high stau mass region in section 8.2.1, the significance is notably decreased, preventing an exclusion at 95% confidence level. It can be observed that the number of events in this region is larger than in region "SR DS HadHad High Mass" (see table 8.7), especially for  $W$ +jets. Although  $W$ +jets is still the dominant background, the main restriction for sensitivity in this case is due to the high statistical uncertainties, this time originating from  $Z$ +jets, the dibosonic and the QCD background. The dibosonic statistical uncertainty is due to two events with negative weights with large absolute value compared to the other event weights.  $\tau$ -promotion seems to have improved the uncertainty of  $W$ +jets, while the statistical uncertainty of QCD is still very high.

points exceeding a z-value of 1.64 (see figure 9.8). Additionally, the  $W$ +jets statistical uncertainty is below 50% (see table 9.8), indicating that the  $\tau$ -promotion method indeed decreases the statistical uncertainty even if rather tight signal region cuts are applied. Despite the promising result, there is a drawback due to the statistical uncertainties of other background processes, especially from the diboson dataset. The N-1 plots are shown in figures D.3 and D.4.

SR DS HadHad High Mass with TP in  $W$ +jets

$$\begin{aligned}
 E_{\text{T}}^{\text{miss}} &> 70 \text{ GeV} \\
 m_{\text{T}2}(\tau_1, \tau_2) &> 75 \text{ GeV} \\
 p_{\text{T}}(\tau_2) &> 50 \text{ GeV} \\
 m_{\text{inv}}(\tau_1, \tau_2) &> 140 \text{ GeV} \\
 m_{\text{T}}(\tau_1) + m_{\text{T}}(\tau_2) &> 300 \text{ GeV} \\
 |\cos \theta_U| &< 0.7
 \end{aligned}$$

Table 9.7.: Cut combination obtained in an optimization for a stau mass of 200 GeV and a neutralino mass of 1 GeV, using a  $W$ +jets dataset which includes  $\tau$ -promotion.

In general the statistical uncertainty of the  $W$ +jets background has decreased due to the  $\tau$ -promotion method, which can be observed e.g. from table 9.8 when comparing it to table 8.7. Both of the signal regions are optimized for the same signal mass point with the same samples, the only difference being the application of  $\tau$ -promotion in the  $W$ +jets

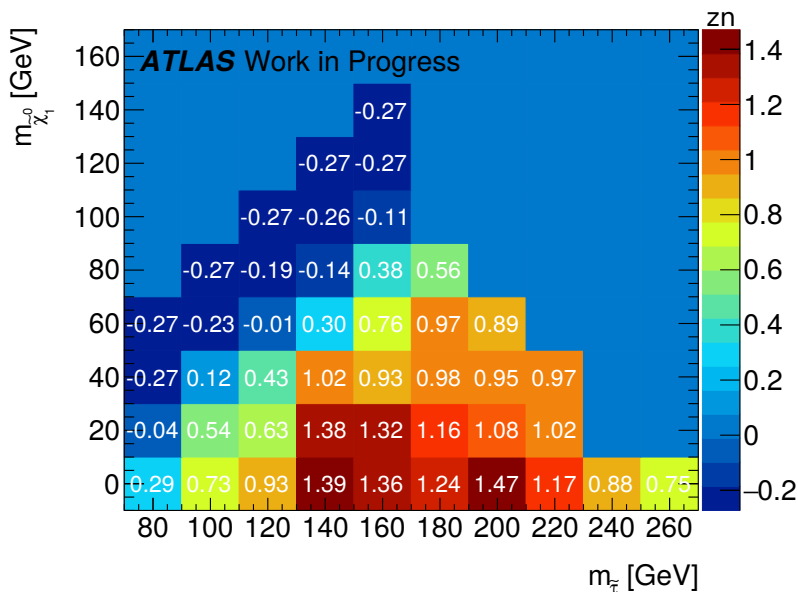


Figure 9.7.: Significances per signal mass point in the cut region defined in table 9.5 for  $W$ +jets and QCD containing events with promoted  $\tau$ -leptons at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The largest  $z$ -values are reached for stau masses between 140 GeV and 200 GeV. However, there is no mass point for which the significance exceeds a value of 1.64, preventing an exclusion at 95% CL.

dataset in the former. This not only leads to different cut combinations, but also to a decrease of the statistical  $W$ +jets uncertainty from 88.89% to 48.00%. Comparing tables 9.6 and 9.8 it can be noted that the statistical uncertainty of the QCD background yield is even larger if  $\tau$ -promotion is included in the QCD estimation than if the nominal dataset is used. There is one main reason, why the QCD estimate's performance is diminished when promoting container  $\tau$ -leptons to loose ID. The control regions A and C are gaining many events with loose  $\tau$ -leptons while the number of events in regions B and D stays the same as at least two medium  $\tau$ -leptons are required. Although the overall reweighting in the  $\tau$ -promotion restores the total number of events, it leads to an imbalance of the numbers of loose and medium (fake)  $\tau$ -leptons that becomes visible if the dataset is split into these two categories and which affects the transfer factor calculation. A possible way to solve this would be to split the  $W$ +jets sample into two parts at random. Then, to one of them the  $\tau$ -promotion to loose ID is applied while for the other the container  $\tau$ -leptons are promoted to medium ID.

### 9.3.2. Boosted Decision Trees

For the training of boosted decision trees a certain amount of events is required. However, it is not reasonable to let all background samples be dominated by one or two datasets with an increased number of unweighted events as the discrimination power of the BDT response is reduced concerning the other background processes. Therefore, instead of including the promoted samples in a new BDT training procedure, the BDT training of section 8.3.7 is reused and evaluated with  $W$ +jets and QCD samples which contain

Process	Events (statistical uncertainty)
$W$ +jets	1.75 (48.00%)
$Z$ +jets	0.17 (58.82%)
Diboson	0.66 (96.97%)
Top	0.67 (73.68%)
QCD	2.52 (45.24%)
total background	5.77 (28.42%)
DS signal	10.24 (7.42%)

Table 9.8.: Event yields for each background and for the signal that the cut combination in table 9.7 was optimized for, at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The statistical uncertainty on the number of events for  $W$ +jets is notably decreased due to  $\tau$ -promotion, which considerably improves the significance of many signal mass points. However, the statistical uncertainties are (still) large for some of the other backgrounds, especially for dibosonic processes.

$\tau$ -promoted events.

### With Tau Promotion in QCD Estimation

Both the  $\tau$ -promoted  $W$ +jets and QCD samples are used in the evaluation shown in figure 9.9. The cut optimization is performed with the requirement of at least two background events with a maximum total statistical uncertainty of 50% and each background yielding more than 0 events. The resulting cut is at a BDT score of 0.78, but with surprisingly low significances, which is shown in figure 9.10. The reason for this can be observed in table 9.9. The QCD estimate for this cut yields by far the largest statistical uncertainty, which does not only influence the total background uncertainty in this particular region, but also prevents the selection of a tighter cut with a better signal to background ratio. The reason for this is the reduced performance of the ABCD estimation of the QCD background, as explained previously.

### Without Tau Promotion in QCD Estimation

The re-evaluation of the high stau mass BDT presented in section 8.3.7 is performed with  $\tau$ -promotion applied in  $W$ +jets but with the nominal QCD dataset. The obtained distribution is depicted in figure 9.11. It shows less fluctuations in the  $W$ +jets distribution than in the evaluation with the nominal  $W$ +jets dataset (see figure 8.17). When optimizing the cut on the BDT response, the total statistical background uncertainty was required to be smaller than 50%, there should be at least two weighted background events left after the cut and each single background process should yield more than 0 events. The resulting cut value is a BDT score of 0.9, which is the same cut resulting for the high stau mass BDT in the optimization in section 8.3.7. The event yields, including the modified  $W$ +jets yields, are given in table 9.10. The statistical uncertainty of  $W$ +jets and consequently the total background uncertainty are notably decreased by the application of  $\tau$ -promotion. This allows for much higher significances, as shown in figure 9.12. Considering the  $z$ -

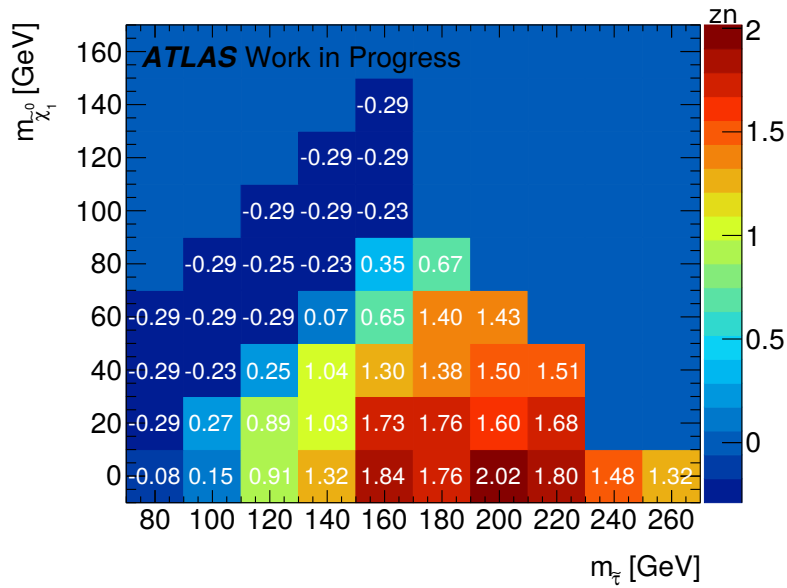


Figure 9.8.: Significances in the region defined by the cuts listed in table 9.7 for each signal mass point at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . There are seven points which show z-values above 1.64, making an exclusion an 95% CL possible.

values, it would be possible to exclude 14 signal mass points. However, the issue of large uncertainties is not entirely solved as other backgrounds than  $W$ +jets, namely the QCD multijet and dibosonic background processes still show uncertainties of 80% and larger. At this point, one can exploit the fact that the BDT tends to accumulate signal events near to a value of +1 and background events at -1. If the cut at 0.9 is softened a little bit, the number of background events will increase but their uncertainties might decrease. This was not feasible before, as the z-values usually were already very low and softening the BDT cut would have even reduced them further. In table 9.11, the yields of a cut at a BDT score of 0.88 are shown. The resulting significances are plotted in figure 9.13. As expected, the significances of all mass points have decreased due to the lower cut, but some of the backgrounds show smaller statistical uncertainties. However, the uncertainty is still about 80% for the QCD multijet background. As it has already been stated in section 6.2.5, the accuracy of the QCD estimation could be improved by using different estimation methods.

Process	Events (statistical uncertainty)
$W$ +jets	32.55 (15.96%)
$Z$ +jets	16.89 (71.69%)
Diboson	13.95 (11.01%)
Top	6.12 (23.25%)
QCD	0.05 (137.31%)
Total Background	69.57 (20.13%)
Signal	21.11 (5.24%)

Table 9.9.: Event yields for a cut on a high stau mass BDT at 0.78, where the evaluation was performed with  $\tau$ -promotion applied to the  $W$ +jets dataset and the loose  $W$ +jets sample that was used in the QCD estimation. The event yields are scaled to an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . It can be noted that the resulting statistical uncertainty is very large (caused by negative event weights), which prevents further tightening of the cut and keeps the z-values for all mass points low (see 9.10).

Process	Events (statistical uncertainty)
$W$ +jets	0.88 (66.85%)
$Z$ +jets	0.84 (51.10%)
Diboson	0.80 (81.56%)
Top	0.19 (45.67%)
QCD	0.46 (125.03%)
Total Background	3.17 (35.95%)
Signal	10.14 (8.23%)

Table 9.10.: Event yields for a lower cut on a high-stau-mass-BDT score at 0.9 for an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The event yields entirely correspond to those in table 8.23 (because the signal region is in fact identical), apart from  $W$ +jets, for which the number of events is slightly reduced (from 1.05 to 0.88). More importantly, the relative statistical uncertainty is reduced by a factor of approximately 2, leading to a decrease of the total statistical uncertainty of about 15%.

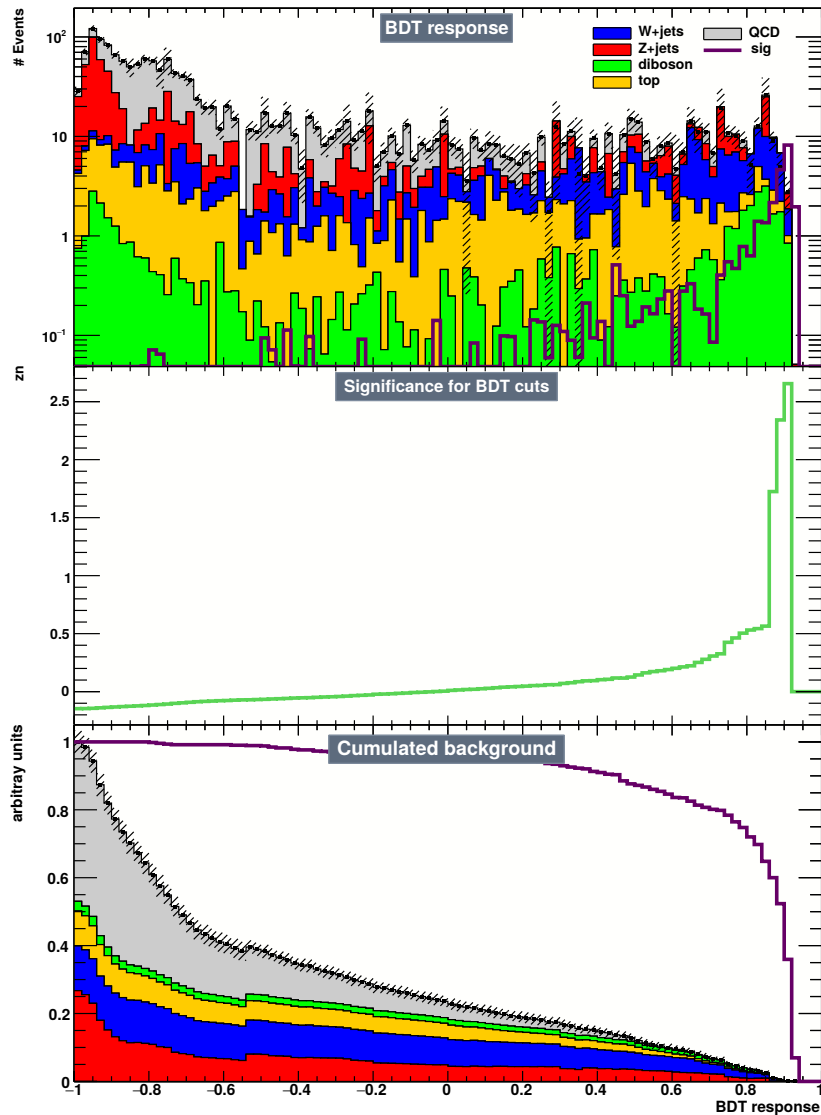


Figure 9.9.: BDT evaluated with  $\tau$ -promoted  $W$ +jets and QCD datasets at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . The distribution for the signal mass point ( $m_{\tau} = 200 \text{ GeV}$ ,  $m_{\chi_1^0} = 1 \text{ GeV}$ ) is depicted by the violet line. In the cumulated background plot a kink can be observed at a BDT score of approximately  $-0.55$ . This is due to negatively weighted events contained in that particular bin, which lowered the cumulated distribution to its left.



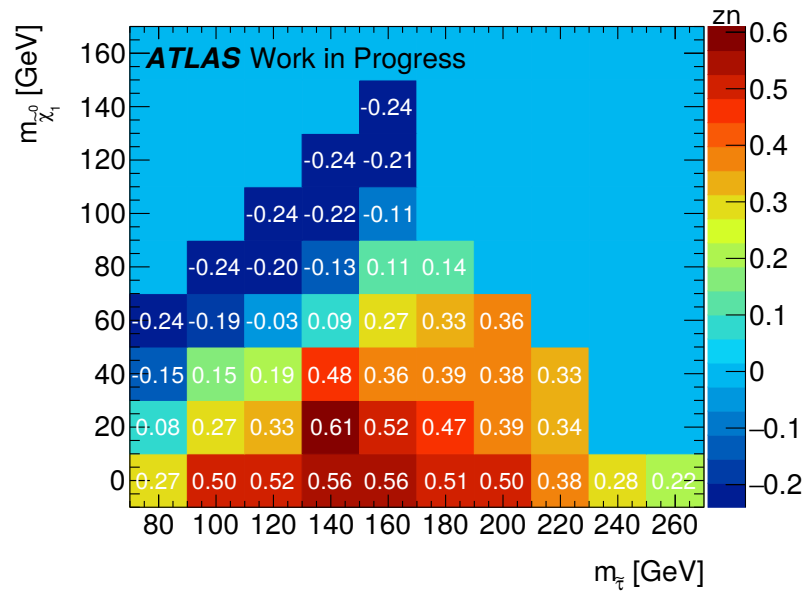


Figure 9.10.: Significances per signal mass point for a lower cut at a high-stau-mass BDT score of 0.78 at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . In comparison to 8.18, the significances are notably reduced due to the large statistical uncertainty of the QCD multijet background.

Process	Events (statistical uncertainty)
$W$ +jets	5.44 (20.05%)
$Z$ +jets	0.92 (68.49%)
Diboson	2.55 (30.18%)
Top	0.66 (55.23%)
QCD	1.29 (79.27%)
Total Background	10.86 (16.87%)
Signal	14.78 (6.53%)

Table 9.11.: Event yields for a cut on a high stau mass BDT at 0.88 for an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . Compared to table 9.10, all event yields are increased. The statistical uncertainties have been lowered for  $W$ +jets, diboson, and QCD. The total statistical uncertainty is reduced to 17%. Still, the largest uncertainty originates from QCD, followed by  $Z$ +jets.

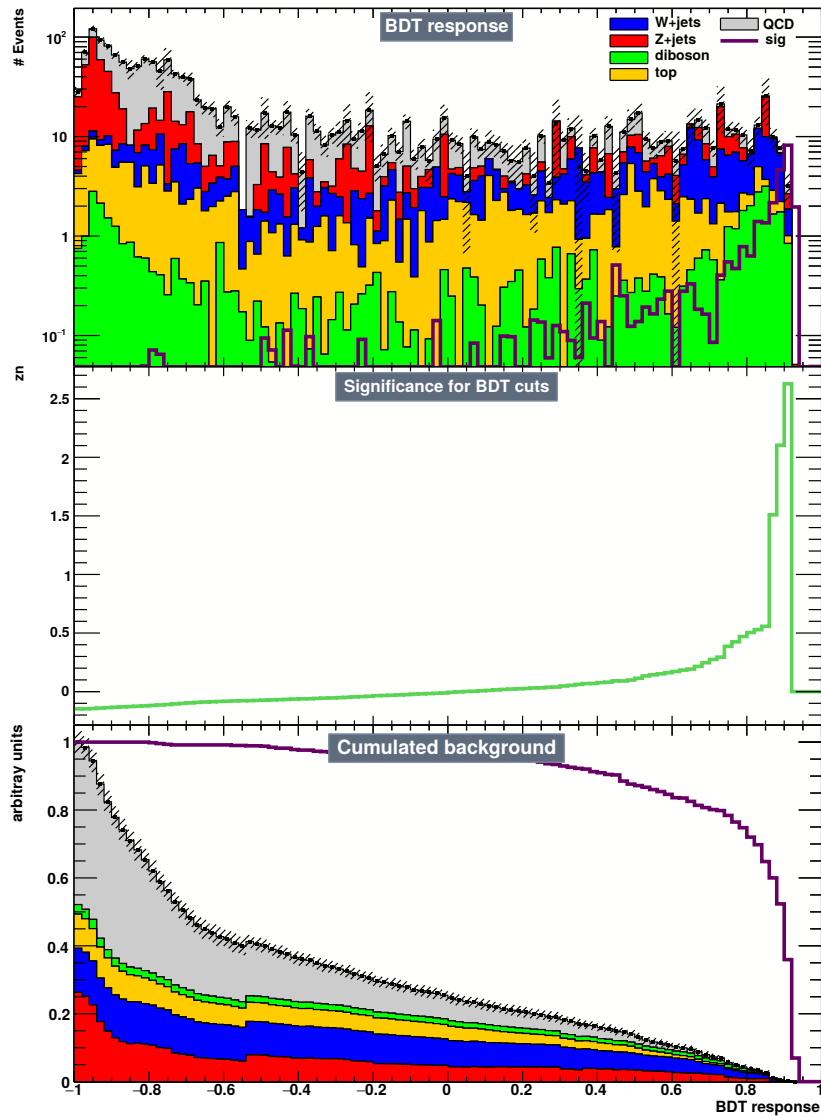


Figure 9.11.: BDT evaluated with a  $\tau$ -promoted  $W$ +jets dataset at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . Already from the upper plot, a reduction of the fluctuations of the  $W$ +jets distribution (blue) is visible compared to figure 8.17.

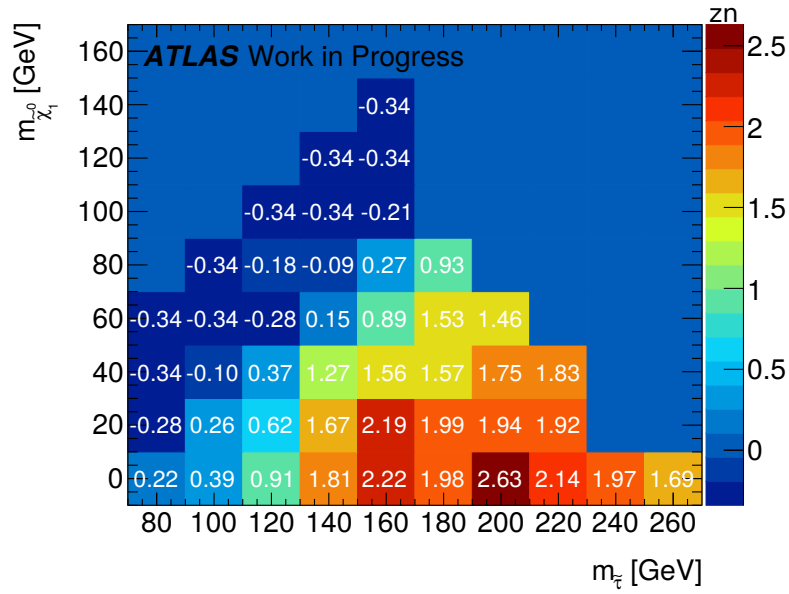


Figure 9.12.: Significances per signal mass point for a lower cut at a high-stau-mass BDT score of 0.9 at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . In comparison to 8.18, the z-values have notably increased for several mass points. In total, 14 mass points exceed the critical z-value of 1.64.

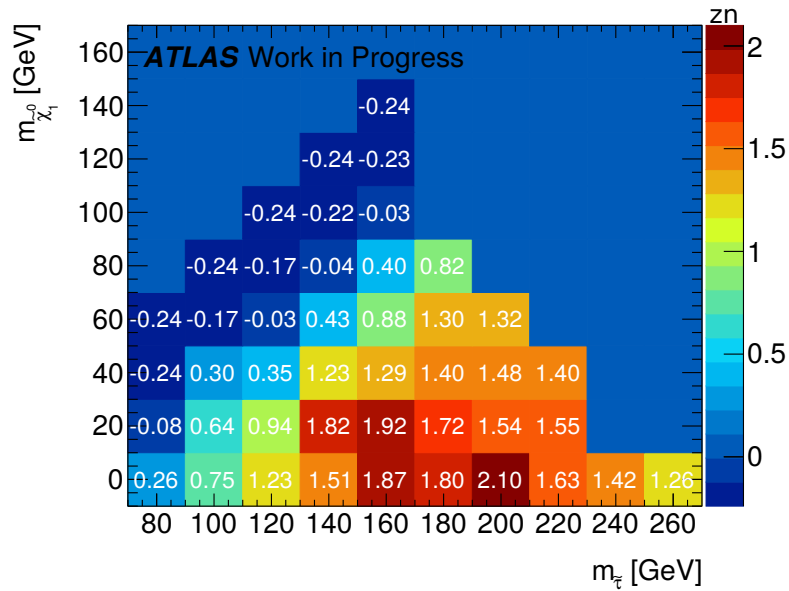


Figure 9.13.: Significances per signal mass point for a lower cut at a high-stau-mass BDT score of 0.88 at an integrated luminosity of  $36.1 \text{ fb}^{-1}$ . In comparison to figure 9.12, the significances are reduced, as expected. Still, six mass points show significances above 1.64 and would allow for exclusion.



## 10. Summary and Interpretation of Results

In total, ten different attempts at signal-region definitions have been shown in these studies. In the following, the corresponding observations together with their reasons and consequences shall be summarized.

The LepHad-channel was examined in a standard cut-and-count approach for signal region optimization. However, this attempt failed due to two main reasons: First, the light signal lepton as a  $\tau$ -decay product is softer than the corresponding visible part of a hadronic  $\tau$ -lepton. Second, the branching fraction especially for  $W$ +jets is increased compared to the HadHad-channel. In these studies, it was decided to concentrate on the sensitivity for the HadHad-channel. However, there are possibilities to improve the LepHad sensitivity, e.g. by using multivariate methods.

The cut-and-count approach also was employed for the HadHad-channel. Cut combinations optimized on three different signal mass points have been presented. Although in run 1 the largest sensitivity was observed at stau masses of 80 GeV and 100 GeV, this does not seem to be the case in run 2. Instead, the sensitivity is larger for high stau masses which can also be observed in the BDT studies. One reason is that the cross-sections of the particular signal and background processes are not increased by the same factor when raising the center-of-mass energy of the proton-proton collisions. Moreover, tight restrictions on the transverse  $\tau$ -momenta placed by the trigger selection tend to have a stronger impact on the low stau mass region. Furthermore,  $m_{T2}$  is lacking discrimination power for low stau masses (which has already been the case in run 1). To be able to exclude low stau mass points, modified triggers are needed with lower selection thresholds and better reconstruction efficiencies for  $\tau$ -leptons.

The high stau mass region in general shows good results in terms of sensitivity. But it has to be taken into account that the statistical uncertainties can grow up to 100% and more for some backgrounds. This can be explained by an insufficient number of unweighted events or a few negatively weighted events dominating the region of phase space for this particular background. Limited statistics is, as has already been mentioned, one of the major issues of this analysis. There are in principle two ways of overcoming this problem. One could either provide even larger simulations with more unweighted events, which might exceed reasonable computing time and resources, or one applies a reweighting technique like the  $\tau$ -promotion, which has been proven to reduce the  $W$ +jets statistical uncertainty notably. In the latter case, one has to deal then with the intrinsic systematic uncertainties of the method. This has not been taken into account in these studies but it would definitely be necessary along with the determination of other systematics, which have been roughly covered by a flat value of 30% of the background yield. Moreover, in these studies, only the improvement of the  $W$ +jets statistics has been undertaken, whereas also other important background processes like  $Z$ +jets and dibosonic processes would need an improvement in statistics by some means. Additionally, the QCD multijet contribution has been determined by an ABCD approach which proved to be sufficient in these studies.

However, to further improve the direct stau sensitivity in the HadHad-channel also the uncertainties arising from the QCD estimation (systematic and statistical) need to be reduced. This could be achieved by alternative, more sophisticated estimation methods like jet smearing [35] or the fake-factor method (described in [34]).

The summary of the results is provided in table 10.1.

Methods		zn HadHad (mass points with $zn > 1.64$ )			zn LepHad (mass points with $zn > 1.64$ )
Separation	Tau Promotion	Low Mass	Medium Mass	High Mass	
C&C	-	1.19 (0)	1.56 (0)	<b>1.78</b> (2)	-0.07 (0)
BDT	-	0.63 (0)		<b>2.16</b> (4)	-
C&C	$W$ +jets	-	-	<b>2.02</b> (7)	-
BDT	$W$ +jets	-	-	<b>2.63</b> (14)	-
C&C	$W$ +jets, QCD	-	-	1.47 (0)	-
BDT	$W$ +jets, QCD	-	-	0.61 (0)	-

Table 10.1.: In this table the best z-values obtained with the respective methods of cut-and-count (C&C) and boosted decision trees (BDT), are listed. Note, that the z-values do not have to be the ones for the mass points the cuts were optimized or the BDT was trained for. For the BDT training, the low and intermediate stau-mass regions were combined to gain more statistics. The best significances are obtained for high stau masses when applying  $\tau$ -promotion to the  $W$ +jets dataset (only).

## 11. Conclusion

In run 1 of the LHC, the search for the pair production of  $\tau$ -sleptons with ATLAS data delivered no sensitivity for this low-cross-section process. In run 2, the center-of-mass energy has increased as well as the amount of available statistics. However, the search for direct stau production still shows some difficult challenges.

The LepHad-channel is completely out of reach, if no better way of analysis is found, due to an overwhelming amount of expected background events and the difficulties that come with the leptonic  $\tau$ -decay, leading to an accumulation of light leptons at low transverse momenta. Also an optimization of cuts could not provide any sensitivity for this channel.

The situation for the HadHad-channel turned out to be better than for the LepHad-channels, although there are also some issues to overcome. The available lowest un-prescaled triggers for two hadronic  $\tau$ -leptons do not only suffer from lacking discrimination between  $\tau$ -leptons and electrons, they also have very high online and offline thresholds, already cutting away the majority of signal events. Therefore, additional cut optimizations only showed little sensitivity for some mass points. Additionally, an attempt to train boosted decision trees for different mass regimes was made, also investigating the possibility of parameterized BDTs. As it turned out, in the analysis the lack of simulated events is not only a problem for the signal but also for some of the background datasets, especially  $W$ +jets. In order to solve this, a newly developed method, called  $\tau$ -promotion, was applied to increase the available statistics for  $W$ +jets and thereby reduce the statistical uncertainties. Although there are some promising results, also indicating sensitivity for several signal mass points, there is still need for a reduction of statistical uncertainties for datasets like the ones for  $Z$ +jets and dibosonic processes. Furthermore, the estimate for the QCD multijet background could be done with other methods, which might also reduce the statistical uncertainty. Finally, the search for direct stau pair production will be continued with the full run 2 data for which about  $120 \text{ fb}^{-1}$  are expected. The increased integrated luminosity could improve the sensitivity for this process notably.





# A. Appendix I: List of Samples

## A.1. Signal Samples

Dataset ID	$m_{\tilde{\tau}}$ [GeV]	$m_{\tilde{\chi}_1^0}$ [GeV]	$\sigma_{\tilde{\tau}_1}^{\text{eff}}$ [pb]	$\sigma_{\tilde{\tau}_2}^{\text{eff}}$ [pb]
393600	80	1	0.3081	0.1090
393601	80	20	0.2982	0.1055
393602	80	40	0.2582	0.0913
393603	80	60	0.1548	0.0548
393604	100	1	0.1591	0.0578
393605	100	20	0.1578	0.0573
393606	100	40	0.1483	0.0538
393607	100	60	0.1266	0.0460
393608	100	80	0.0756	0.0274
393609	120	1	0.0914	0.0337
393610	120	20	0.0913	0.0337
393611	120	40	0.0887	0.0328
393612	120	60	0.0822	0.0304
393613	120	80	0.0703	0.0260
393614	120	100	0.0410	0.0151

Table A.1.: Here the signal mass points simulated up to a stau mass of 120 GeV are listed which are used for this analysis. The cross-sections for each mass point are given, also considering the filter efficiencies (hence  $\sigma^{\text{eff}}$ ). The filter efficiencies are needed to account for certain preselections (filters) at generator level.

Dataset ID	$m_{\tilde{\tau}}$ [GeV]	$m_{\tilde{\chi}_1^0}$ [GeV]	$\sigma_{\tilde{\tau}_1}^{\text{eff}}$ [pb]	$\sigma_{\tilde{\tau}_2}^{\text{eff}}$ [pb]
393615	140	1	0.0566	0.0260
393616	140	20	0.0560	0.0211
393617	140	40	0.0556	0.0209
393618	140	60	0.0531	0.0198
393619	140	80	0.0497	0.0186
393620	140	100	0.0421	0.0157
393621	140	120	0.0240	0.0089
393622	160	1	0.0365	0.0137
393623	160	20	0.0364	0.0137
393624	160	40	0.0359	0.0135
393625	160	60	0.0353	0.0133
393626	160	80	0.0339	0.0128
393627	160	100	0.0314	0.0118
393628	160	120	0.0266	0.0100
393629	160	140	0.0151	0.0057
393630	180	1	0.0245	0.0093
393631	180	20	0.0245	0.0093
393632	180	40	0.0243	0.0092
393633	180	60	0.0239	0.0091
393634	180	80	0.0234	0.0088
393635	200	1	0.0171	0.0065
393636	200	20	0.0170	0.0065
393637	200	40	0.0170	0.0065
393638	200	60	0.0168	0.0064
393639	220	1	0.0122	0.0047
393640	220	20	0.0121	0.0046
393641	220	40	0.0121	0.0046
393642	240	1	0.0088	0.0034
393643	260	1	0.0066	0.0025

Table A.2.: Listed here are the signal mass points simulated for stau masses between 140 GeV and 260 GeV which are used for this analysis. The cross-sections for each mass point are given, also considering the filter efficiencies (hence  $\sigma^{\text{eff}}$ ). The filter efficiencies are needed to account for certain preselections (filters) at generator level.

## A.2. Background Samples

Dataset ID	$W$ -decay	$\max(H_T, p_T(W))$ [GeV]	filters
364156	$W \rightarrow \mu\nu$	0-70	light-jet filter
364157	$W \rightarrow \mu\nu$	0-70	c-jet filter
364158	$W \rightarrow \mu\nu$	0-70	b-jet filter
364159	$W \rightarrow \mu\nu$	70-140	light-jet filter
364160	$W \rightarrow \mu\nu$	70-140	c-jet filter
364161	$W \rightarrow \mu\nu$	70-140	b-jet filter
364162	$W \rightarrow \mu\nu$	140-280	light-jet filter
364163	$W \rightarrow \mu\nu$	140-280	c-jet filter
364164	$W \rightarrow \mu\nu$	140-280	b-jet filter
364165	$W \rightarrow \mu\nu$	280-500	light-jet filter
364166	$W \rightarrow \mu\nu$	280-500	c-jet filter
364167	$W \rightarrow \mu\nu$	280-500	b-jet filter
364168	$W \rightarrow \mu\nu$	500-1000	-
364169	$W \rightarrow \mu\nu$	> 1000	-
364170	$W \rightarrow e\nu$	0-70	light-jet filter
364171	$W \rightarrow e\nu$	0-70	c-jet filter
364172	$W \rightarrow e\nu$	0-70	b-jet filter
364173	$W \rightarrow e\nu$	70-140	light-jet filter
364174	$W \rightarrow e\nu$	70-140	c-jet filter
364175	$W \rightarrow e\nu$	70-140	b-jet filter
364176	$W \rightarrow e\nu$	140-280	light-jet filter
364177	$W \rightarrow e\nu$	140-280	c-jet filter
364178	$W \rightarrow e\nu$	140-280	b-jet filter
364179	$W \rightarrow e\nu$	280-500	light-jet filter
364180	$W \rightarrow e\nu$	280-500	c-jet filter
364181	$W \rightarrow e\nu$	280-500	b-jet filter
364182	$W \rightarrow e\nu$	500-1000	-
364183	$W \rightarrow e\nu$	> 1000	-

Table A.3.: Used samples for the  $W$ +jets background. Each sample represents a different leptonic decay channel (the decays into  $\tau$ -leptons are given in table A.4) of the  $W$ -boson in another transverse momentum regime with either a filter for light jets (originating from up-, down-, or strange-quarks) or filters for charm- or bottom-jets. Where no filter is given, the samples have been calculated inclusively for all jet types. The total cross-section for  $W$ +jets processes is 20079.97 pb, also including higher orders. The samples have been produced for a center-of-mass energy of 13 TeV with Sherpa 2.2.1.

Dataset ID	$W$ -decay	$\max(H_T, p_T(W))$ [GeV]	filters
364184	$W \rightarrow \tau\nu$	0-70	light-jet filter
364185	$W \rightarrow \tau\nu$	0-70	c-jet filter
364186	$W \rightarrow \tau\nu$	0-70	b-jet filter
364187	$W \rightarrow \tau\nu$	70-140	light-jet filter
364188	$W \rightarrow \tau\nu$	70-140	c-jet filter
364189	$W \rightarrow \tau\nu$	70-140	b-jet filter
364190	$W \rightarrow \tau\nu$	140-280	light-jet filter
364191	$W \rightarrow \tau\nu$	140-280	c-jet filter
364192	$W \rightarrow \tau\nu$	140-280	b-jet filter
364193	$W \rightarrow \tau\nu$	280-500	light-jet filter
364194	$W \rightarrow \tau\nu$	280-500	c-jet filter
364195	$W \rightarrow \tau\nu$	280-500	b-jet filter
364196	$W \rightarrow \tau\nu$	500-1000	-
364197	$W \rightarrow \tau\nu$	> 1000	-

Table A.4.: Used samples for the  $W$ +jets background. The samples given in this table represent the decays of the  $W$ -boson into a  $\tau$ -lepton and its neutrino in different transverse momentum regimes with either a filter for light jets (originating from up-, down-, or strange-quarks) or filters for charm- or bottom-jets. Where no filter is given, the samples have been calculated inclusively for all jet types. The total cross-section for  $W$ +jets processes is 20079.97 pb, also including higher orders. The samples have been produced for a center-of-mass energy of 13 TeV with Sherpa 2.2.1. These samples are also used for the container  $\tau$ -promotion method (see chapter 9).

Dataset ID	$Z$ -decay	$\max(H_T, p_T(Z))$ [GeV]	filters
364100	$Z \rightarrow \mu\mu$	0-70	light-jet filter
364101	$Z \rightarrow \mu\mu$	0-70	c-jet filter
364102	$Z \rightarrow \mu\mu$	0-70	b-jet filter
364103	$Z \rightarrow \mu\mu$	70-140	light-jet filter
364104	$Z \rightarrow \mu\mu$	70-140	c-jet filter
364105	$Z \rightarrow \mu\mu$	70-140	b-jet filter
364106	$Z \rightarrow \mu\mu$	140-280	light-jet filter
364107	$Z \rightarrow \mu\mu$	140-280	c-jet filter
364108	$Z \rightarrow \mu\mu$	140-280	b-jet filter
364109	$Z \rightarrow \mu\mu$	280-500	light-jet filter
364110	$Z \rightarrow \mu\mu$	280-500	c-jet filter
364111	$Z \rightarrow \mu\mu$	280-500	b-jet filter
364112	$Z \rightarrow \mu\mu$	500-1000	-
364113	$Z \rightarrow \mu\mu$	> 1000	-
364114	$Z \rightarrow ee$	0-70	light-jet filter
364115	$Z \rightarrow ee$	0-70	c-jet filter
364116	$Z \rightarrow ee$	0-70	b-jet filter
364117	$Z \rightarrow ee$	70-140	light-jet filter
364118	$Z \rightarrow ee$	70-140	c-jet filter
364119	$Z \rightarrow ee$	70-140	b-jet filter
364120	$Z \rightarrow ee$	140-280	light-jet filter
364121	$Z \rightarrow ee$	140-280	c-jet filter
364122	$Z \rightarrow ee$	140-280	b-jet filter
364123	$Z \rightarrow ee$	280-500	light-jet filter
364124	$Z \rightarrow ee$	280-500	c-jet filter
364125	$Z \rightarrow ee$	280-500	b-jet filter
364126	$Z \rightarrow ee$	500-1000	-
364127	$Z \rightarrow ee$	> 1000	-

Table A.5.: List of used  $Z$ +jets samples. The samples are divided according to the leptonic decay channels of the  $Z$ -boson (here the decays to electrons and muons are given, for the other decay channels see table A.6), the transverse momentum range and the applied filters as already explained for  $W$ +jets in table A.3. The samples are produced with Sherpa 2.2.1 with a total cross-section of 6321 pb.

Dataset ID	$Z$ -decay	$\max(H_T, p_T(Z))$ [GeV]	filters
364128	$Z \rightarrow \tau\tau$	0-70	light-jet filter
364129	$Z \rightarrow \tau\tau$	0-70	c-jet filter
364130	$Z \rightarrow \tau\tau$	0-70	b-jet filter
364131	$Z \rightarrow \tau\tau$	70-140	light-jet filter
364132	$Z \rightarrow \tau\tau$	70-140	c-jet filter
364133	$Z \rightarrow \tau\tau$	70-140	b-jet filter
364134	$Z \rightarrow \tau\tau$	140-280	light-jet filter
364135	$Z \rightarrow \tau\tau$	140-280	c-jet filter
364136	$Z \rightarrow \tau\tau$	140-280	b-jet filter
364137	$Z \rightarrow \tau\tau$	280-500	light-jet filter
364138	$Z \rightarrow \tau\tau$	280-500	c-jet filter
364139	$Z \rightarrow \tau\tau$	280-500	b-jet filter
364140	$Z \rightarrow \tau\tau$	500-1000	-
364141	$Z \rightarrow \tau\tau$	> 1000	-
364142	$Z \rightarrow \nu\nu$	0-70	light-jet filter
364143	$Z \rightarrow \nu\nu$	0-70	c-jet filter
364144	$Z \rightarrow \nu\nu$	0-70	b-jet filter
364145	$Z \rightarrow \nu\nu$	70-140	light-jet filter
364146	$Z \rightarrow \nu\nu$	70-140	c-jet filter
364147	$Z \rightarrow \nu\nu$	70-140	b-jet filter
364148	$Z \rightarrow \nu\nu$	140-280	light-jet filter
364149	$Z \rightarrow \nu\nu$	140-280	c-jet filter
364150	$Z \rightarrow \nu\nu$	140-280	b-jet filter
364151	$Z \rightarrow \nu\nu$	280-500	light-jet filter
364152	$Z \rightarrow \nu\nu$	280-500	c-jet filter
364153	$Z \rightarrow \nu\nu$	280-500	b-jet filter
364154	$Z \rightarrow \nu\nu$	500-1000	-
364155	$Z \rightarrow \nu\nu$	> 1000	-

Table A.6.: List of used  $Z$ +jets samples. The samples are divided according to the leptonic decay channels of the  $Z$ -boson (here the decay to  $\tau$ -leptons and neutrinos are given, for the other decay channels see table A.5), the transverse momentum range and the applied filters as explained for  $W$ +jets in table A.3. The samples are produced with Sherpa 2.2.1 with a total cross-section of 6321 pb.

Dataset ID	diboson channel / final state	cross-section [pb]
361069	$ll\nu\nu + 2$ jets	0.0235
361070	$ll\nu\nu + 2$ jets	0.0374
361071	$lll\nu + 2$ jets	0.0366
361072	$lll + 2$ jets	0.0274
361073	$lll + 2$ gluon jets	0.0321
361077	$ll\nu\nu + 2$ gluon jets	0.7773
363355	$Z \rightarrow qq, Z \rightarrow \nu\nu$	4.346
363356	$Z \rightarrow qq, Z \rightarrow ll$	2.182
363357	$W \rightarrow qq, Z \rightarrow \nu\nu$	6.795
363358	$W \rightarrow qq, Z \rightarrow ll$	3.433
363359	$W^+ \rightarrow qq, W^- \rightarrow l\nu$	24.719
363360	$W^+ \rightarrow l\nu, W^- \rightarrow qq$	112.740
363489	$W \rightarrow l\nu, Z \rightarrow qq$	11.421
363490	$lll$	1.256
363491	$lll\nu$	4.588
363492	$ll\nu\nu$	12.465
363493	$l\nu\nu\nu$	3.227
363494	$\nu\nu\nu\nu$	0.604

Table A.7.: List of used diboson samples. Some samples are generated including all processes resulting in the same final states, whereas in other samples particular decays of electroweak vector bosons are simulated. The cross-sections are given for each sample separately as the processes can be very different.

Dataset ID	filter	$H_T$ - or $E_T^{\text{miss}}$ -slice [GeV]	cross-section [pb]
407009	$\geq 1$ lepton	600-1000 ( $H_T$ )	19.068
407010	$\geq 1$ lepton	1000-1500 ( $H_T$ )	2.666
407011	$\geq 1$ lepton	$> 1500$ ( $H_T$ )	0.470
407012	$\geq 1$ lepton	$> 200$ ( $E_T^{\text{miss}}$ )	67.670
407322	$\geq 1$ lepton	$> 300$ ( $E_T^{\text{miss}}$ )	1.162
407323	$\geq 1$ lepton	$> 400$ ( $E_T^{\text{miss}}$ )	0.261
410000	non all hadronic	-	452.694
410007	all hadronic	-	380.211

Table A.8.: List of used samples for top quark pair production. At least one lepton in the decay chain is required by the filter. Two different slicings have been used: One according to  $H_T$  and one according to  $E_T^{\text{miss}}$ . The cross-section is given for each sample separately (without filter efficiency).

Dataset ID	$t/\bar{t}$	channel	$H_T$ - or $E_T^{\text{miss}}$ -slice [GeV]	cross-section [pb]
407018	top	-	$> 500 (H_T)$	3.171
407019	top	-	$> 200 (E_T^{\text{miss}})$	0.404
407020	antitop	-	$> 500 (H_T)$	3.168
407021	antitop	-	$> 200 (E_T^{\text{miss}})$	0.403
410011	top	t-channel	-	44.152
410012	antitop	t-channel	-	26.276
410013	top	-	-	35.845
410014	antitop	-	-	35.824
410025	top	s-channel	-	2.061
410026	antitop	s-channel	-	1.289
410215	both	$tWZ$	-	0.016
410049	both	$tZ$ trilepton	-	0.009
410050	both	$tZ$ non all hadronic	-	0.240

Table A.9.: List of used samples for single top quark production. Apart from the samples representing the associated production of a single top quark with one or more electroweak vector boson the simulations for s- and t-channel as well as for the mentioned  $H_T$ - and  $E_T^{\text{miss}}$ -slices are separated into a sample for top and antitop quark production each.

Dataset ID	channel	cross-section [pb]
410066	$ttW$ with 0 partons	0.233
410067	$ttW$ with 1 parton	0.186
410068	$ttW$ with 2 partons	0.181
410073	$ttZ\nu\nu qq$ with 0 partons	0.255
410074	$ttZ\nu\nu qq$ with 1 parton	0.243
410075	$ttZ\nu\nu qq$ with 2 partons	0.254
410081	$ttWW$	0.010
410111	$ttee$ with 0 partons	0.013
410112	$ttee$ with 1 parton	0.022
410113	$tt\mu\mu$ with 0 partons	0.013
410114	$tt\mu\mu$ with 1 parton	0.022
410115	$tt\tau\tau$ with 0 partons	0.014
410116	$tt\tau\tau$ with 1 parton	0.022

Table A.10.: List of used samples for top quark pair production with associated vector boson radiation. The particular channels are given in absence of or in combination with a certain number of parton jets.



## B. Appendix II: N-1 Plots

### B.1. SR DS HadHad Low Mass

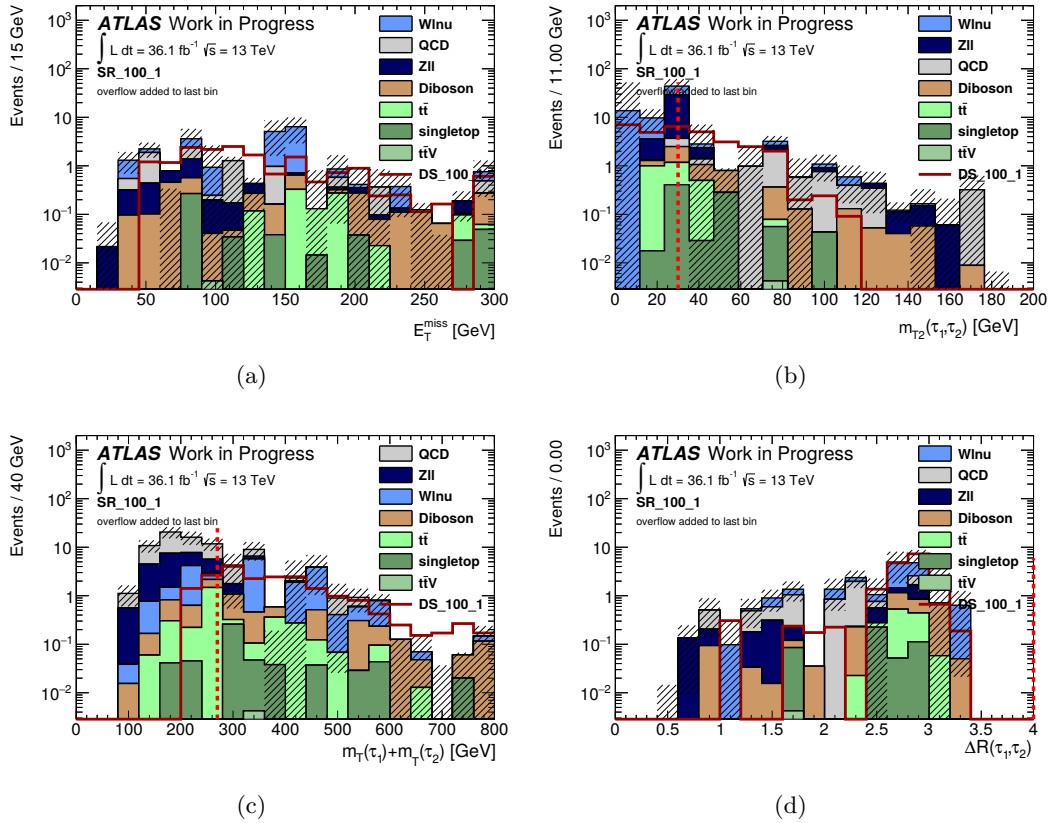


Figure B.1.: Distributions of the variables used for the signal region optimization for stau mass of 100 GeV and neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.

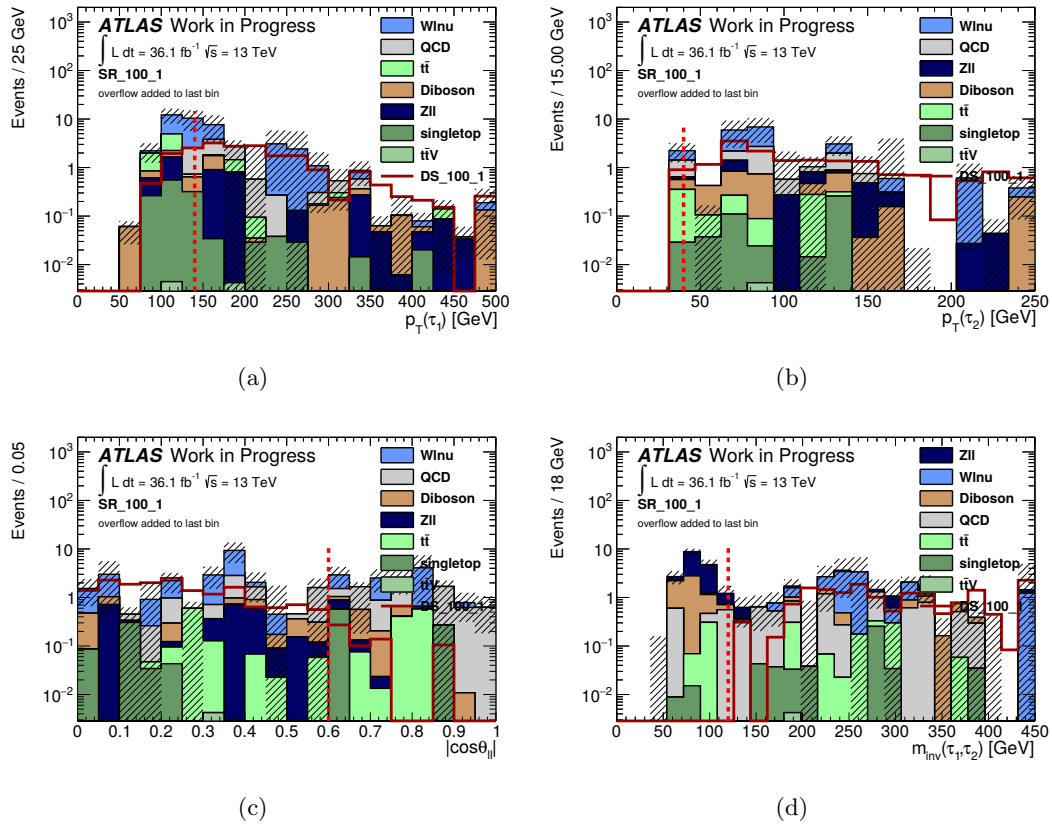


Figure B.2.: Distributions of the variables used for the signal region optimization for stau mass of 100 GeV and neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.

## B.2. SR DS HadHad Intermediate Mass

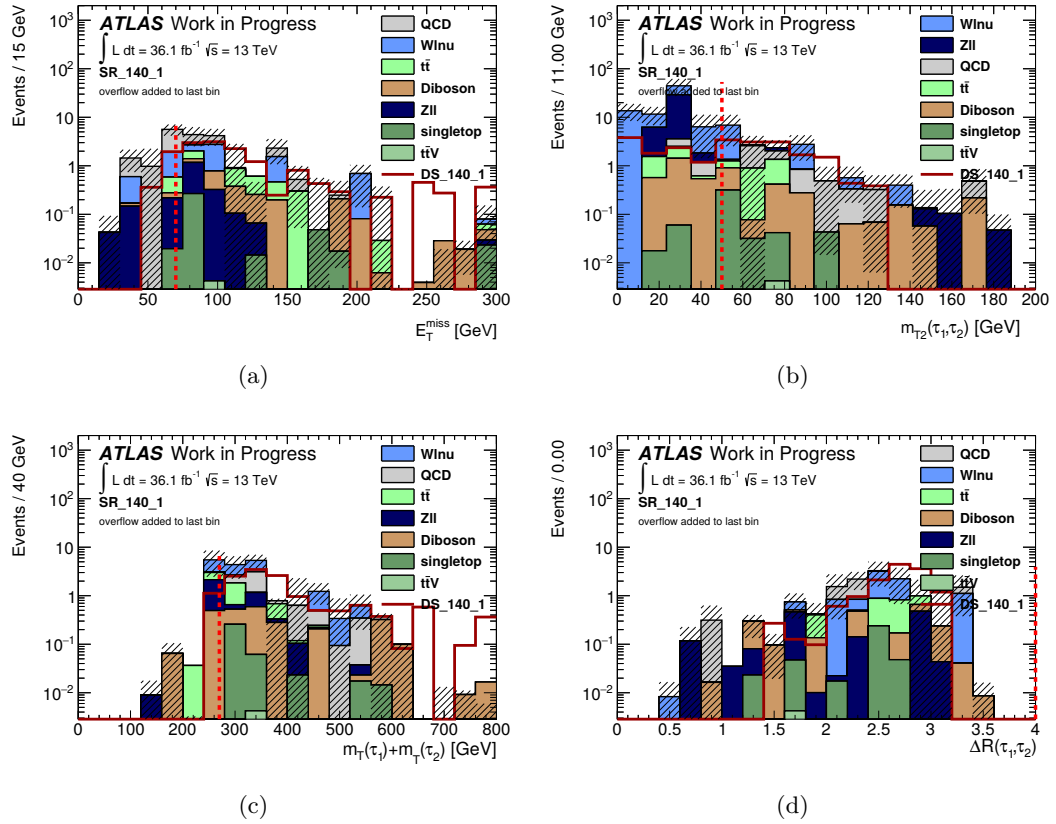


Figure B.3.: Distributions of the variables used for the signal region optimization for a stau mass 140 GeV and a neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.

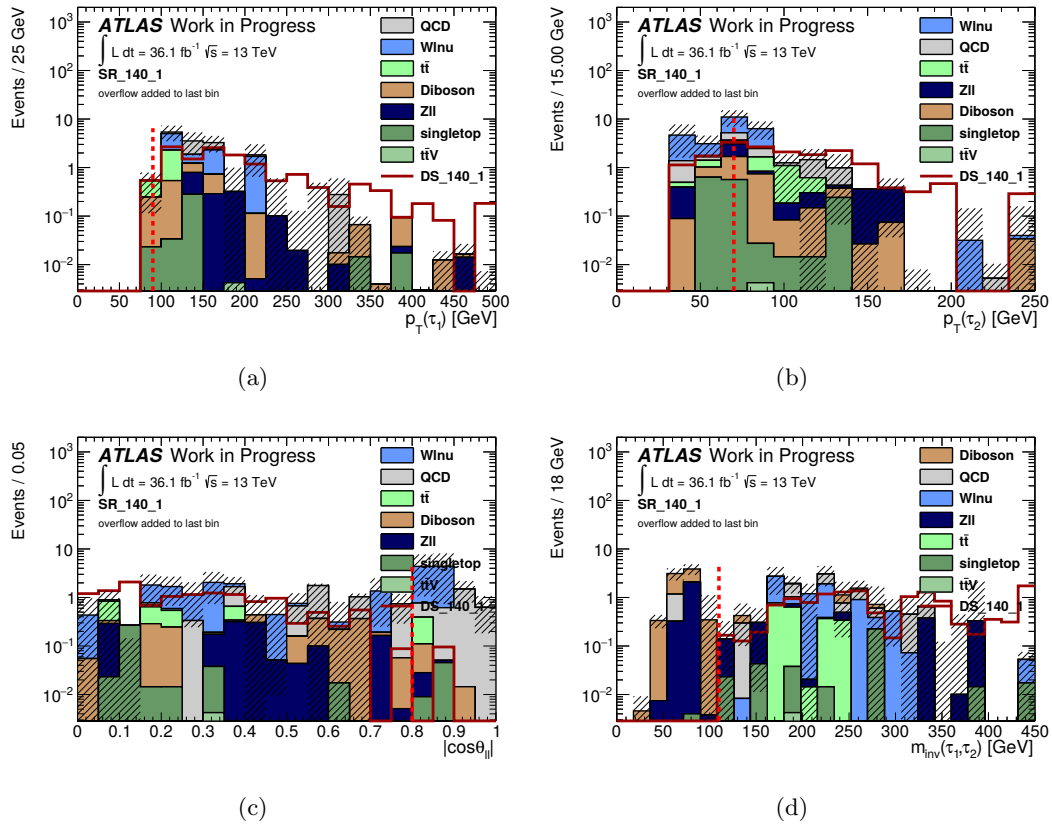


Figure B.4.: Distributions of the variables used for the signal region optimization for a stau mass 140 GeV and a neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.

## B.3. SR DS HadHad High Mass

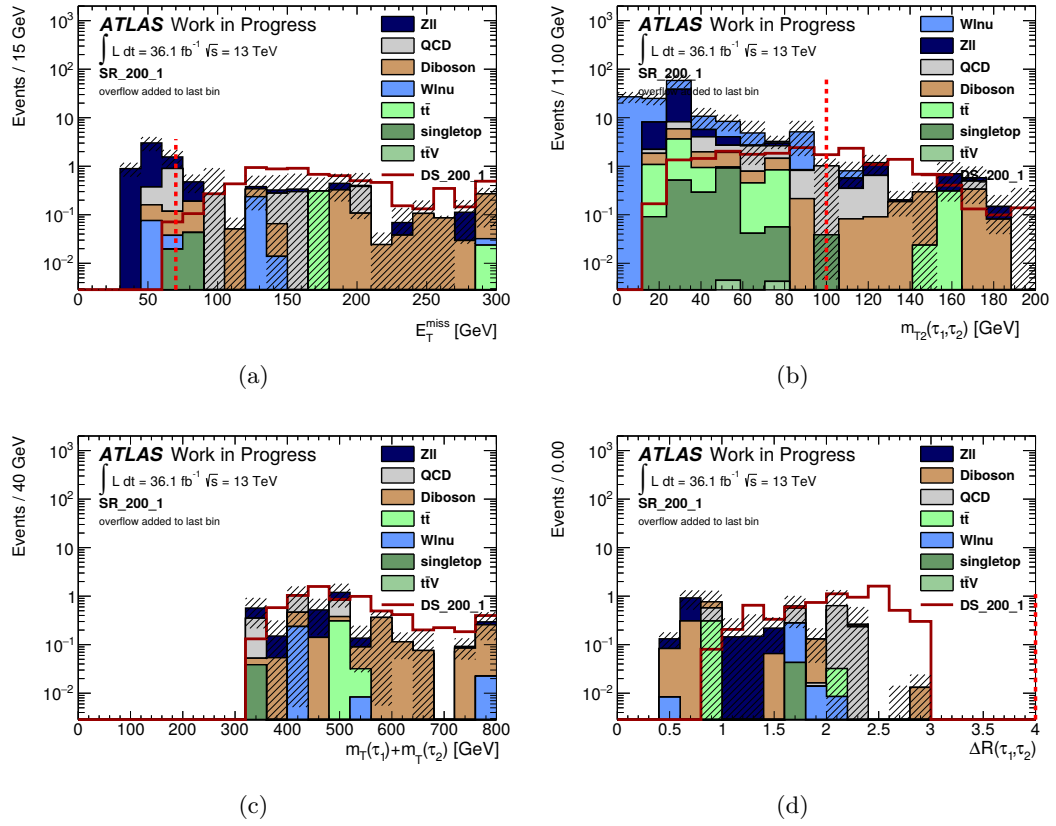


Figure B.5.: Distributions of the variables used for the signal region optimization for a stau mass 200 GeV and a neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.

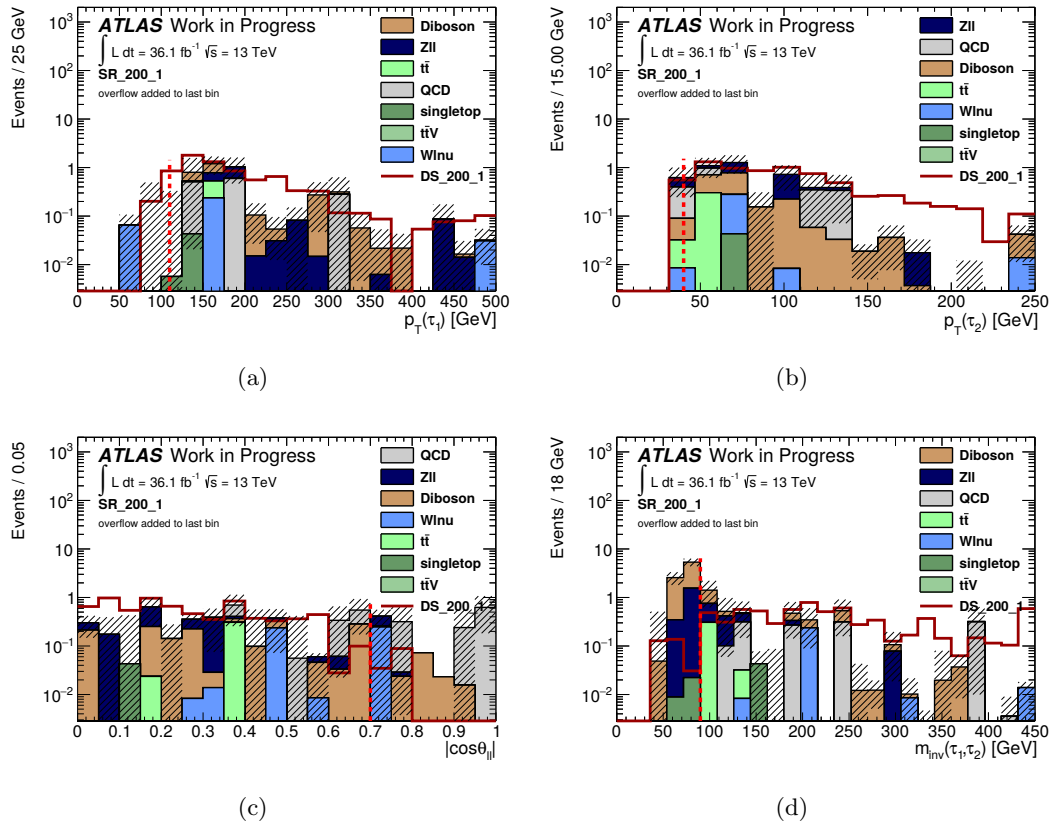


Figure B.6.: Distributions of the variables used for the signal region optimization for a stau mass 200 GeV and a neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.

## B.4. SR DS LepHad

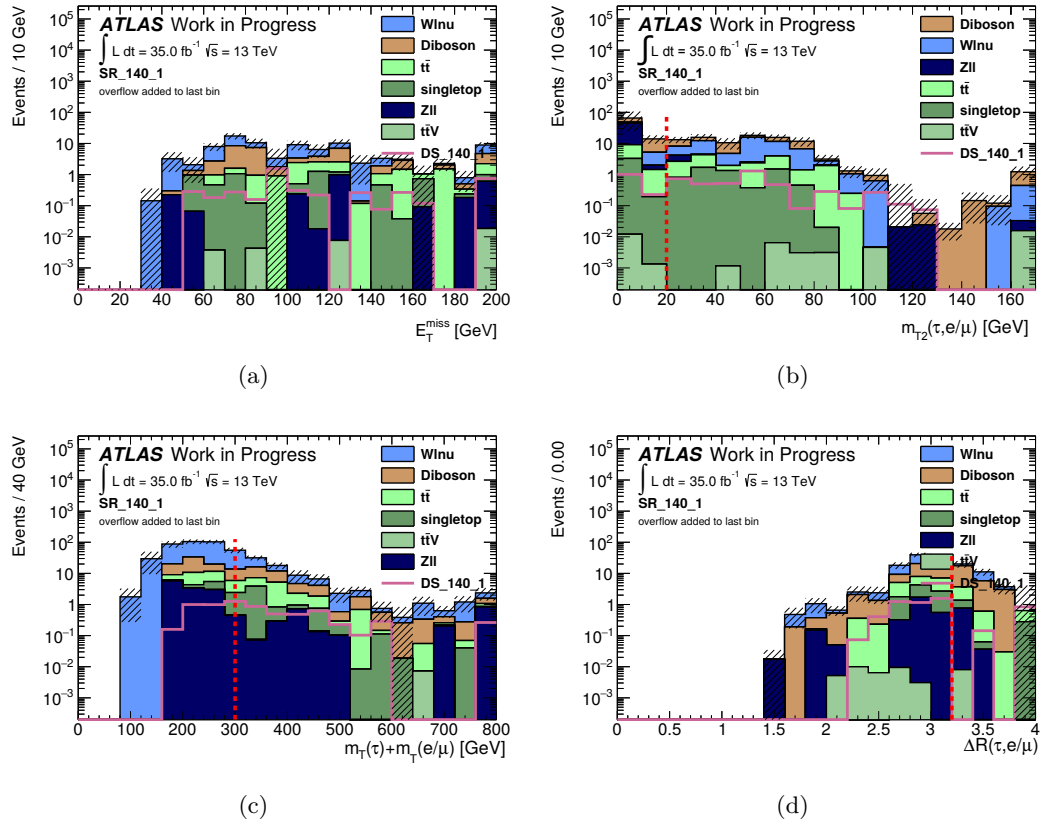


Figure B.7.: Distributions of the variables used for the signal region optimization for the LepHad-channel with a stau mass of 140 GeV and a neutralino mass of 0 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied. Only backgrounds estimated from MC simulations are taken into account.

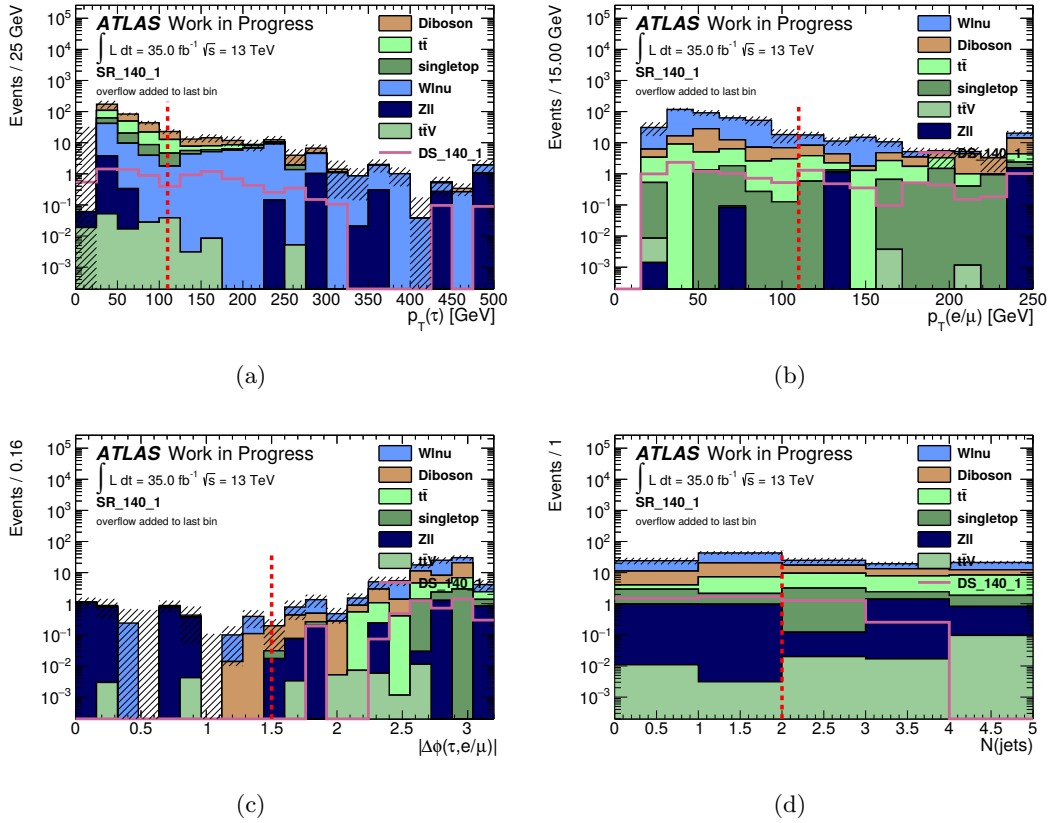
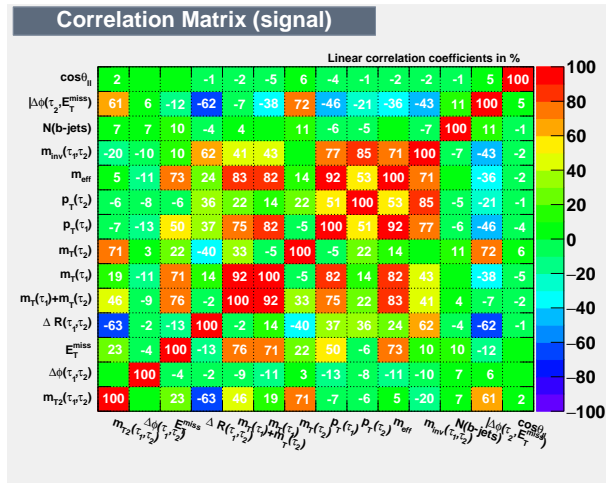


Figure B.8.: Distributions of the variables used for the signal region optimization for the LepHad-channel with a stau mass of 140 GeV and a neutralino mass of 0 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied. Only backgrounds estimated from MC simulations are taken into account.

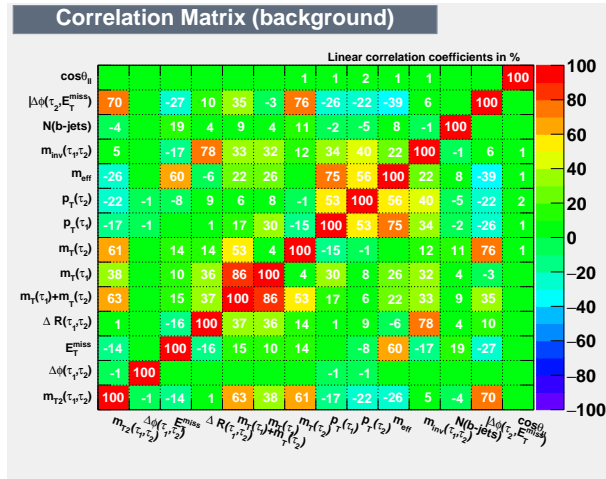


## C. Appendix III: Correlation Plots

### C.1. BDT trained on ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ , $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ )



(a)



(b)

Figure C.1.: The linear correlation factors between each pair of variables for a BDT trained on the signal mass point ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) are depicted here in the form of a matrix for signal (a) and for background (b). In general, the variables seem to be more correlated for the signal samples.

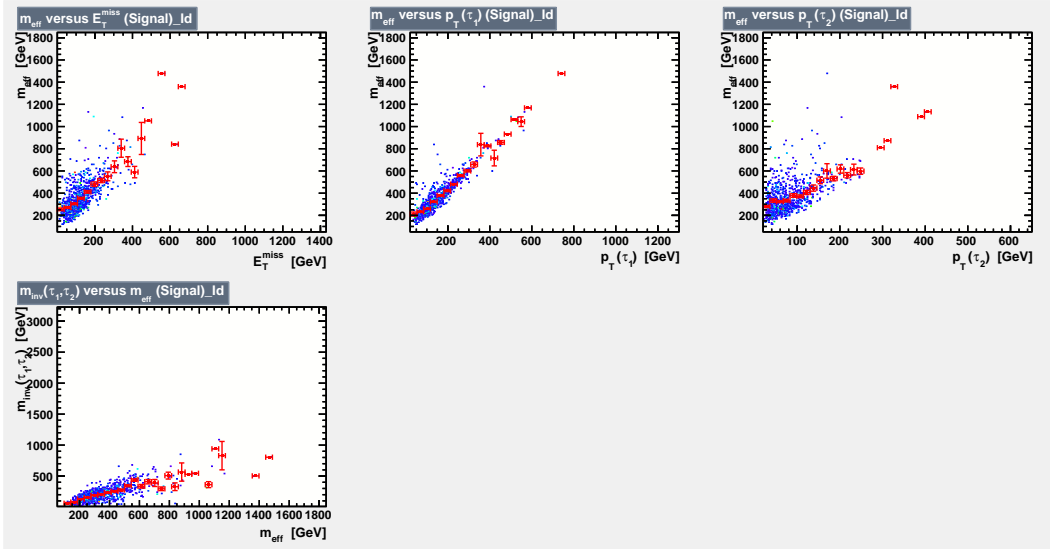


Figure C.2.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the signal samples before PCA transformation for a BDT trained on the mass point ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ). The four plotted variables show an approximately linear dependence on the effective mass.

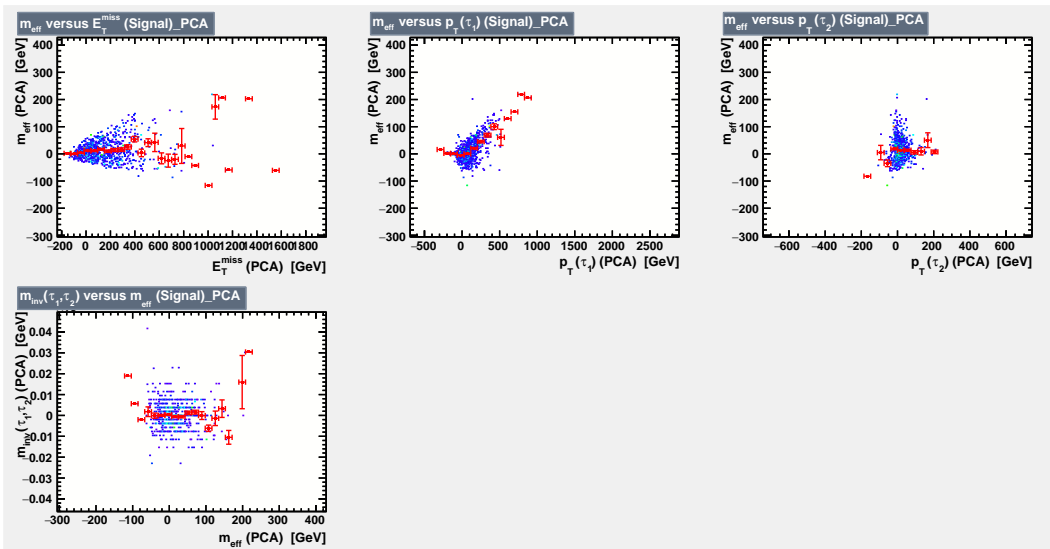


Figure C.3.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the signal samples after PCA transformation for a BDT trained on the mass point ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ).

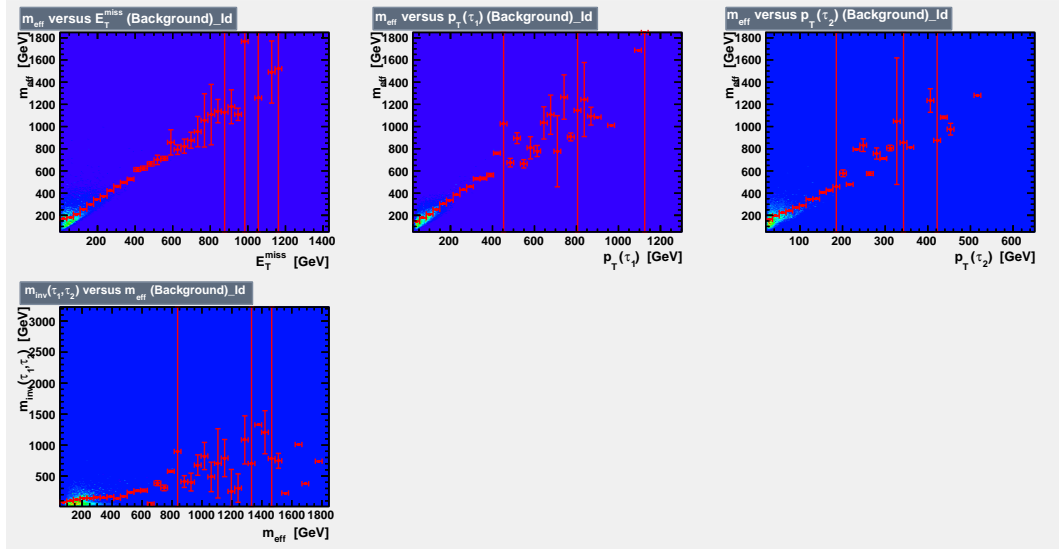


Figure C.4.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the background samples before PCA transformation for a BDT trained on the ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) mass point. The four plotted variables show an approximately linear dependence on the effective mass.

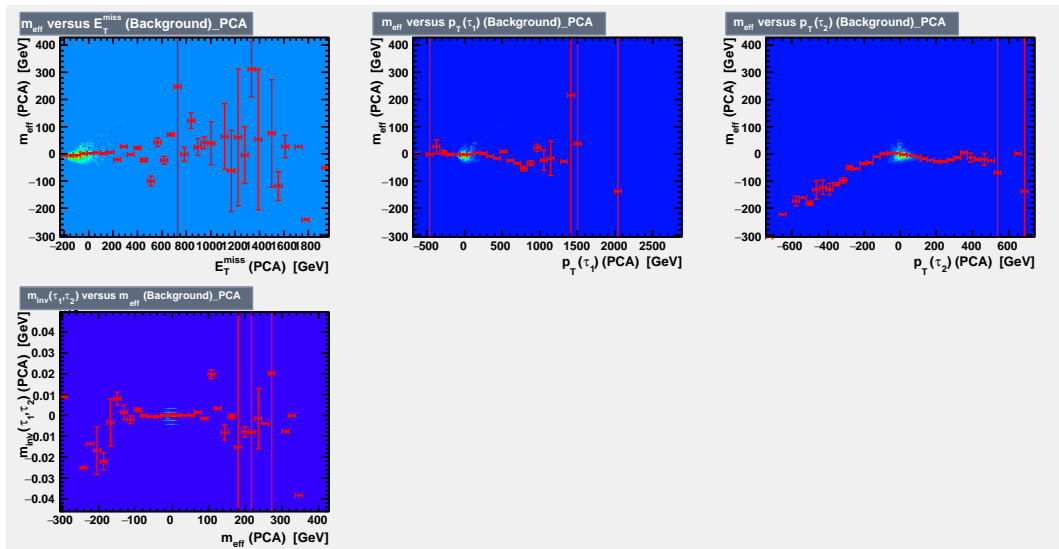


Figure C.5.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the background samples after PCA transformation for a BDT trained on the ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ ,  $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) mass point.

## C.2. Low Stau Mass BDT

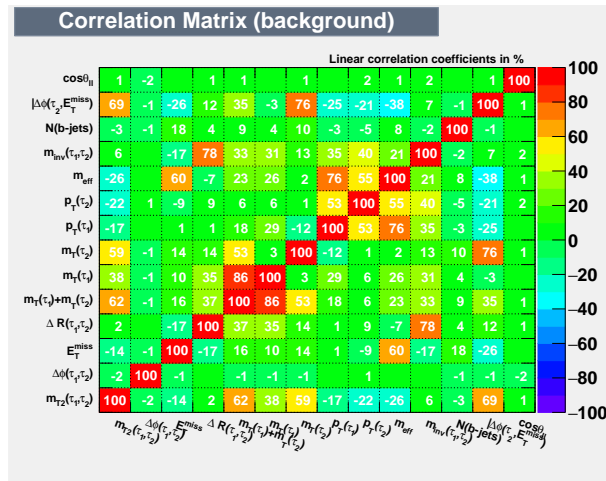
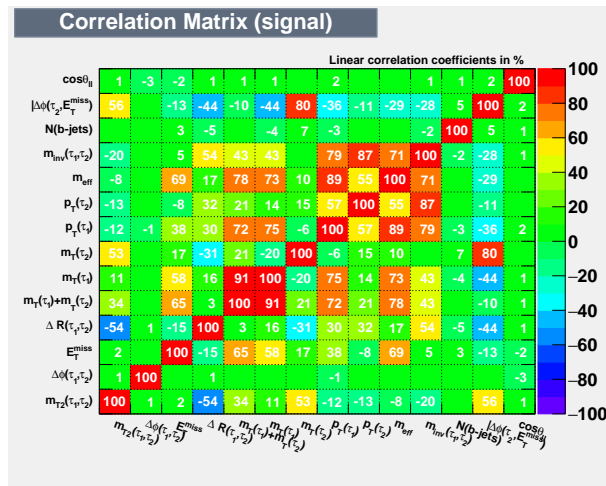


Figure C.6.: The linear correlation factors between each pair of variables for a BDT trained on low stau masses are depicted here in the form of a matrix for signal (a) and for background (b).

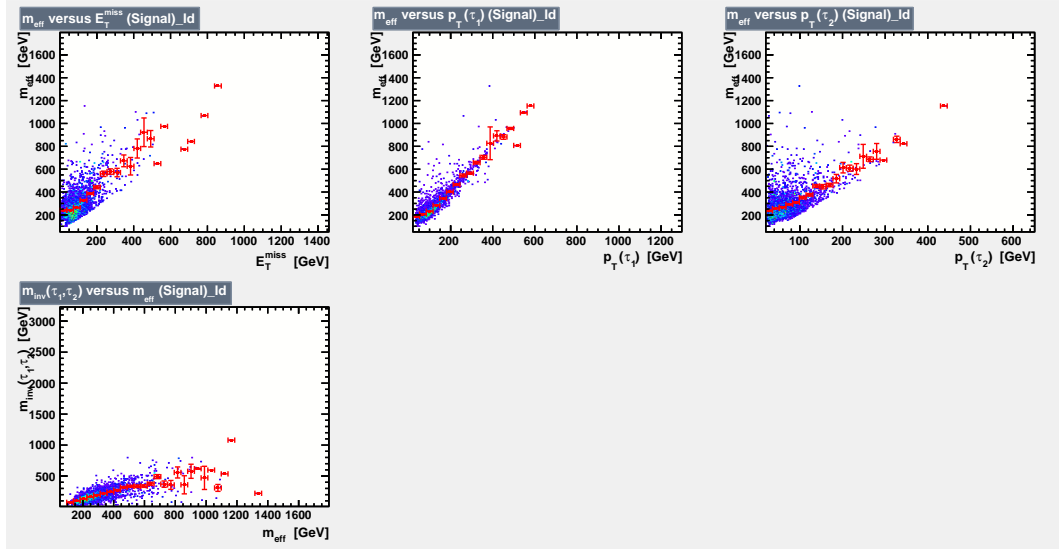


Figure C.7.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the signal samples before PCA transformation for a low-stau-mass BDT. The four plotted variables show an approximately linear dependence on the effective mass.

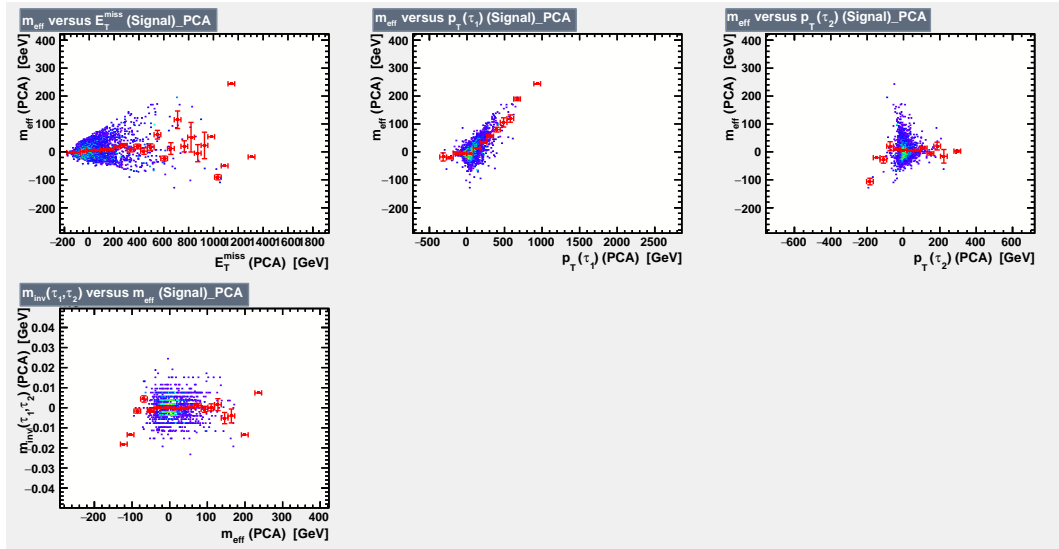


Figure C.8.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the signal samples after PCA transformation for a low-stau-mass BDT. The PCA transformation has notably smoothed the correlations although it is still visible for  $m_{\text{eff}}$  and  $p_T(\tau_1)$ .

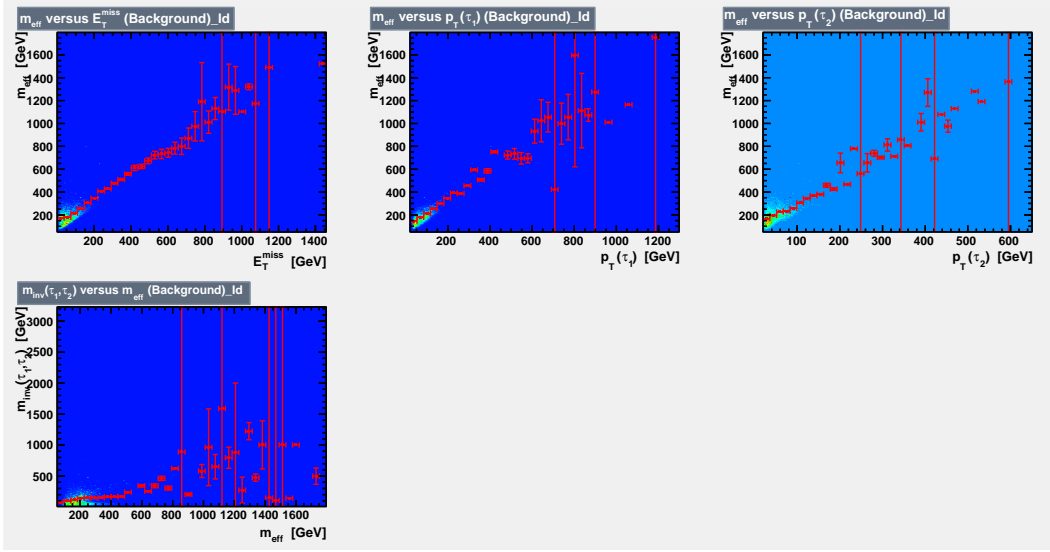


Figure C.9.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the background samples before PCA transformation for a low-stau-mass BDT. The linear dependence of the plotted variables to  $m_{\text{eff}}$  is clearly visible.

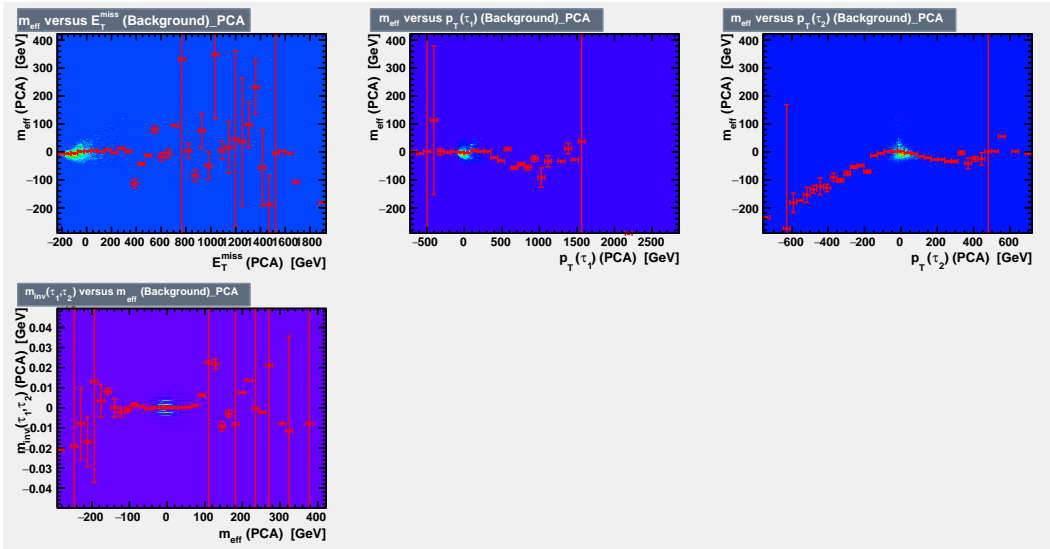
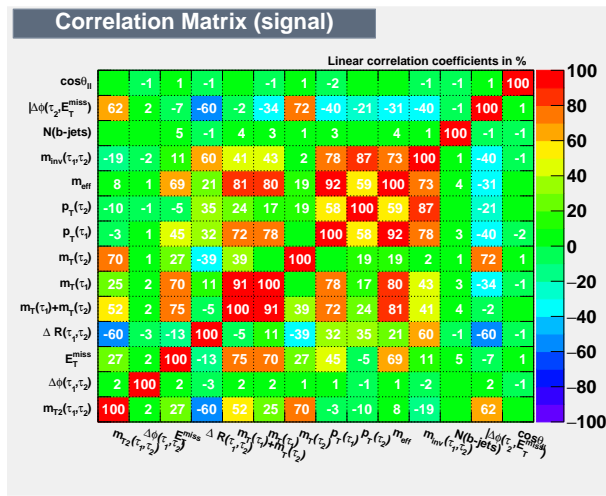
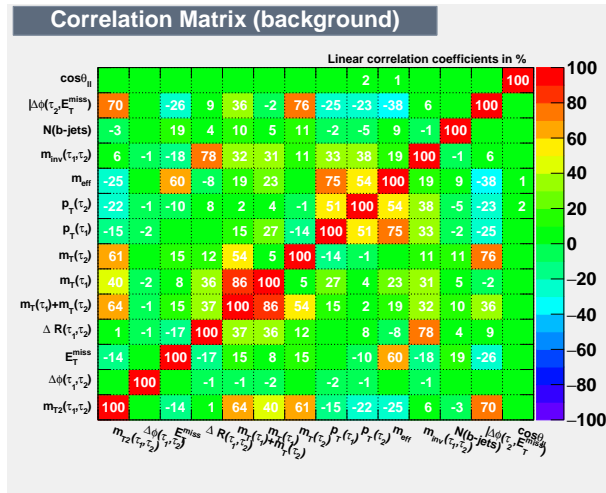


Figure C.10.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the background samples after PCA transformation for a low-stau-mass BDT. Due to the PCA transformation the dependences have been weakened compared to figure C.9.

### C.3. High Stau Mass BDT



(a)



(b)

Figure C.11.: The linear correlation factors between each pair of variables for a BDT trained on high stau masses are depicted here in the form of a matrix for signal (a) and for background (b).

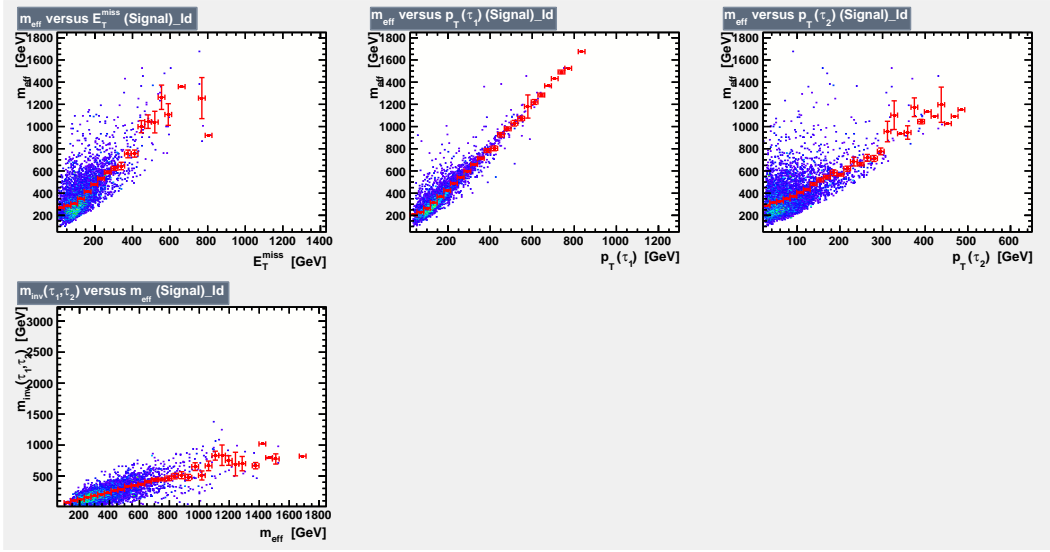


Figure C.12.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the signal samples before PCA transformation for a high-stau-mass BDT. For each of these variables an approximately linear dependence on the effective mass is observed.

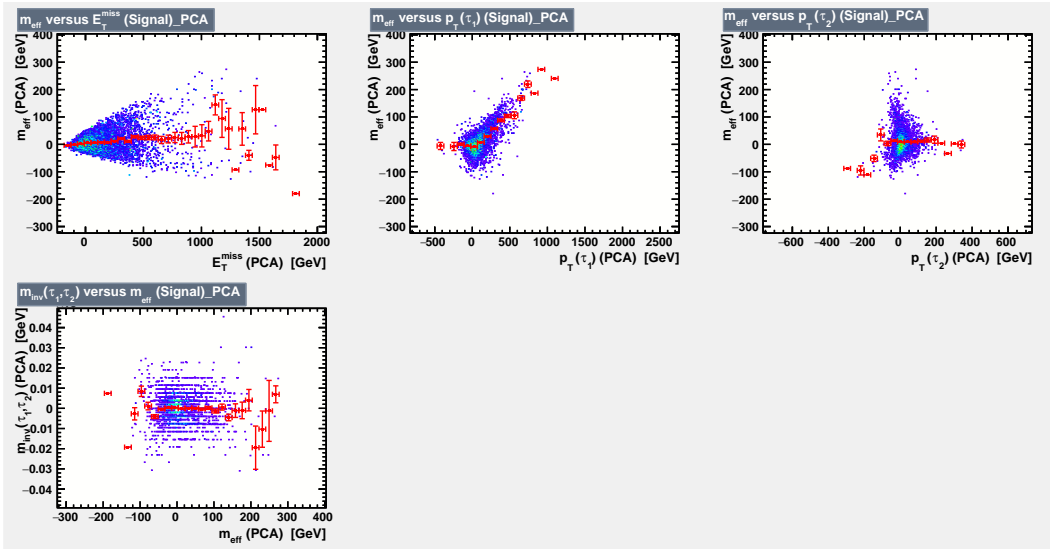


Figure C.13.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the signal samples after PCA transformation for a high-stau-mass BDT. Compared to figure C.12, the correlations to the effective mass are notably weakened. There is still a dependence to be observed between the  $m_{\text{eff}}$  and  $p_T(\tau_1)$ , although it is also weaker than it was before the transformation.



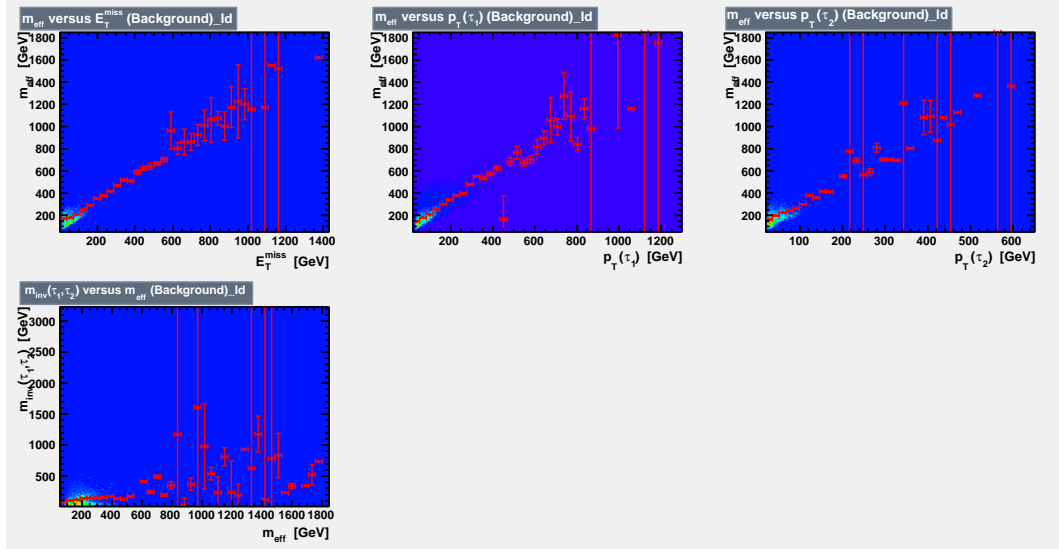


Figure C.14.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the background samples before PCA transformation for a high-stau-mass BDT. Similar to figure C.12, there is a visible (linear) dependence between the effective mass and the other variables.

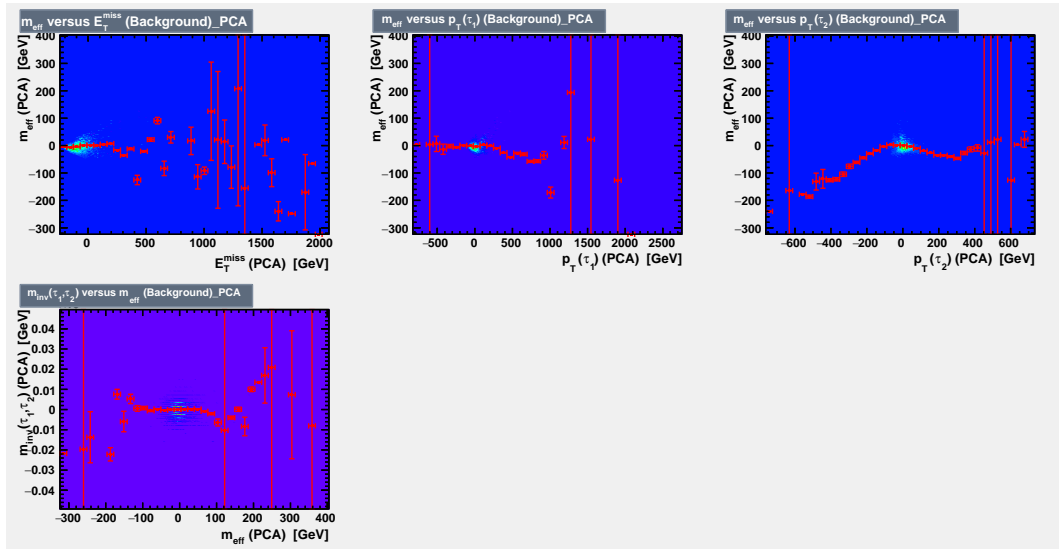


Figure C.15.: Correlation scatter plots for  $m_{\text{eff}}$ ,  $m_{\text{inv}}(\tau_1, \tau_2)$ ,  $p_T(\tau_1)$ ,  $p_T(\tau_2)$  and  $E_T^{\text{miss}}$  for the background samples before PCA transformation for a high-stau-mass BDT. The PCA transformation has flattened out most of the dependences between the effective mass and the other variables.



## D. Appendix IV: N-1 Plots with Tau Promotion

### D.1. SR DS HadHad High Mass with Tau Promotion in $W$ +jets and QCD

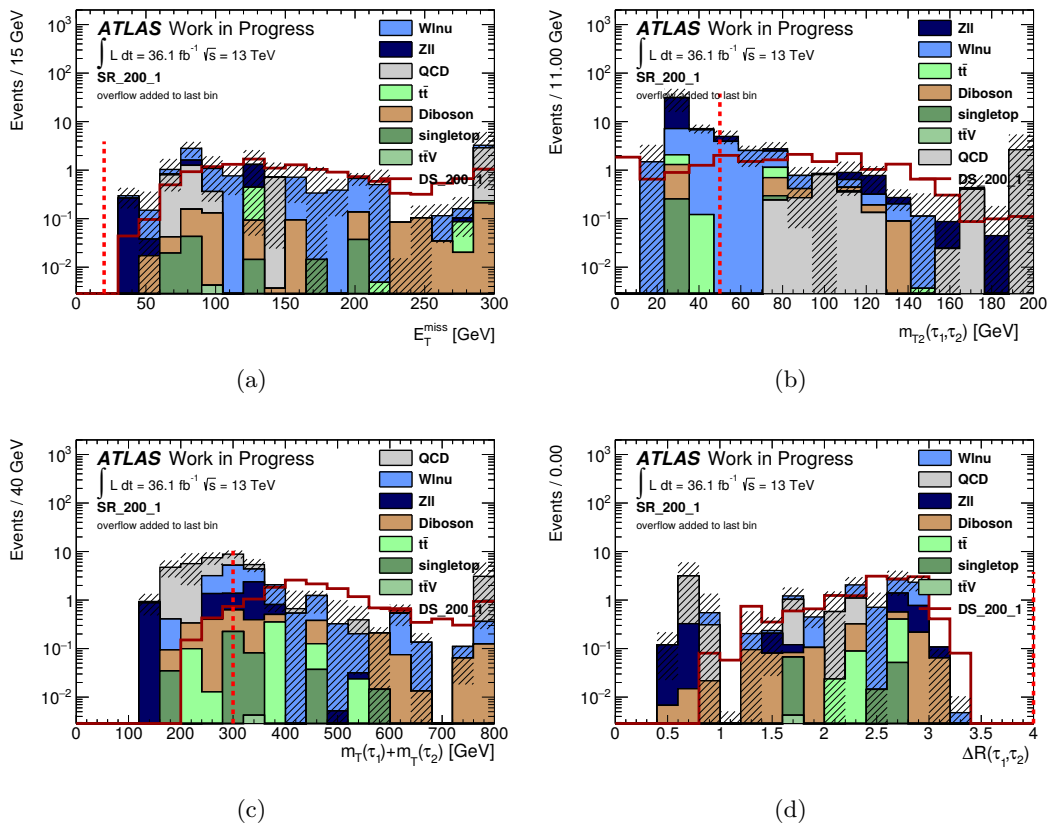


Figure D.1.: Distributions of the variables used for the signal region optimization for a stau mass 200 GeV and a neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.  $\tau$ -promotion has been included in the  $W$ +jets dataset as well as in the estimation of the QCD background.

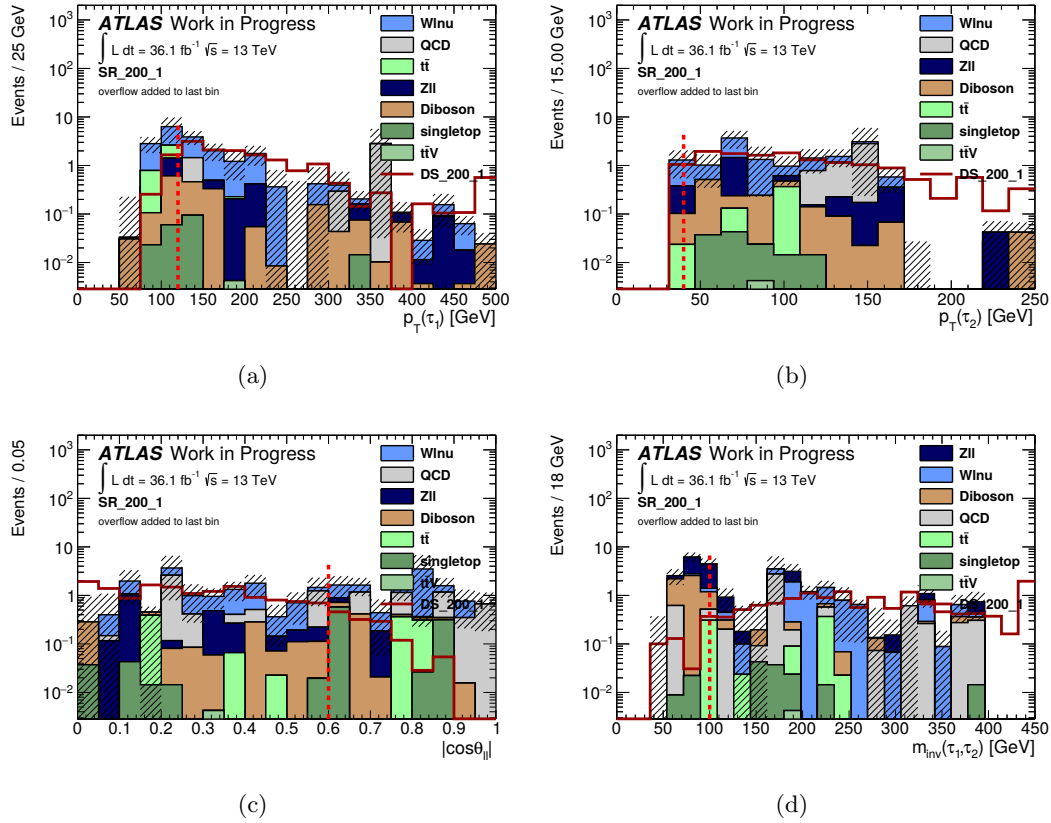


Figure D.2.: Distributions of the variables used for the signal region optimization for a stau mass 200 GeV and a neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.  $\tau$ -promotion has been included in the  $W$ +jets dataset as well as in the estimation of the QCD background.

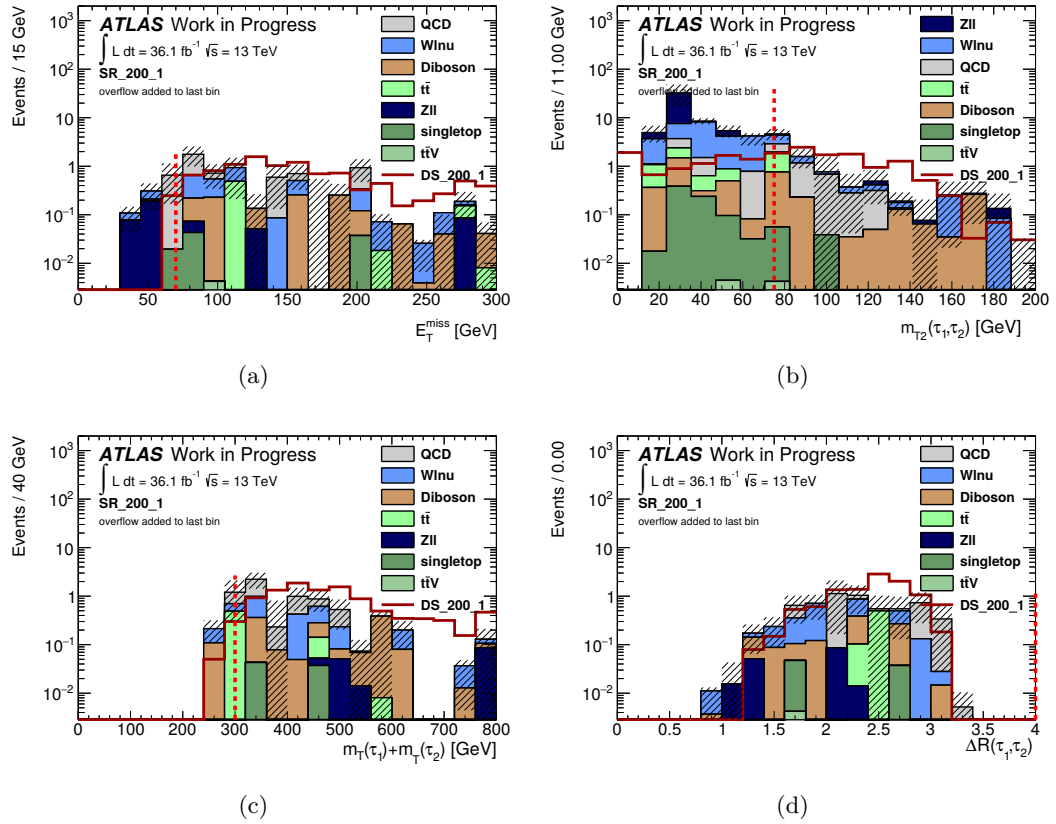
D.2. SR DS HadHad High Mass with Tau Promotion in  $W$ +jets

Figure D.3.: Distributions of the variables used for the signal region optimization for a stau mass 200 GeV and a neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.  $\tau$ -promotion has only been applied to the  $W$ +jets dataset.

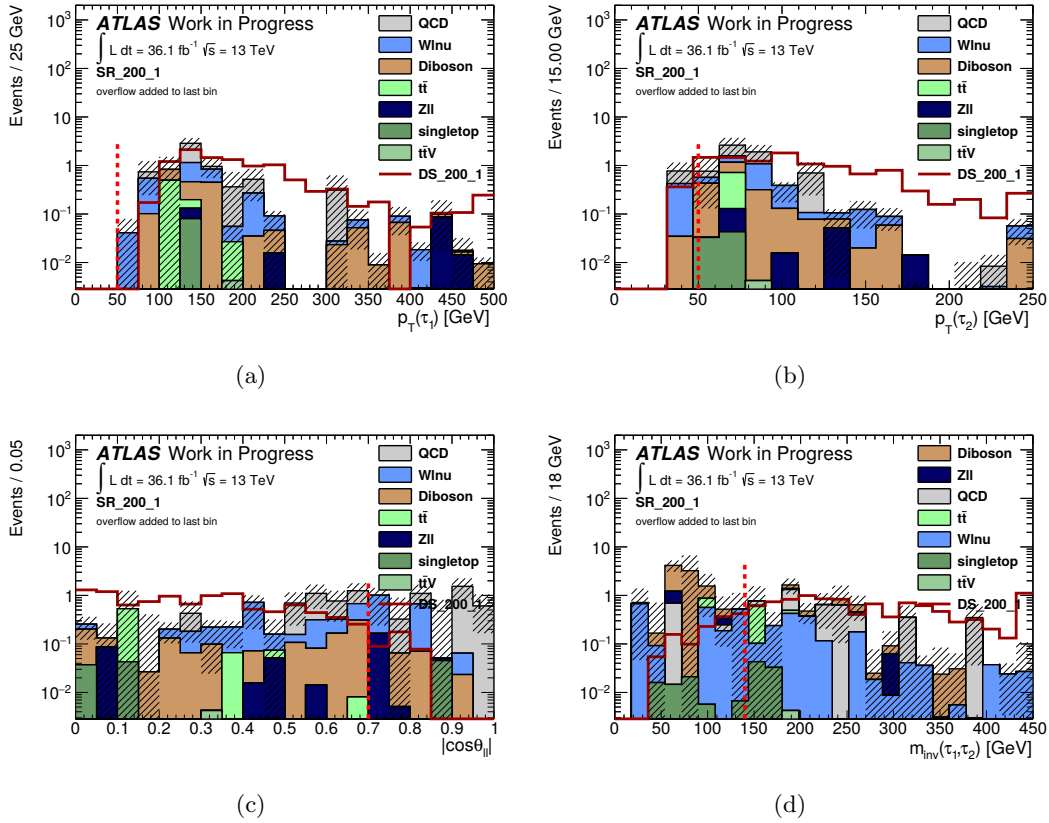


Figure D.4.: Distributions of the variables used for the signal region optimization for a stau mass 200 GeV and a neutralino mass 1 GeV, each with all cuts but the one on the variable itself or, if there is no cut placed on the variable, with the total number of cuts applied.  $\tau$ -promotion has only been applied to the  $W$ +jets dataset.

## Bibliography

- [1] Georges Aad et al. “Search for pair production of gluinos decaying via stop and sbottom in events with  $b$ -jets and large missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *Phys. Rev. D* 94.3 (2016), p. 032003. DOI: 10.1103/PhysRevD.94.032003. arXiv: 1605.09318 [hep-ex].
- [2] Morad Aaboud et al. “Search for squarks and gluinos in events with an isolated lepton, jets and missing transverse momentum at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: (2017). arXiv: 1708.08232 [hep-ex].
- [3] Morad Aaboud et al. “Search for the direct production of charginos and neutralinos in  $\sqrt{s} = 13$  TeV  $pp$  collisions with the ATLAS detector”. In: (2017). arXiv: 1708.07875 [hep-ex].
- [4] Georges Aad et al. “Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”. In: *JHEP* 05 (2014), p. 071. DOI: 10.1007/JHEP05(2014)071. arXiv: 1403.5294 [hep-ex].
- [5] Georges Aad et al. “Search for the direct production of charginos, neutralinos and staus in final states with at least two hadronically decaying taus and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”. In: *JHEP* 10 (2014), p. 096. DOI: 10.1007/JHEP10(2014)096. arXiv: 1407.0350 [hep-ex].
- [6] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013. ISBN: 9781107034266.
- [7] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. New York, NY: Wiley, 2008.
- [8] Stephen P. Martin. “A Supersymmetry primer”. In: (1997). [Adv. Ser. Direct. High Energy Phys.18,1(1998)]. DOI: 10.1142/9789812839657\_0001, 10.1142/9789814307505\_0001. arXiv: hep-ph/9709356 [hep-ph].
- [9] Katherine Garrett and Gintaras Duda. “Dark Matter: A Primer”. In: *Adv. Astron.* 2011 (2011), p. 968283. DOI: 10.1155/2011/968283. arXiv: 1006.2483 [hep-ph].
- [10] N.M.A. 2014 (2017). *The 2004 nobel prize in physics - popular information*. URL: [http://www.nobelprize.org/nobel\\_prizes/physics/laureates/2004/popular.html](http://www.nobelprize.org/nobel_prizes/physics/laureates/2004/popular.html).
- [11] Tianyou Xie. *Supersymmetry And Gauge Couplings Correction*. 2016. URL: [https://www.researchgate.net/publication/311843425\\_Supersymmetry\\_And\\_Gauge\\_Couplings\\_Correction](https://www.researchgate.net/publication/311843425_Supersymmetry_And_Gauge_Couplings_Correction).
- [12] A. Djouadi et al. “The Minimal supersymmetric standard model: Group summary report”. In: *GDR (Groupement De Recherche) - Supersymetrie Montpellier, France, April 15-17, 1998*. 1998. arXiv: hep-ph/9901246 [hep-ph]. URL: [https://inspirehep.net/record/481987/files/arXiv:hep-ph\\_9901246.pdf](https://inspirehep.net/record/481987/files/arXiv:hep-ph_9901246.pdf).

- [13] N.M.A. 2016 (2017). *SUSY Feynman Diagrams*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SUSYFeynmanDiagrams>.
- [14] Melissa van Beekveld et al. “Supersymmetry with Dark Matter is still natural”. In: *Phys. Rev. D* 96.3 (2017), p. 035015. DOI: 10.1103/PhysRevD.96.035015. arXiv: 1612.06333 [hep-ph].
- [15] *Search for pair production of tau sleptons in  $\sqrt{s} = 13$  TeV pp collisions in the all-hadronic final state*. Tech. rep. CMS-PAS-SUS-17-003. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2273395>.
- [16] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.
- [17] N.M.A. (2017). *LHC Facts*. URL: <http://www.lhc-facts.ch/>.
- [18] Joao Pequenao 2015 (2017). *CERN Document Server*. URL: <https://cds.cern.ch/images/CERN-GE-0803012-01>.
- [19] G. Aad et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.
- [20] Yu Nakahama. “The ATLAS Trigger System: Ready for Run-2”. In: *J. Phys. Conf. Ser.* 664.8 (2015), p. 082037. DOI: 10.1088/1742-6596/664/8/082037.
- [21] N.M.A. 2014 (2017). *ATLAS Data Summary*. URL: <https://atlas.web.cern.ch/Atlas/GROUPS/DATAPREPARATION/DataSummary/2016/>.
- [22] Torbjorn Sjostrand. “Monte Carlo Generators”. In: *High-energy physics. Proceedings, European School, Aronsborg, Sweden, June 18-July 1, 2006*. 2006, pp. 51–74. arXiv: hep-ph/0611247 [hep-ph]. URL: <http://weplib.cern.ch/abstract?CERN-LCGAPP-2006-06>.
- [23] M. A. Dobbs et al. “Les Houches guidebook to Monte Carlo generators for hadron collider physics”. In: *Physics at TeV colliders. Proceedings, Workshop, Les Houches, France, May 26-June 3, 2003*. 2004, pp. 411–459. arXiv: hep-ph/0403045 [hep-ph]. URL: [http://lss.fnal.gov/cgi-bin/find\\_paper.pl?conf-04-183](http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-04-183).
- [24] A. Rimoldi and A. Dell’Acqua. “The Full detector simulation for the ATLAS experiment: Status and outlook”. In: *eConf* C0303241 (2003), TUMT001. arXiv: physics/0306086 [physics].
- [25] The ATLAS collaboration. “Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data”. In: (2016).
- [26] *Electron identification measurements in ATLAS using  $\sqrt{s} = 13$  TeV data with 50 ns bunch spacing*. Tech. rep. ATL-PHYS-PUB-2015-041. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2048202>.
- [27] Georges Aad et al. “Muon reconstruction performance of the ATLAS detector in proton-proton collision data at  $\sqrt{s} = 13$  TeV”. In: *Eur. Phys. J. C* 76.5 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].
- [28] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The Anti-k(t) jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].
- [29] A. Hrynevich. “ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2”. In: *JINST* 12.06 (2017), p. C06038. DOI: 10.1088/1748-0221/12/06/C06038.



- [30] The ATLAS collaboration. *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at  $\sqrt{s} = 13$  TeV*. Tech. rep. ATLAS-CONF-2017-029. 2017.
- [31] *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*. Tech. rep. ATL-PHYS-PUB-2015-045. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2064383>.
- [32] A. J. Barr. “Measuring slepton spin at the LHC”. In: *JHEP* 02 (2006), p. 042. DOI: 10.1088/1126-6708/2006/02/042. arXiv: hep-ph/0511115 [hep-ph].
- [33] C. G. Lester and D. J. Summers. “Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders”. In: *Phys. Lett.* B463 (1999), pp. 99–103. DOI: 10.1016/S0370-2693(99)00945-4. arXiv: hep-ph/9906349 [hep-ph].
- [34] ATLAS Collaboration. “A search for high-mass resonances decaying to  $\tau^+\tau^-$  in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”. In: *JHEP* 07.CERN-PH-EP-2015-018. CERN-PH-EP-2015-018 (2015). 26 pages plus author list + cover page (50 pages total), 9 figures, 4 tables, submitted to JHEP, All figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/EXOT-2014-05/>, 157. 49 p. URL: <http://cds.cern.ch/record/1994408>.
- [35] S Owen. *Data-driven estimation of the QCD multijet background to SUSY searches with jets and missing transverse momentum at ATLAS using jet smearing*. Tech. rep. ATL-PHYS-INT-2012-008. Geneva: CERN, 2012. URL: <https://cds.cern.ch/record/1423310>.
- [36] Lorenzo Moneta et al. “The RooStats Project”. In: *PoS ACAT2010* (2010), p. 057. arXiv: 1009.1003 [physics.data-an].
- [37] Yoav Freund and Robert E Schapire. “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”. In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139. ISSN: 0022-0000. DOI: <http://dx.doi.org/10.1006/jcss.1997.1504>. URL: <http://www.sciencedirect.com/science/article/pii/S002200009791504X>.
- [38] A. Hoecker et al. “TMVA - Toolkit for Multivariate Data Analysis”. In: *ArXiv Physics e-prints* (Mar. 2007). eprint: physics/0703039.
- [39] Robert E. Schapire. *Explaining AdaBoost*. Jan. 2013, pp. 37–52. ISBN: 978-3-642-41135-9.
- [40] Leo Breiman. “Prediction Games and Arcing Algorithms”. In: 11 (Nov. 1999), pp. 1493–517.
- [41] Jerome Friedman. “Greedy Function Approximation: A Gradient Boosting Machine”. In: 29 (Oct. 2001), pp. 1189–1232.
- [42] A. Mayr et al. “The Evolution of Boosting Algorithms - From Machine Learning to Statistical Modelling”. In: *ArXiv e-prints* (Mar. 2014). arXiv: 1403.1452 [stat.ME].
- [43] J. Shlens. “A Tutorial on Principal Component Analysis”. In: *ArXiv e-prints* (Apr. 2014). arXiv: 1404.1100 [cs.LG].

- 
- [44] A. N. Kolmogorov. “Sulla Determinazione Empirica di una Legge di Distribuzione”. In: *Giornale dell’Istituto Italiano degli Attuari* 4 (1933), pp. 83–91.
- [45] Pierre Baldi et al. “Parameterized neural networks for high-energy physics”. In: *Eur. Phys. J. C* 76.5 (2016), p. 235. DOI: 10.1140/epjc/s10052-016-4099-4. arXiv: 1601.07913 [hep-ex].

# List of Figures

2.1.	Unification of fundamental forces . . . . .	8
2.2.	Loop corrections to the Higgs mass . . . . .	9
2.3.	Feynman diagram of the direct stau pair production process . . . . .	12
2.4.	Upper limits on the cross-sections for direct stau production in run 1 . . . . .	14
3.1.	The large hadron collider system . . . . .	16
3.2.	The ATLAS detector . . . . .	17
4.1.	Total delivered and recorded luminosity in 2016 . . . . .	22
6.1.	Available signal mass points for the direct stau process in the plane of stau and neutralino masses . . . . .	33
6.2.	Schematic principle of the ABCD method . . . . .	37
6.3.	Correlation of $m_{T2}$ and the $\tau$ -charge sign and quality . . . . .	38
6.4.	Comparison of real and estimated QCD multijet contribution in validation region F . . . . .	39
6.5.	Background distributions in regions A, E and C . . . . .	40
6.6.	Background distributions in region D, including the QCD estimate . . . . .	41
7.1.	Comparison of signal and background distributions in the HadHad- and LepHad-channels I . . . . .	47
7.2.	Comparison of signal and background distributions in the HadHad- and LepHad-channels II . . . . .	48
7.3.	Transverse momenta of a light lepton in the LepHad-channel and of hadronic $\tau$ -leptons in the HadHad-channel . . . . .	49
8.1.	Significances for SR DS HadHad Low Mass . . . . .	55
8.2.	Significances for SR DS HadHad Intermediate Mass . . . . .	56
8.3.	Significances for SR DS HadHad High Mass . . . . .	57
8.4.	Significances for SR DS LepHad . . . . .	60
8.5.	Schematic principle of a decision tree . . . . .	62
8.6.	BDT misidentification rate for training and testing with increasing number of decision trees . . . . .	64
8.7.	Response of BDT trained on single mass point . . . . .	68
8.8.	Response of BDT trained on a single high statistics mass point . . . . .	70
8.9.	Evaluation of a BDT trained with a high statistics signal sample of the mass point ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ , $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) . . . . .	74
8.10.	Significances for a cut on the BDT response at 0.58 for all signal mass points, with a BDT trained on a high statistics dataset of the mass point . . . . .	75
8.11.	Comparison of responses for a BDT trained on eight signal samples with and without parameterization through the stau mass . . . . .	76

8.12. Comparison of the $m_{T2}$ distribution for three signal mass points and all background samples . . . . .	78
8.13. Response for a BDT trained on low stau masses . . . . .	79
8.14. Evaluation of a BDT trained on low stau masses . . . . .	81
8.15. Significances for a cut on the response of a BDT trained on low stau masses . . . . .	82
8.16. Response of a BDT trained on high stau masses . . . . .	82
8.17. Evaluation of a BDT trained on high stau masses . . . . .	84
8.18. Significances for a cut on a BDT trained on high stau masses . . . . .	85
9.1. Schematic principle of the $\tau$ -promotion method . . . . .	89
9.2. Comparison of distributions for $W$ +jets before and after promoting one container $\tau$ -lepton to medium ID . . . . .	93
9.3. Original binning for the fake efficiency measurement in the $p_T(\tau)$ - $E_T^{\text{miss}}$ -plane . . . . .	94
9.4. Modified binning for the fake efficiency measurement in the $p_T(\tau)$ - $E_T^{\text{miss}}$ -plane . . . . .	94
9.5. Comparison of the distributions for $W$ +jets before and after promoting one container $\tau$ -lepton to medium ID with modified $p_T(\tau)$ -bins for the fake efficiency measurement . . . . .	95
9.6. Comparison of distributions for $W$ +jets before and after promoting one container $\tau$ -lepton to loose ID . . . . .	98
9.7. Significances in SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets and QCD datasets . . . . .	102
9.8. Significances in SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets dataset . . . . .	104
9.9. Evaluation of a BDT trained on high stau masses with $\tau$ -promotion in $W$ +jets and QCD . . . . .	106
9.10. Significances for a cut on the response of a BDT trained on high stau masses evaluated with $\tau$ -promoted $W$ +jets and QCD samples . . . . .	107
9.11. Evaluation of a BDT trained on high stau masses with $\tau$ -promotion in $W$ +jets . . . . .	108
9.12. Significances for a cut at 0.9 on the response of a BDT trained on high stau masses evaluated with a $\tau$ -promoted $W$ +jets sample . . . . .	109
9.13. Significances for a cut at 0.88 on the response of a BDT trained on high stau masses evaluated with a $\tau$ -promoted $W$ +jets sample . . . . .	109
B.1. N-1 plots for SR DS HadHad Low Mass I . . . . .	123
B.2. N-1 plots for SR DS HadHad Low Mass II . . . . .	124
B.3. N-1 plots for SR DS HadHad Intermediate Mass I . . . . .	125
B.4. N-1 plots for SR DS HadHad Intermediate Mass II . . . . .	126
B.5. N-1 plots for SR DS HadHad High Mass I . . . . .	127
B.6. N-1 plots for SR DS HadHad High Mass II . . . . .	128
B.7. N-1 plots for SR DS LepHad I . . . . .	129
B.8. N-1 plots for SR DS LepHad II . . . . .	130
C.1. Linear correlation factors for a BDT trained on the mass point ( $m_{\tilde{\tau}} = 200$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV) . . . . .	131
C.2. Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the signal samples before PCA transformation for a BDT trained on the ( $m_{\tilde{\tau}} = 200$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV) mass point . . . . .	132

C.3.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the signal samples after PCA transformation for a BDT trained on the ( $m_{\tilde{\tau}} = 200$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV) mass point . . . . .	132
C.4.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the background samples before PCA transformation for a BDT trained on the ( $m_{\tilde{\tau}} = 200$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV) mass point . . . . .	133
C.5.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the background samples after PCA transformation for a BDT trained on the ( $m_{\tilde{\tau}} = 200$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV) mass point . . . . .	133
C.6.	Linear correlation factors for a low-stau-mass BDT . . . . .	134
C.7.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the signal samples before PCA transformation for a low-stau-mass BDT . . . . .	135
C.8.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the signal samples after PCA transformation for a low-stau-mass BDT . . . . .	135
C.9.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the background samples before PCA transformation for a low-stau-mass BDT . . . . .	136
C.10.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the background samples after PCA transformation for a low-stau-mass BDT . . . . .	136
C.11.	Linear correlation factors for a high-stau-mass BDT . . . . .	137
C.12.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the signal samples before PCA transformation for a high-stau-mass BDT . . . . .	138
C.13.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the signal samples after PCA transformation for a high-stau-mass BDT . . . . .	138
C.14.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the background samples before PCA transformation for a high-stau-mass BDT . . . . .	139
C.15.	Correlation scatter plots for $m_{\text{eff}}$ , $m_{\text{inv}}(\tau_1, \tau_2)$ , $p_T(\tau_1)$ , $p_T(\tau_2)$ and $E_T^{\text{miss}}$ for the background samples after PCA transformation for a high-stau-mass BDT . . . . .	139
D.1.	N-1 plots for SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets and QCD datasets I . . . . .	141
D.2.	N-1 plots for SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets and QCD datasets II . . . . .	142
D.3.	N-1 plots for SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets dataset I . . . . .	143
D.4.	N-1 plots for SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets dataset II . . . . .	144



# List of Tables

2.1.	Standard Model lepton charges and masses . . . . .	3
2.2.	Standard Model quark charges and masses . . . . .	4
2.3.	Chiral and gauge supermultiplets in the MSSM . . . . .	10
6.1.	Yields in each region used for the estimation of the QCD multijet background	37
7.1.	Preselection cutflow for the HadHad-channel . . . . .	45
7.2.	Preselection cutflow for the LepHad-channel . . . . .	46
8.1.	Cut-and-count: List of cuts (HadHad-channel) . . . . .	53
8.2.	Cut combination SR DS HadHad Low Mass . . . . .	53
8.3.	Event yields in SR DS HadHad Low Mass . . . . .	54
8.4.	Cut combination SR DS HadHad Intermediate Mass . . . . .	54
8.5.	Event Yields SR DS HadHad Intermediate Mass . . . . .	55
8.6.	Cut combination for SR DS HadHad High Mass . . . . .	56
8.7.	Event yields in SR DS HadHad High Mass . . . . .	57
8.8.	Cut-and-count: List of cuts (LepHad-channel) . . . . .	58
8.9.	Cut combination SR DS LepHad . . . . .	59
8.10.	Event yields in SR DS LepHad . . . . .	59
8.11.	BDT training parameters . . . . .	67
8.12.	Ranking of input variables for a BDT trained on a single mass point . . . . .	69
8.13.	Overtraining check for a BDT trained on a single mass point . . . . .	69
8.14.	Overtraining check for a BDT trained on a single high statistics mass point	70
8.15.	Event Yields for an optimized cut on the response of a BDT trained on the mass point ( $m_{\tilde{\tau}} = 200 \text{ GeV}$ , $m_{\tilde{\chi}_1^0} = 1 \text{ GeV}$ ) with a high statistics dataset . .	71
8.16.	Ranking for a parameterized BDT . . . . .	73
8.17.	Comparison of the overtraining checks for a parameterized and an unpa- rameterized BDT . . . . .	73
8.18.	Ranking of input variables for a BDT trained on low stau masses . . . . .	80
8.19.	Overtraining check for a BDT trained on low stau masses . . . . .	80
8.20.	Event Yields for an optimized cut on the response of a BDT trained on low stau masses . . . . .	80
8.21.	Ranking of input variables for a BDT trained on high stau masses . . . . .	83
8.22.	Overtraining check for the BDT trained on high stau masses . . . . .	83
8.23.	Event yields for a cut on the response of a BDT trained on high stau masses	83
9.1.	Weighted and unweighted event yields before and after the promotion of one container $\tau$ -lepton to medium ID . . . . .	92
9.2.	Event yields before and after $\tau$ -promotion per $W \rightarrow \tau\nu$ -dataset with a modified binning for the fake efficiency measurement . . . . .	96

9.3. Weighted and unweighted event yields before and after promoting one container $\tau$ -lepton to loose ID . . . . .	97
9.4. Event yields in the phase space regions used for QCD estimation including promoted loose $\tau$ -leptons in the $W$ +jets dataset . . . . .	99
9.5. Cut combination SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets and QCD datasets . . . . .	100
9.6. Event yields in SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets and QCD datasets . . . . .	101
9.7. Cut combination for SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets dataset . . . . .	101
9.8. Event yields in SR DS HadHad High Mass with $\tau$ -promotion applied in the $W$ +jets dataset . . . . .	103
9.9. Event yields for a cut on the response of a BDT trained on high stau masses with $\tau$ -promotion in $W$ +jets and QCD . . . . .	105
9.10. Event yields for a cut on the response of a BDT trained on high stau masses with $\tau$ -promotion in $W$ +jets I . . . . .	105
9.11. Event yields for a cut on the response of a BDT trained on high stau masses with $\tau$ -promotion in $W$ +jets II . . . . .	107
10.1. Summary of best significances . . . . .	112
A.1. Signal datasets with stau masses from 80 GeV to 120 GeV . . . . .	115
A.2. Signal datasets with stau masses from 140 GeV to 260 GeV . . . . .	116
A.3. $W$ +jets datasets for $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ . . . . .	117
A.4. $W$ +jets datasets for $W \rightarrow \tau\nu$ . . . . .	118
A.5. $Z$ +jets datasets for $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ . . . . .	119
A.6. $Z$ +jets datasets with $Z \rightarrow \tau\tau$ and $Z \rightarrow \nu\nu$ . . . . .	120
A.7. Diboson datasets . . . . .	121
A.8. Top quark pair production datasets . . . . .	121
A.9. Single top quark production datasets . . . . .	122
A.10. Top quark pair production with vector boson radiation datasets . . . . .	122



# Erklärung

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

Clara Leitgeb

München, den 25. Oktober 2017