

Optimising variables in the selection of $Z \rightarrow \tau\tau$ events in the $\tau_\mu\tau_{\text{had}}$ final state at the ATLAS experiment*

Ricardo Wölker^{1,†}

(Student ID: 8939451)

Supervised by Prof. Terry Wyatt¹,
performed in collaboration with Vilius Čepaitis¹

¹*School of Physics and Astronomy,
University of Manchester, Manchester, UK*

(Dated: May 15, 2017)

Obtaining pure samples of Z bosons decaying into pairs of tau leptons is important in verifying proposed theories beyond the standard model of particle physics. A predicted heavier cousin of the Z boson, the Z' , could have an enhanced coupling to tau leptons. In the semileptonic decay channel with a hadronic tau and muon as final-state particles, the main background to these events comes from $W+\text{jets}$ events, in which the W decays into a muon and a jet fakes a hadronic tau. A selection of event variables is tested for its discriminating power between these sets of events, using full ATLAS physics simulations. A version of the transverse mass that makes use of an approximate boost into the centre-of-mass frame of the decay, $m^*(\mu, E_T^{\text{miss}})$, is shown to be the most discriminating single variable. Additional cuts on the visible mass between the muon and the tau, as well as the muon-tau transverse-momentum asymmetry ratio, are used to construct an event-selection variable which exceeds in its performance the currently recommended set of variables. The ability of this new variable, Γ , to reject background from $W+\text{jets}$ events is compared with other multivariate candidate methods. Predicted distributions from physics simulations are compared with recorded ATLAS data.

* MPhys project final report

† ricardo.woelker@student.man.ac.uk

CONTENTS

I. Introduction	3
II. The ATLAS detector	4
A. Coordinate system	4
B. Missing transverse energy	4
III. Monte Carlo Simulations	5
A. Basics and generators used	5
B. Sample properties	6
C. Merging samples	7
IV. Multivariate techniques	7
A. Boosted decision trees	7
V. Tau lepton properties	7
A. Hadronic decay modes	8
B. Tau reconstruction and identification	8
VI. Event selection	9
A. Physics objects	9
B. Triggers	9
C. Preselection cuts	9
D. Overlap removal	10
VII. Physics analysis	10
A. General approach	10
B. Discriminating variables	11
1. Angular variables	11
2. Mass variables	13
C. ATLAS cuts	15
D. ROC curves from fixed cuts	16
E. ROC curves from two-dimensional cuts	17
1. Rectangular fixed cuts	17
2. Selecting regions based on 2D $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ plots	19
3. Linear cuts and the triangle parameter Δ	19
4. Γ in $\rho - \Delta$ space	19
F. Linear correlations between variables	20
G. BDT results	20
H. Final results	22
1. The significance improvement characteristic, SIC	26
VIII. Data-MC comparison	28
IX. Conclusion and future work	30
Acknowledgments	30
References	31
A. Supplementary figures	33
B. Monte Carlo samples used	43
C. Muon triggers	43
D. Preselection cutflow	43

I. INTRODUCTION

The discovery of a heavy boson by the CMS and ATLAS collaborations consistent with the standard model (SM) Higgs boson in 2012 [1, 2] has provided the last remaining piece of the SM of particle physics. The SM, however, has a number of shortcomings, such as its inability to explain dark matter, to provide an explanation of what happens at very high energies on the grand-unification scale ($\sim 10^{16}$ GeV), or to give an adequate account of why the elementary particles have particular masses. Particle physics is presently concerned with two major avenues: (i), to measure precisely the parameters of the SM in order to find deviations from its theoretical predictions, and (ii), to search for phenomena beyond the standard model (BSM).

An efficient identification of tau leptons is essential for many searches within the standard model and beyond. Theories that try to account for the behaviour at the grand-unification scale foresee heavier partners of the SM carriers of the weak force (Z, W^\pm). These heavier partners are called Z' and $W^{\pm'}$, and are predicted by the minimal theoretical extension of the electroweak symmetry of the SM: the sequential standard model (SSM) [3]. Certain models that aim to explain the mass of the top quark predict that these heavy bosons couple favourably to tau leptons. A discovery of this type would provide evidence for the violation of lepton universality; the expectation that the electroweak interaction couples equally to leptons of all three generations in the SM [4]. ATLAS has recently searched for Z' bosons decaying to a pair of tau leptons, and has excluded Z' bosons with masses less than 2.02 TeV at 95% confidence level [5, 6].

Further, tau leptons play an important role in the search for supersymmetry. The minimal supersymmetric extension of the SM (MSSM) [7–9] is the simplest model to include supersymmetric particles. Unlike the SM, which contains one Higgs boson (h) only, the MSSM foresees five Higgs bosons, whose couplings to tau leptons could be enhanced compared to those to light leptons (electrons and muons). To first order, the model can be parametrised by $\tan\beta$, which is the ratio of the vacuum expectation values of the two Higgs doublets in the MSSM, a quantity related to the ratio of the couplings to up and down-type fermions. For certain values of $\tan\beta$, the couplings of the Higgs bosons to tau leptons are greatly enhanced. The SM decay of $Z \rightarrow \tau\tau$ can be used as a calibration sample, as it will constitute a background to this process.

In this report, we study the decay of a Z boson (or virtual photon, γ^*) into a $\tau^+\tau^-$ pair ($Z \rightarrow \tau\tau$, the *signal*), where one tau decays hadronically (τ_{had}) and the other tau decays leptonically into a muon (τ_μ), with the aim of improving the rejection of *background* arising from $W+\text{jets}$ events. In $W+\text{jets}$ events, one of the incoming quarks may radiate a gluon which, when it hadronises, can leave a similarly signature in the detector to a hadronic tau (it is said to *fake* a τ , τ_{fake}). The Feynman diagrams for the two processes are shown in Figure 1. In either case, a quark-antiquark ($q\bar{q}$) pair annihilates to produce a vector boson: a Z is produced in the case of two up-type or two down-type quarks, whereas a quark pair consisting of an up-type and a down-type quark produces a W . Another important source of background arises from rare multijets which can be misidentified as tau leptons. This source of background cannot be modelled reliably using physics simulations, and hence this study focuses on $W+\text{jets}$ as the sole source of background. *Data-driven* background-estimation techniques can be used to estimate the background contribution arising from multijets, in which the differing rates of *opposite-sign* (OS) and *same-sign* (SS) events in signal and background are exploited.

This study aims to improve the discrimination of $Z \rightarrow \tau\tau$ signal events against background events from $W+\text{jets}$ using Monte Carlo (MC) simulations of the two processes. The samples are analysed using the CERN **ROOT** framework [10] based on C++. This report is structured as follows: the ATLAS experiment is described in Section II, followed by a brief description of the MC event generation in Section III. Section IV introduces some of the multivariate techniques employed, and the properties of the tau lepton and its reconstruction are given in Section V. The selection of events for the final analysis containing one tau and one muon is explained in Section VI. The results of the analysis, demonstrating an improved event selection using new kinematic variables,

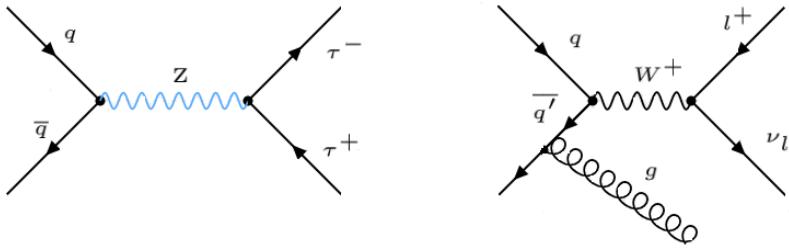


Figure 1: Feynman diagrams for the signal ($Z \rightarrow \tau\tau$, left) and background ($W + \text{jets}$, right) processes. The radiated quark in the background process may leave a similar signature in the detector as a hadronic tau, τ_{had} . Here, ℓ^+ starts for a light lepton (e^+ or μ^+). ν_ℓ is the corresponding neutrino of the same generation.

are presented in Section VII. Section VIII demonstrates the accurate modeling of the recorded data by the MC in a few key distributions. Final remarks and recommendations for future work are made in Section IX.

II. THE ATLAS DETECTOR

The ATLAS (*A Toroidal LHC Apparatus*) detector [11] is a large general-purpose detector located at the CERN Large Hadron Collider (LHC). It is designed to capture the entire solid angle around the proton-proton (pp) collision point. The innermost part of ATLAS is the *inner detector*, which consists of pixel and strip silicon pixel detectors and a *transition radiation tracker*. The inner detector is surrounded by a 2-T solenoid magnetic field, which causes charged particles to follow a curved trajectory. The next layers of the detector are the *electromagnetic* and *hadronic* calorimeters, which measure the energy deposition of electrons, photons and hadrons. The *muon spectrometer* forms the outermost layer of the detector, and is surrounded by a toroidal magnetic field. The muon chambers provide information on muon trajectories with a high spatial resolution, while the triggering chambers have a higher time resolution.

Due to the limited rate at which data can be processed, ATLAS uses a three-level triggering system to record events that are of interest. The Level-1 trigger is hardware-based and reduces the data rate from 40 MHz to 75 kHz. It is reduced further approximately 400 Hz using a two-stage software-trigger system. Figure 2 shows a schematic drawing of the ATLAS detector.

A. Coordinate system

ATLAS uses a right-handed coordinate system in which the x -axis points towards the centre of the LHC ring, the z -axis points along the beam pipe, and the y -axis points upwards. It is convenient to use cylindrical coordinates in the transverse plane, such that $r = \sqrt{x^2 + y^2}$, $\phi = \arctan(\frac{y}{x})$, and $z = z$. The polar angle, θ , is defined as the angle from the positive z -axis. A quantity called the *pseudorapidity*, η , is frequently used in particle physics, and is defined in terms of θ such that $\eta = -\ln[\tan(\frac{\theta}{2})]$. The distance in the space spanned by η and ϕ , ΔR , is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

B. Missing transverse energy

Neutrinos leave the detector unobserved. Their existence, however, can be inferred in the transverse plane from the overall momentum imbalance. In head-on pp collisions, the initial transverse

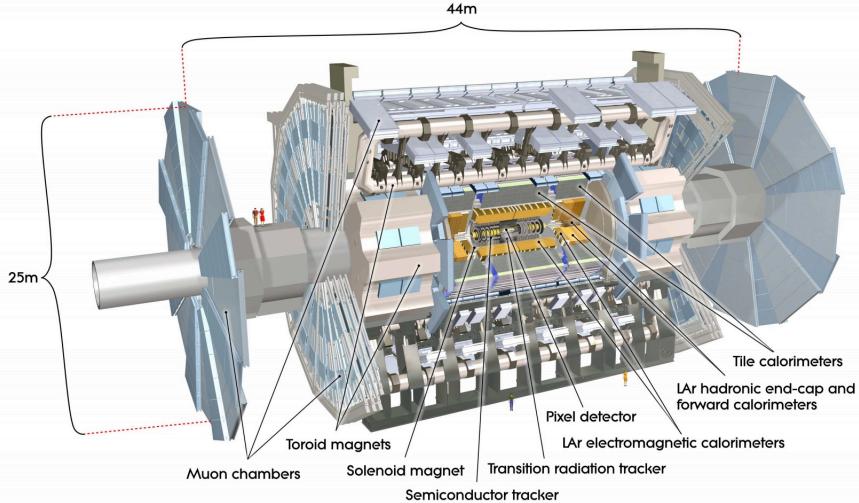


Figure 2: Drawing of the ATLAS detector at the LHC [11]. The detector is comprised of three main parts: the inner detector, the electromagnetic and hardonic calorimeters, and the muon spectrometer. The detector is 44 m long, 25 m in diameter, and weighs a total of 7,000 tonnes.

momentum is zero, which means that any net transverse momentum of all detected particle must be balanced by neutrinos. This is called the missing transverse energy, \vec{E}_T^{miss} , and it is defined as the negative vectorial sum of the transverse momenta of all detected particles, viz.

$$\vec{E}_T^{\text{miss}} = - \sum_{i=1}^N \mathbf{p}_T^i \quad (1)$$

for N detected particles, where the \mathbf{p}_T^i ¹ are the transverse momenta of the particles detected in the event. We will further refer to the magnitude of \vec{E}_T^{miss} as E_T^{miss} . E_T^{miss} is an important observable in this work due to neutrinos' being present in all tau decays. The data files used are equipped with various containers for the missing transverse energy. The `MET_RefFinal` container is used to retrieve E_T^{miss} , which contains for a given event the sum of all constituent E_T^{miss} components.

III. MONTE CARLO SIMULATIONS

A. Basics and generators used

In order to gain a better understanding of the physical process under investigation, it is necessary to predict accurately the distributions of the underlying event variables. In particle physics, the *Monte Carlo method* is widely used to simulate collisions of particles and their interactions with the detector material. These problems require the solution of equations that are too complex to be solved analytically. The Monte Carlo method provides an approach of solving these equations by sampling the underlying distributions using random numbers.

At the base of these simulations lies the *event generation*, which aims to calculate the scattering amplitude of the underlying event, followed by the simulation of parton showering. The partons eventually form colour-neutral hadrons, and enter the detector where they interact with matter. The MC samples used in this study have been provided by the ATLAS MC working group in the `Ntuple` [10] format, and include all physical processes present in data apart from multijet events. The event generation software used is `SHERPA` [12], which is combined with `GEANT4` [13] to simulate

¹ We will throughout this report refer to the magnitude of \mathbf{p}_T as p_T , such that particle x has transverse momentum \mathbf{p}_T^x with magnitude p_T^x .

	preselection cut	muons	electrons	taus
p_T		$> 5 \text{ GeV}$	$> 5 \text{ GeV}$	$> 15 \text{ GeV}$
$ \eta $		< 2.6	< 2.6	\times
N_{tracks}	\times	\times	1 OR 3	
isolation	\times	loose	\times	

Table I: Summary of the loose preselection cuts applied to the $t\bar{t}H$ multilepton samples used in this study. See Section V A for an explanation of N_{tracks} .

	skimming cut	muons	electrons	taus
p_T		$> 5 \text{ GeV}$	$> 5 \text{ GeV}$	$> 15 \text{ GeV}$
leading p_T		$> 15 \text{ GeV}$	$> 15 \text{ GeV}$	\times
$ \eta $		< 2.6	\times	< 2.6
N_{tracks}	\times		1 OR 3	\times
$ q $	\times	\times	\times	1

Table II: Skimming cuts on muons, electrons, and taus. $|q|$ denotes the absolute value of the electric charge.

the response of the detector. In order to perform an unbiased comparison between simulation and data, they are processed and analysed in the same fashion.

B. Sample properties

The samples used in this report have been created originally for a study of the associated production of the Higgs with a top-antitop pair ($t\bar{t}H$), and all events are required to contain at least two light leptons ($\ell = e, \mu$). The objects in the sample have additional loose cuts applied which are summarised in Table I.

The samples have a *loose two-lepton skim* applied which greatly reduces the size of the dataset. The events have to contain at least two light leptons, at least one of which has to pass leading- p_T requirements². Alternatively, events can contain at least one light lepton (passing a leading- p_T requirement) and at least one tau with $\Delta R(\ell, \tau) > 0.4$, which is the case this study focuses on. The objects' skimming cuts are summarised in Table II.

The work presented in the interim report on this study [14] made use of samples that are split according to the number of partons above *tree-level*³ partaking in the interaction, ranging from zero (Np0) to more than four (Np5). Only the zero-parton slice was used, which presented a critical shortcoming in the first stage of the project that has now been mended. Instead, we here use SHERPA-generated samples which are divided up into 24 slices of boson p_T , all of which are combined. The slices are merged by assigning to each sample a weight, w , given by

$$\omega = \frac{\sigma \mathcal{L}_{\text{int}} K}{N}, \quad (2)$$

where N is the number of simulated events in the sample, σ is the cross-section of the simulated process, and \mathcal{L}_{int} is the integrated luminosity of the entire sample (usually matched to a certain

² Leading p_T refers to the p_T of the lepton with the higher p_T if more than one is detected. For these events, the acceptance thresholds are different for the leading lepton and the sub-leading lepton.

³ *Tree level* denotes processes described by Feynman diagrams without closed loops. Higher-order diagrams can have a number of closed loops, but contribute less to the overall cross-section of the process.

data-taking period), where K , also known as the K -factor, is the ratio of next-to-leading-order (NLO) to leading-order (LO) predictions.

C. Merging samples

This work uses a combination of samples generated by **SHERPA** at NLO as well as at next-to-next-to-leading order (NNLO) so as to increase the available statistics⁴. Each sample is normalised to the expected number of events in the total ATLAS dataset recorded in 2015 and 2016 ($\mathcal{L}_{\text{int}} = 36.1 \text{ fb}^{-1}$). Figure A.8 shows a comparison between the different samples in a few important kinematic distributions. It is evident that they are nearly identical to one another, which allows us to merge them together, using a table of cross-sections available at Reference [15].

IV. MULTIVARIATE TECHNIQUES

Characterising a particle-physics event usually requires multiple kinematic quantities to be considered. Examples of these *event variables* include the particles' transverse momenta, their invariant mass, or the angle in the transverse plane between two particles, and are generally correlated with one another. In order to obtain results from these event variables most efficiently, it is useful to consider them simultaneously using multivariate analysis methods [16].

A central concept in multivariate analysis methods is that of *machine learning*, a computing paradigm for algorithms that learn autonomously from the input data. Within the application of background discrimination in particle physics, the algorithm is trained using two separate samples (one for signal and one for background) – these are called the *training samples* and are usually obtained from Monte Carlo simulations. The goal of the training is to find a set of parameters that optimally discriminate the signal from the background. Once the training is completed, the algorithm is applied to a *test sample*, which could be either recorded data or a different set of Monte Carlo-simulated data.

A. Boosted decision trees

A *boosted decision tree* (BDT) is an instance of a multivariate technique [17]. Given a set of variables, the BDT algorithm is trained by placing a one-dimensional cut on the most discriminating variable (the *root node*). This step is repeated in the resulting two subsets using the second most discriminating variable. This is continued recursively until all variables are exhausted. The whole process is repeated, and events that have been classified wrong are assigned a higher weight in the next iteration (they are *boosted*). The output of a BDT can be drawn as a tree, where each leaf corresponds to a certain *purity*, i.e. the probability that an event in a particular leaf is signal. Figure A.1 shows an example output of a BDT.

The BDT analysis used in this report is performed using the Toolkit for Multivariate Analysis in ROOT (TMVA) [18]. It provides, among other multivariate algorithms, a BDT library.

V. TAU LEPTON PROPERTIES

The tau lepton is a spin-1/2 charged lepton of the third generation, and is about 3,500 times heavier than the electron. It has a mass of $m_\tau = 1.777 \text{ GeV}$ [19], which makes it the heaviest lepton of the SM. The tau lepton decays very quickly with a mean lifetime of $t_{1/2} = 2.9 \times 10^{-13} \text{ s}$. The τ^- decays leptonically with a probability of 35% into a neutrino (ν_τ) and a W^- , which in turn decays

⁴ A more detailed overview of the samples used is given in Table A1.

into a light lepton (e, μ) and a corresponding antineutrino ($\bar{\nu}_e, \bar{\nu}_\mu$). A leptonically decaying tau is referred to as τ_{lep} . The τ^- mainly decays hadronically with a branching ratio of 65% into a ν_τ and a W^- , which decays into quarks that subsequently hadronise. Thus, a hadronic tau is denoted by τ_{had} . A Feynman graph of the decay modes of the tau is shown in Figure 3. The τ^+ decays similarly into the relevant antimatter equivalents of the particles mentioned [20].

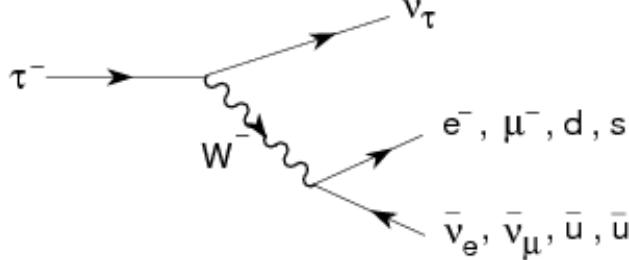


Figure 3: Decay modes of a tau lepton. The τ^- always decays into a ν_τ and a W^- . The W^- can decay either into a same-generation lepton-neutrino pair or into a different-generation quark pair.

As any tau-lepton decay produces at least one neutrino, which leaves the detector unobserved, there is always some non-zero missing transverse energy.

A. Hadronic decay modes

As the overall tau decay conserves electric charge, a τ_{had} can only decay into an odd number of charged hadrons, and, in principle, any number of neutral hadrons. The number of charged hadrons in a τ_{had} candidate is referred to as the *track multiplicity*⁵, or N_{tracks} . We here consider the two most common cases with either one or three charged hadrons.

B. Tau reconstruction and identification

Their decay length of $ct_{1/2} = 87 \mu\text{m}$ means that taus decay before reaching the detector, and are only identified via their decay products [21]. τ_{had} decays predominantly produce charged pions, and in 78% of cases at least one neutral pion. The ensemble of neutral and charged pions of the τ_{had} decay make up the visible part of the decay (the neutrino is missing), and is thus normally referred to as $\tau_{\text{had-vis}}$ ⁶. The main background to τ_{had} candidates arises from hadrons coming from quarks and gluons.

The τ_{had} reconstruction is *seeded* by selecting energy deposits in the calorimeter that have been identified as jets with $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$ (the so-called *jet seed*). The resulting vertex must have at least three associated tracks, and must be within a cone of $\Delta R < 0.2$ around the direction of the jet seed. At this stage, the algorithm does not yet make a distinction between jets and τ_{had} candidates. Jets resulting from hadronic taus must have a reconstructed visible mass that is smaller than m_τ , and typically contain a small number of hadrons (less than four) [22]. They further tend to be more collimated than hadronic jets. The final identification is performed using a BDT algorithm, which assigns a score between 0 and 1, corresponding to the probability that the tau candidate is a real tau. The *working point* for tau candidates in this analysis was chosen to be > 0.3 , which is from now on referred to as a **medium** tau ID. The efficiency for a **medium** BDT score is roughly 60% for 1-prong tau decays and 40% for 3-prong decays [21].

⁵ The track multiplicity is also denoted by the term *prong*, so that, for instance, a 1-prong tau candidate indicates that it has decayed into one charged hadron (and potentially a number of neutral hadrons).

⁶ For simplicity, $\tau_{\text{had-vis}}$ is throughout this report referred to as τ_{had} , as only the visible part of the decay enters into the final analysis as a reconstructed object.

VI. EVENT SELECTION

At the beginning of every measurement, the events that are likely to be interest have to be selected. This is done by placing certain selection criteria on all events before they are chosen for the analysis. Every event consists of a set of reconstructed particles, called *objects*, which are created using sophisticated identification algorithms (cf. Section V B). As this study is concerned with the $\mu\tau_{\text{had}}$ final state, we can place a number of constraints on the objects that should be present in the event, and on some of their kinematic properties.

A. Physics objects

The possible tau decay modes described in Section V dictate that the final-state particles that should be considered in this work are hadronic taus, electrons, and muons, and hence these are the main physics objects under investigation. However, as this study is primarily concerned with a leptonic tau decaying into a muon, electrons can be ignored. In addition, as was explained before (Section V), some non-zero missing transverse energy is always to be expected, which means that $E_{\text{T}}^{\text{miss}}$ also enters as an important quantity. Photons and b -jets are not used as they lack an obvious relation to the final state studied.

B. Triggers

In order to reduce the high data rates in the high-luminosity environment of the LHC, potential events of interest are *triggered*, and subsequently recorded and reconstructed. To recreate the same conditions as in recorded data, five different muon triggers are applied, selecting events with one muon in the final state. They fire at 20, 24, 26, 40, and 50 GeV, respectively, and are summarised in Table A2.

C. Preselection cuts

Events are selected for the final analysis if they contain at least one τ_{had} , exactly one muon and no electrons. If more than one τ_{had} is present, the candidate with the highest p_{T} is selected. Muons must satisfy $p_{\text{T}} > 27$ GeV, while hadronic taus are required to have $p_{\text{T}} > 20$ GeV. Muons and taus are accepted in the forward region up to $|\eta| < 2.47$. They are excluded if they are found in a *crack* region, where the barrel and end-cap calorimeters meet, as the energy resolution there is poor. Muons are required to have a medium identification strength, while a similar requirement in the τ_{had} case translates to a BDT score of at least 0.3. As the Z is electrically neutral, the τ_{had} and the muon as its decay products are required to be oppositely charged.

In addition, muons must satisfy isolation requirements that ensure a vetoing of muons with too much excess energy around their actual track. This allows muons from W decays to be distinguished from those that arise within hadronic jets. The two variables used to this end are, (i), the track-based variable $p_{\text{T}}\text{VarCone30}/p_{\text{T}}$. It is defined as the sum of the transverse momenta of all tracks within a cone of $\Delta R < 0.3$ not associated with the muon, divided by the muon- p_{T} , p_{T}^{μ} , i.e.

$$p_{\text{T}}\text{VarCone30}/p_{\text{T}} = \frac{\sum_{\text{non}-\mu \text{ tracks}} |p_{\text{T}}^{\text{track}}|}{|p_{\text{T}}^{\mu}|}. \quad (3)$$

There is, (ii), a similar calorimeter-based variable that measures activity around the muon track called $E_{\text{T}}\text{TopoCone20}/p_{\text{T}}$. It is defined within a cone of $\Delta R < 0.2$ of calorimeter clusters around

the muon track and takes into account the transverse energy from *pile-up events*⁷, viz.

$$E_T \text{TopoCone}20/p_T = \frac{\sum_{\text{clusters}} E_T^{\text{cluster}} - E_T^\mu - E_T^{\text{pileup}}}{|p_T^\mu|}. \quad (4)$$

The preselection cuts are summarised in Table III.

preselection cut	muons	taus
p_T	$> 27 \text{ GeV}$	$> 20 \text{ GeV}$
$ \eta $	< 2.47	< 2.47
$ \eta_{\text{crack}} $	$< 1.37 \text{ OR } > 1.52$	$< 1.37 \text{ OR } > 1.52$
N_{tracks}	\times	$1 \text{ OR } 3$
ID strength	medium ID	BDT ID > 0.3
isolation	$p_T \text{VarCone}30/p_T < 0.01$ $E_T \text{TopoCone}20/p_T < 0.05$	\times

Table III: Summary of preselection cuts on muons and taus.

D. Overlap removal

In order to avoid the simultaneous reconstruction of a particle as, for instance, both a muon and an electron, an *overlap removal* is performed. Overlap is an unphysical consequence of the often imperfect particle reconstruction and must be accounted for accordingly. Overlapping objects are preferentially selected as (1) muons, (2) electrons, and (3) taus. The order of priorities is related to the efficiencies with which particles can be reconstructed from the detector signatures: the muon identification (ID) is more efficient than the electron ID, which is more efficient than the tau ID. The removal happens via a measurement of ΔR between the particles: If a muon track is too close to an electron track ($\Delta R(e, \mu) < 0.01$), the electron is likely to have come from bremsstrahlung photons that have been misidentified, and the electron is removed. Similarly, if the ΔR between an electron track (muon track) is too close to that of a tau ($\Delta R(e/\mu, \tau) < 0.15$), the lepton is selected and the tau is removed. The overall cutflow including the number of events passing each stage of the selection is given in Table A3.

VII. PHYSICS ANALYSIS

A. General approach

In order to separate signal events of interest from background, it is important to investigate a diverse set of event variables that ideally capture and enhance the features that are most different between the two categories of events. An event variable, or *observable*, is said to be *discriminating* if, given some cut on it, more signal than background events are selected. Central to the study of discriminating power is the concept of *efficiency*. The fraction of signal events that pass a certain cut on the variable in question is referred to as the *signal efficiency*, ϵ_{sig} , whereas the fraction of background events passing the cut is called the *background efficiency*, ϵ_{bkg} .

⁷ Pile-up is any activity arising from additional pp events that can interfere with the reconstruction of the underlying event of interest. A typical crossing of the two proton beams in ATLAS creates tens of pp interactions.

B. Discriminating variables

1. Angular variables

It is in general not possible to determine the net longitudinal momentum in collisions at hadron colliders, as the partons that participate in the interaction carry only some unknown fraction of the momentum of the incoming hadrons. The momentum in the plane transverse to the beam direction, however, is initially very close to zero⁸. This makes it useful to consider variables in the transverse plane which are invariant under a boost in the longitudinal direction (the z -direction in cylindrical coordinates). Some important variables in the transverse plane are shown in Figure 4.

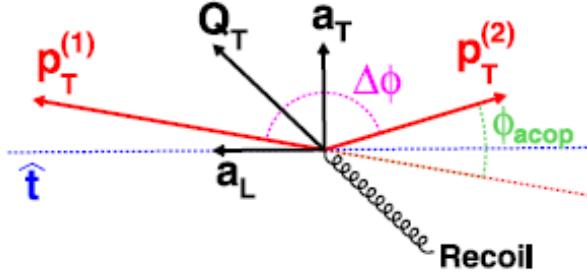


Figure 4: Schematic drawing showing important variables defined in the plane that lies transverse to the beam direction [23]. The total boson transverse momentum is \mathbf{Q}_T , with a component along and perpendicular to the event axis \hat{t} (a_L and a_T , respectively). The boson decays into two leptons with transverse momenta $\mathbf{p}_T^{(1)}$ and $\mathbf{p}_T^{(2)}$.

The Z boson carries transverse momentum \mathbf{Q}_T , which subsequently decays into two tau leptons, each carrying transverse momentum $\mathbf{p}_T^{(1)}$ and $\mathbf{p}_T^{(2)}$. It follows then, by momentum conservation, that $\mathbf{Q}_T = \mathbf{p}_T^{(1)} + \mathbf{p}_T^{(2)}$. It is useful to construct an event axis given by

$$\hat{t} = \frac{\mathbf{p}_T^{(1)} - \mathbf{p}_T^{(2)}}{|\mathbf{p}_T^{(1)} - \mathbf{p}_T^{(2)}|}. \quad (5)$$

The component of \mathbf{Q}_T transverse to \hat{t} is a_T ⁹, and a component parallel to \hat{t} is denoted by a_L . $\Delta\phi$ ¹⁰ is the $\tau\tau$ opening angle, and is often used as $\phi_{acop} \equiv \pi - \Delta\phi(\mu, \tau_{had})$, which quantifies the deviation from a back-to-back emission of the two tau leptons via their visible final states (a muon and a hadronic tau).

The use of a purely angular variable in studying the transverse momentum of a dilepton system, ϕ_η^* , has been demonstrated recently [23]. The construction of ϕ_η^* involves an approximate boost into the dilepton centre-of-mass frame using angular quantities only. This boost yields an angle θ_η^* , which is approximately given by the polar angle relative to the beam direction of both leptons in the centre-of-mass frame of the dilepton system. It is defined as

$$\cos(\theta_\eta^*) = \tanh\left(\frac{\eta^- - \eta^+}{2}\right), \quad (6)$$

where η^\pm are the pseudorapidities of the positively and negatively charged leptons, respectively. ϕ_η^* is then defined as

$$\phi_\eta^* \equiv \tan\left(\frac{\phi_{acop}}{2}\right) \cdot \sin(\theta_\eta^*). \quad (7)$$

⁸ By the time a quark-quark interaction takes place to produce, say, a Z boson, the net transverse momentum may be non-zero due to gluon radiation.

⁹ If the $\tau\tau$ opening angle, $\Delta\phi$, is less $\pi/2$, a_T is set to be equal to the boson p_T , i.e. $a_T = Q_T$ if $\Delta\phi < \pi/2$.

¹⁰ To avoid confusion with other $\Delta\phi$ -type variables, this quantity is in the case of the $\tau_{had}\tau_\mu$ final state referred to as $\Delta\phi(\mu, \tau_{had})$.

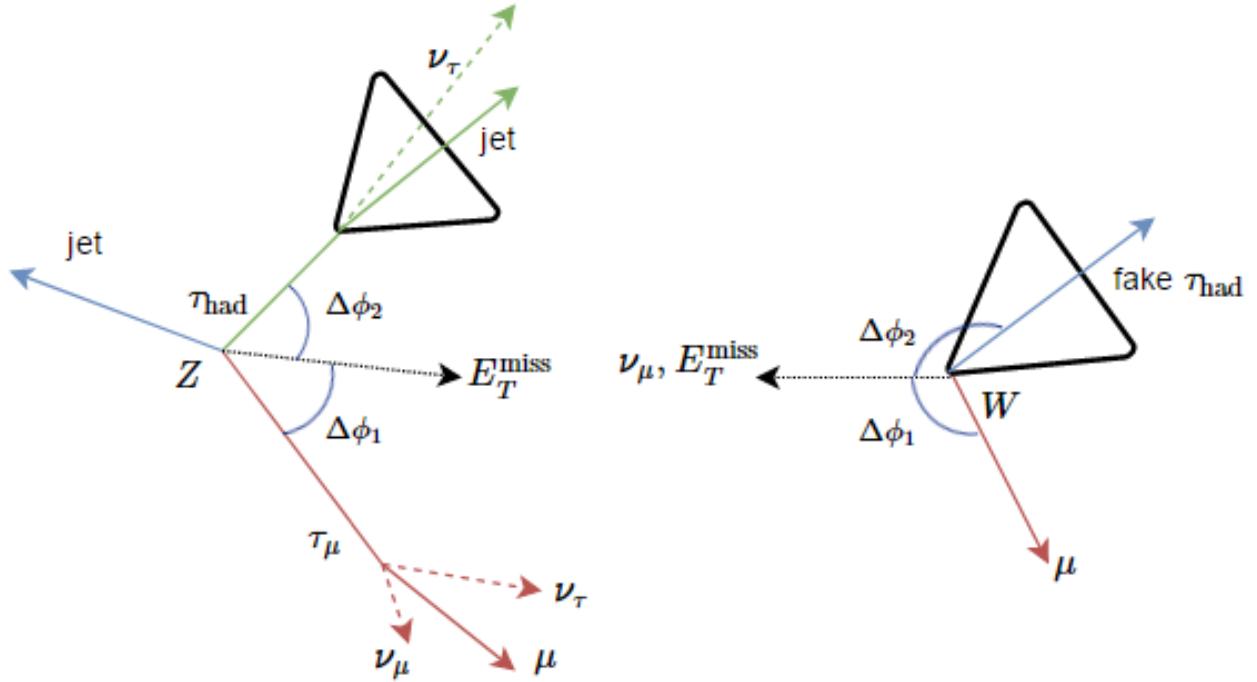


Figure 5: The decay topology for $Z \rightarrow \tau\tau$ and $W + \text{jets}$ events. The red lines correspond to the production of the final-state muon, whereas the blue lines symbolise the recoiling jet. Here, $\Delta\phi_1 = \Delta\phi(E_T^{\text{miss}}, \mu)$ and $\Delta\phi_2 = \Delta\phi(E_T^{\text{miss}}, \tau_{\text{had}})$.

Another important angular quantity is the opening angle in the transverse plane between \vec{E}_T^{miss} and the transverse-momentum vector of the hadronically decaying tau, \mathbf{p}_T^τ , referred to as $\Delta\phi(\tau_{\text{had}}, E_T^{\text{miss}})$. The opening angle in the transverse plane between \vec{E}_T^{miss} and the p_T of the leptonically decaying tau, \mathbf{p}_T^μ , is referred to as $\Delta\phi(\mu, E_T^{\text{miss}})$. In $Z \rightarrow \tau\tau$ signal events, \vec{E}_T^{miss} is expected to lie within the opening angle of the semileptonically decaying $\tau\tau$ system. This is because the components of \vec{E}_T^{miss} carried by either the τ_{had} or the muon are split roughly equal, and any significantly non-zero \vec{E}_T^{miss} arises from recoiling jets radiated from the Z boson before it decays. In $W + \text{jets}$ background events, on the other hand, the neutrino from the W decay tends to point away from both the jet $\rightarrow \tau_{\text{fake}}$ and the muon. Figure 5 shows schematically the event topologies for typical $Z \rightarrow \tau\tau$ and $W + \text{jets}$ decays. An event variable given by the sum of the cosines of the two opening angles relative to \vec{E}_T^{miss} aims to capture this fundamental difference in decay topology. Referred to as SumCos, it is defined as

$$\text{SumCos} = \cos(\Delta\phi(E_T^{\text{miss}}, \mu)) + \cos(\Delta\phi(E_T^{\text{miss}}, \tau_{\text{had}})), \quad (8)$$

and is expected to take on values near zero for the signal decay. In the $W + \text{jets}$ background process, the distribution in SumCos is expected to shift towards negative values.

Similarly, conclusions about the transverse momenta of the final-state particles can be drawn from the different event topologies. Z bosons are produced at rest on average, which means the transverse momenta of the tau lepton and the muon, p_T^τ and p_T^μ , are very similar to one another. The τ_{fake} in the background decay is not kinematically constrained, and therefore has in general a very different transverse momentum from the muon. A p_T -asymmetry ratio, ρ , aims to capture this difference, and is given by

$$\rho = \frac{p_T^\tau - p_T^\mu}{p_T^\tau + p_T^\mu}. \quad (9)$$

Figures 6 and 7 show the distributions in these variables for events that pass the basic event

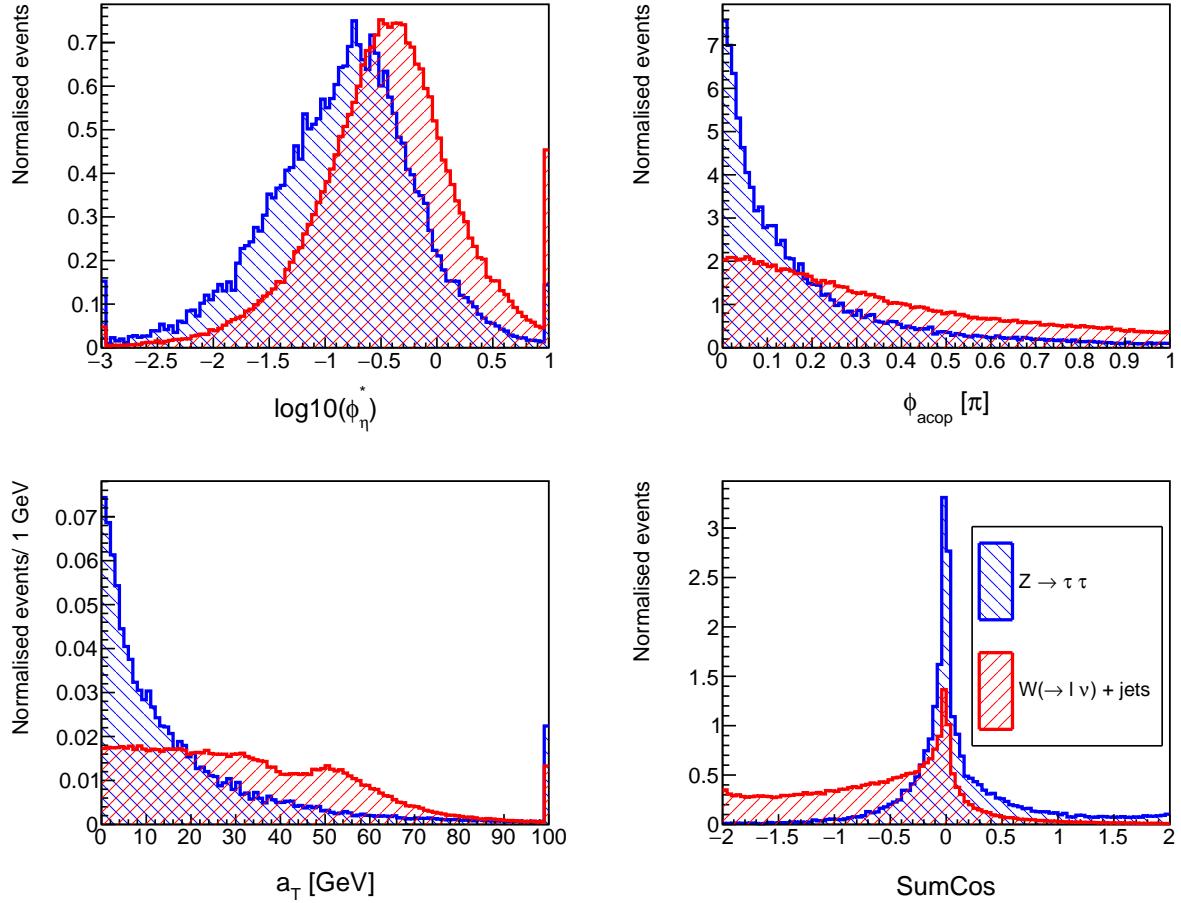


Figure 6: Distributions in four of the angular event variables for $Z \rightarrow \tau\tau$ (blue) and $W + \text{jets}$ (red). Note that these distributions are normalised to unit area to facilitate a simple visual comparison.

selection criteria (as given in Section VI and Table I).

2. Mass variables

For final states involving neutrinos, it is impossible to fully reconstruct the invariant mass of the decaying system, as the forward component of the missing energy is unknown. The simplest approach to obtaining a mass-related quantity is to form the invariant mass between the final-state muon and the visible part of the τ_{had} (not involving the ν_τ). It is given by

$$m_{\text{vis}}^2(\mu, \tau_{\text{had}}) = 2 p_T^\mu p_T^\tau (\cosh \Delta\eta(\mu, \tau_{\text{had}}) - \cos \Delta\phi(\mu, \tau_{\text{had}})), \quad (10)$$

where $\Delta\eta(\mu, \tau_{\text{had}})$ is the difference in pseudorapidity between the muon and the tau. When constrained to the transverse plane, it is possible to include neutrinos in mass-related variables through E_T^{miss} [24]. In general, this transverse mass, m_T follows from the energy-momentum relation in the transverse plane, and is given by

$$m_T^2 = E_T^2 - \mathbf{p}_T^2, \quad (11)$$

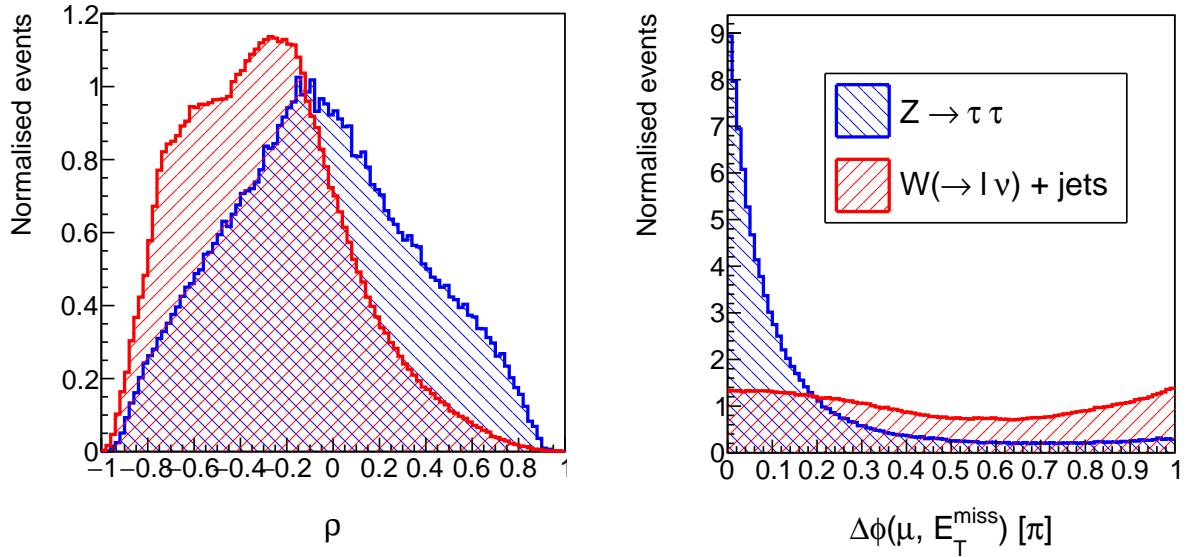


Figure 7: Distributions in two additional angular event variables. These distributions are also normalised to unit area.

where E_T is the total energy in the transverse plane, and \mathbf{p}_T is the sum of all transverse momenta. It is straightforward to form transverse masses between the muon and E_T^{miss} , viz.

$$m_T^2(\mu, E_T^{\text{miss}}) = 2 p_T^\mu E_T^{\text{miss}} (1 - \cos \Delta\phi(\mu, E_T^{\text{miss}})). \quad (12)$$

The transverse mass formed by the muon and the tau is constructed similarly via

$$m_T^2(\mu, \tau_{\text{had}}) = 2 p_T^\mu p_T^\tau (1 - \cos \Delta\phi(\mu, \tau_{\text{had}})). \quad (13)$$

A version of the transverse mass involving all three momenta of the muon, the tau, and E_T^{miss} is given by

$$m_T^2(\mu, \tau_{\text{had}}, E_T^{\text{miss}}) = m_T^2(\mu, E_T^{\text{miss}}) + m_T^2(\mu, \tau_{\text{had}}) + m_T^2(\tau_{\text{had}}, E_T^{\text{miss}}), \quad (14)$$

where $m_T^2(\tau_{\text{had}}, E_T^{\text{miss}})$ is the transverse mass between the muon and the tau. It is given by

$$m_T^2(\tau_{\text{had}}, E_T^{\text{miss}}) = 2 p_T^\tau E_T^{\text{miss}} (1 - \cos \Delta\phi(\tau_{\text{had}}, E_T^{\text{miss}})). \quad (15)$$

We propose a new variable related to Equation (12), which takes into account a boost in the longitudinal direction through θ_η^{*11} (see Equation (6)). It is given by

$$m^*(\mu, E_T^{\text{miss}}) = \frac{m_T(\mu, E_T^{\text{miss}})}{\sin \theta_\eta^*}. \quad (16)$$

Similarly, we extend the notion of a transverse mass involving three particles (see Equation (14)) to give

$$m^*(\mu, \tau_{\text{had}}, E_T^{\text{miss}}) = \frac{m_T(\mu, \tau_{\text{had}}, E_T^{\text{miss}})}{\sin \theta_\eta^*}. \quad (17)$$

¹¹ As the longitudinal component of the neutrinos is unknown, θ_η^* is constructed using the muon and tau momenta, i.e. $\theta_\eta^* = \theta_\eta^*(\mu, \tau_{\text{had}})$.

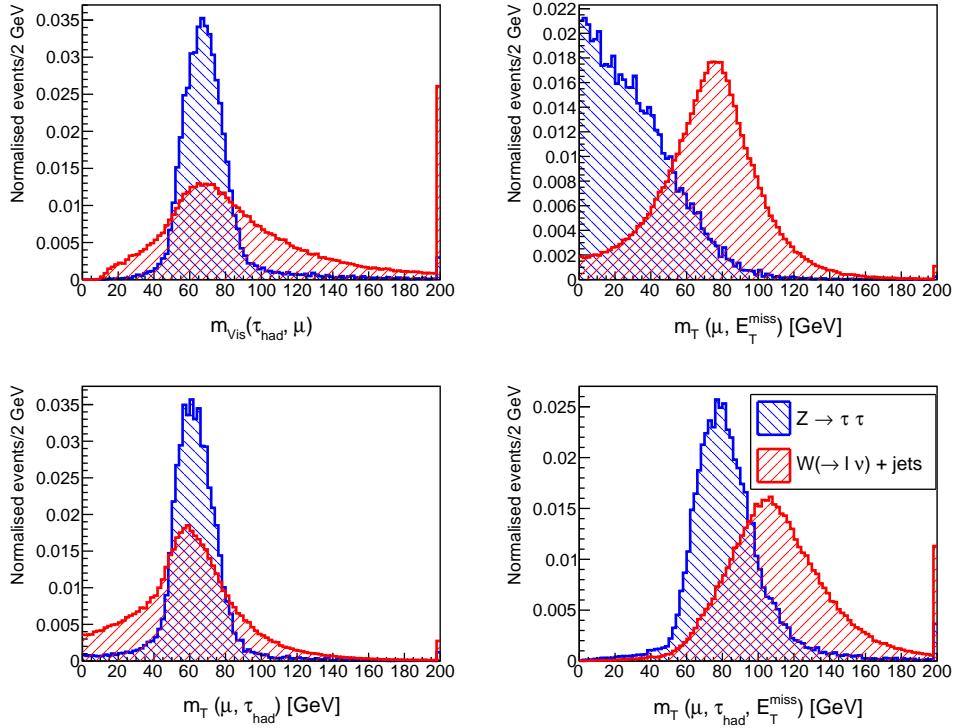


Figure 8: Distributions in the mass variables defined by Equations (12) to (15). Out of the four distributions shown, $m_T(\mu, E_T^{\text{miss}})$ looks to be most discriminating.

For completeness, we also mention the construction of

$$m^*(\mu, \tau_{\text{had}}) = \frac{m_T(\mu, \tau_{\text{had}})}{\sin \theta_\eta^*} \quad (18)$$

and

$$m^*(\tau_{\text{had}}, E_T^{\text{miss}}) = \frac{m_T(\tau_{\text{had}}, E_T^{\text{miss}})}{\sin \theta_\eta^*}. \quad (19)$$

The distributions in these mass variables are shown in Figures 8 and 9. As is evident from these distributions, $m^*(\mu, E_T^{\text{miss}})$ is more discriminating than all other variables considered so far, which is demonstrated more explicitly in what follows.

C. ATLAS cuts

The ATLAS tau working group currently recommend a set of five fixed¹² cuts for the $Z \rightarrow \tau\tau$ signal selection [21], which are summarised in Table IV. The operating point of these cuts lies at $\epsilon_{\text{sig}} = 37.77\%$ and $\epsilon_{\text{bkg}} = 1.97\%$ for MC-generated events that pass the basic event selection criteria (Section VI).

¹² They are fixed in the sense that each cut is constant and independent of any other event variables.

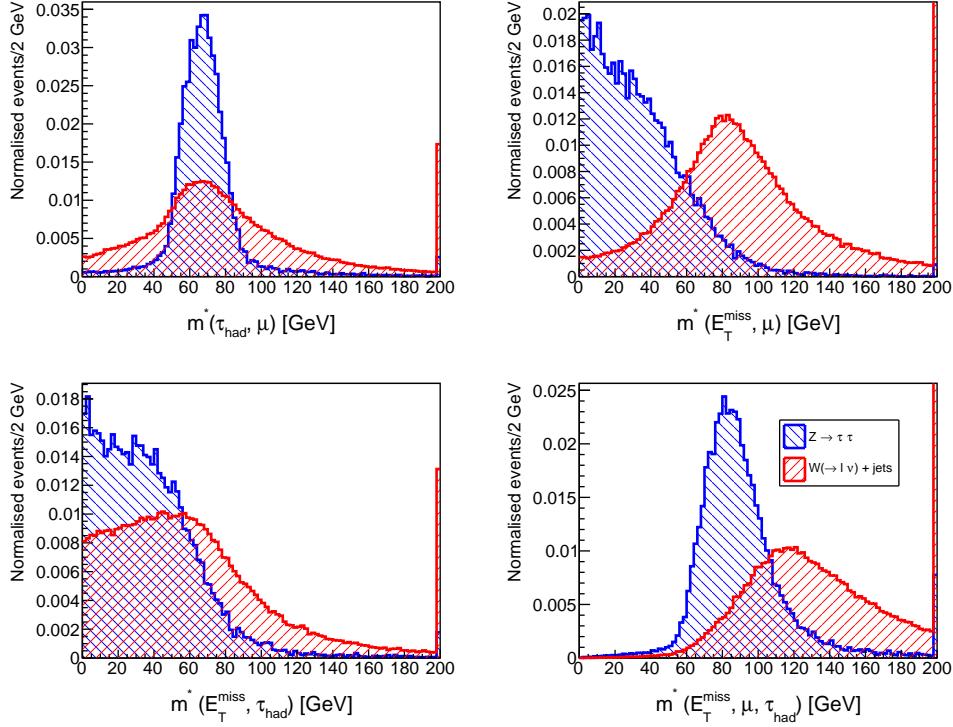


Figure 9: Distributions in the m^* -type variables, defined in Equations (16) to (19). $m^*(\mu, E_T^{\text{miss}})$ is the most discriminating variable out of the variables considered up to now.

selection cut
$m_T(\mu, E_T^{\text{miss}}) < 50 \text{ GeV}$
$45 \text{ GeV} < m_{\text{vis}}(\mu, \tau_{\text{had}}) < 85 \text{ GeV}$
SumCos > -0.15
$p_T^\mu < 40 \text{ GeV}$
$\Delta\phi(\mu, \tau_{\text{had}}) > 2.4$

Table IV: Summary of the $Z \rightarrow \tau\tau$ signal-selection cuts recommended by the ATLAS tau working group.

D. ROC curves from fixed cuts

It is convenient to visualise the performance of an event variable in its ability to discriminate between two samples as a cut on it is varied. These *efficiency curves*, also referred to as *receiver operating characteristic (ROC) curves* show for a given ϵ_{sig} the corresponding ϵ_{bkg} . The ROC curves for a range of variables are summarised in Figure 10. The magenta curve in Figure 10 shows that $m^*(\mu, E_T^{\text{miss}})$ (cf. Equation (16)) provides the best discriminating power out of the variables shown for fixed cuts that are varied along the ROC curve. It performs significantly better than the next best variable, $m_T(\mu, E_T^{\text{miss}})$ (green curve), on which it is based. The ATLAS point is shown for completeness and is not directly comparable, as it consists of a set of five cuts (see Table IV). It is important to note that the ROC curves shown in Figure 10 have been obtained by varying a fixed cut from the lower ends of the distributions shown in Figures 8 and 9, and forming the cumulative up to a given cut value. The only exception to this procedure is $m_{\text{vis}}(\mu, \tau_{\text{had}})$; its peaks for signal and background overlap (see Figure 8, top left), and instead the cuts start at the peak of the signal

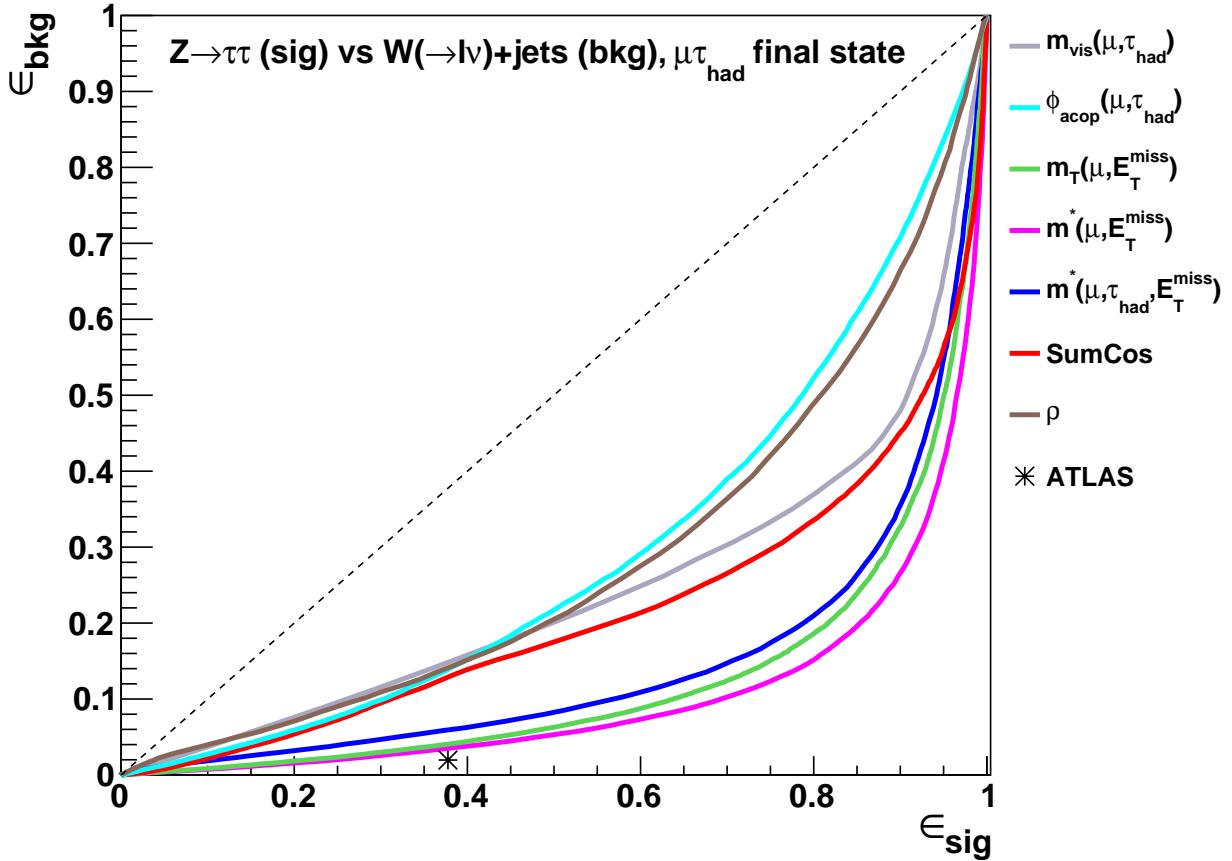


Figure 10: ROC curves obtained from varying cuts on event variables. Clearly, $m^*(\mu, E_T^{\text{miss}})$ (magenta) performs better than all other candidate variables, however not as well as the ATLAS operating point (black asterisk). Note that the black dashed line described by the line $\epsilon_{\text{bkg}} = \epsilon_{\text{sig}}$ in this one and the following figures corresponds to ineffective cuts where the background probability is the same as the signal probability.

distribution (67 GeV), and are symmetrically extended to either side until all entries are captured. The ROC curve for ρ (brown curve in Figure 10) is obtained by starting from the upper end of the ρ distribution (cf. left plot in Figure 7), as the bulk of the signal lies to the right of the background.

E. ROC curves from two-dimensional cuts

1. Rectangular fixed cuts

Having identified $m^*(\mu, E_T^{\text{miss}})$ as the best-performing single variable, the next natural step is to search for variables that possess discriminating power in addition to $m^*(\mu, E_T^{\text{miss}})$. By studying correlations between the variables given in Section VII B, it emerges that the 2D histograms showing $m_{\text{vis}}(\mu, \tau_{\text{had}})$ vs. $m^*(\mu, E_T^{\text{miss}})$ look most different for signal and background. The two distributions are shown in Figure 11, and it is clear that signal and background are peaked at very different values in the 2D plane spanned by $m_{\text{vis}}(\mu, \tau_{\text{had}})$ and $m^*(\mu, E_T^{\text{miss}})$.

In order to explore the possibility of an improved performance by combining the two variables, we consider all possible ROC curves obtained from rectangular cuts in the $m_{\text{vis}}(\mu, \tau_{\text{had}})$ - $m^*(\mu, E_T^{\text{miss}})$ plane, where two sides of the rectangle are held fixed at $m_{\text{vis}}(\mu, \tau_{\text{had}}) = m^*(\mu, E_T^{\text{miss}}) = 0$ GeV. Figure 12 shows a set of 400 such curves, where the colour is indicative of the end-point of the rectangle on the $m^*(\mu, E_T^{\text{miss}})$ -axis. It ranges from blue for a cut at the low end to yellow for a cut at the upper end of the $m^*(\mu, E_T^{\text{miss}})$ -axis. The envelope described by the outer band of

this set of curves is the best possible ROC curve attainable from two-dimensional fixed cuts on the two variables shown, starting from the lower ends of the two respective distributions. The

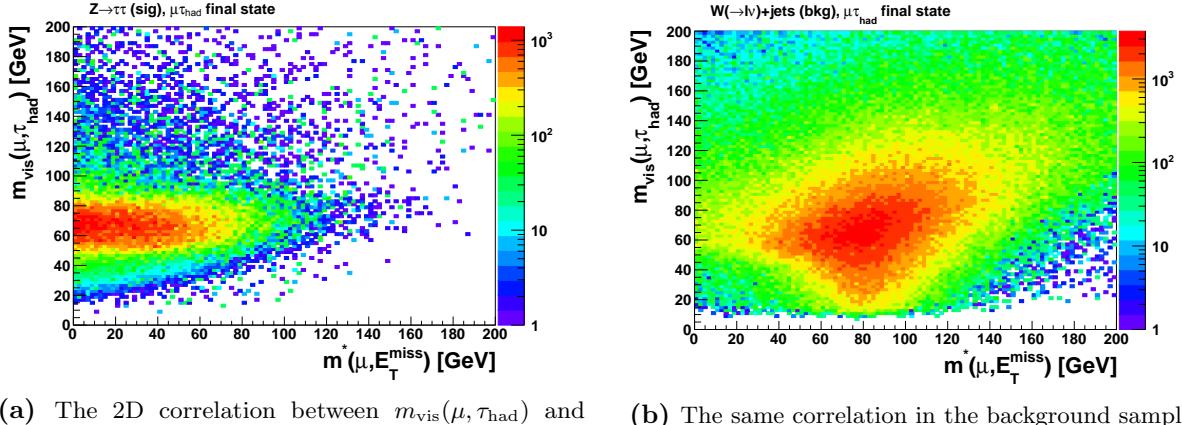


Figure 11: 2D correlations between $m_{\text{vis}}(\mu, \tau_{\text{had}})$ and $m^*(\mu, E_T^{\text{miss}})$ for signal (a) and background (blue). These plots are used as a basis for the 2D $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ plot below (Figure 13).

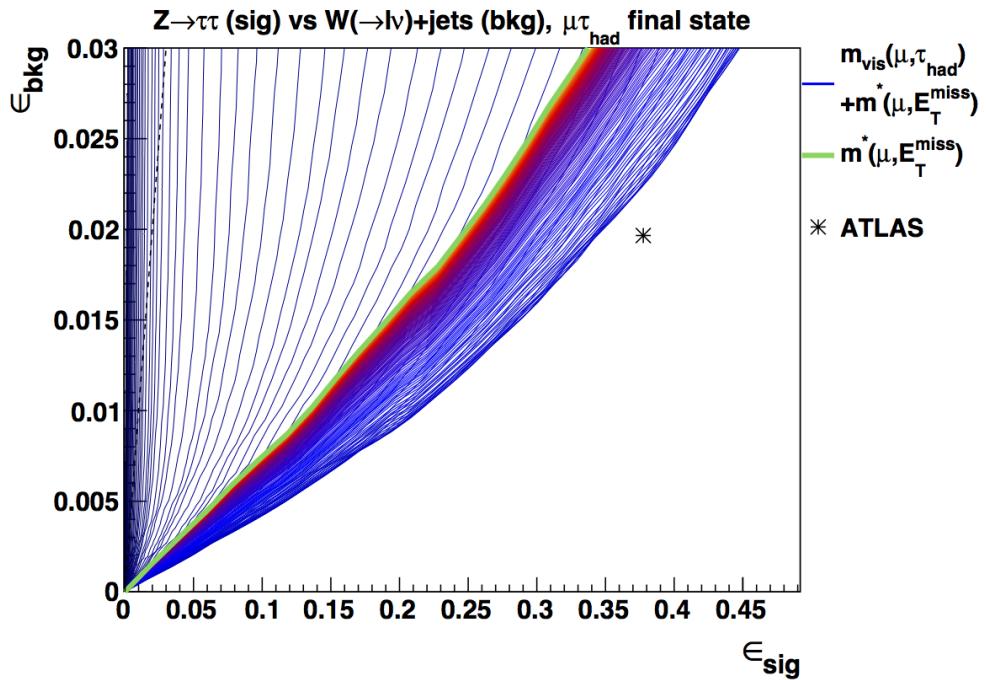


Figure 12: The curves encompass a region which corresponds to all possible rectangular cuts on the two variables with one edge held fixed at the origin. For comparison, the ROC curve obtained from simple cuts on $m^*(\mu, E_T^{\text{miss}})$ is shown in green.

outer envelope in Figure 12 performs significantly better than $m^*(\mu, E_T^{\text{miss}})$ alone, which means a substantial improvement in discriminating power can be gained from adding $m_{\text{vis}}(\mu, \tau_{\text{had}})$ as a second variable to $m^*(\mu, E_T^{\text{miss}})$.

2. Selecting regions based on 2D $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ plots

While rectangular cuts were explored in Section VII E 1, this section will lay out attempts to improve on the cuts in the $m_{\text{vis}}(\mu, \tau_{\text{had}})$ - $m^*(\mu, E_{\text{T}}^{\text{miss}})$ plane further by looking in more detail at the correlations for signal and background (cf. Figure 11). We perform a bin-by-bin division of the two histograms in Figure 11 to obtain a two-dimensional map of $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ for every value of $m_{\text{vis}}(\mu, \tau_{\text{had}})$ and $m^*(\mu, E_{\text{T}}^{\text{miss}})$, which is shown in Figure 13.

3. Linear cuts and the triangle parameter Δ

An interesting region emerges in the two-dimensional plot of $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ for $m^*(\mu, E_{\text{T}}^{\text{miss}})$ and $m_{\text{vis}}(\mu, \tau_{\text{had}})$ (Figure 13). The red area in the centre-left region looks roughly triangular in shape, which can be exploited by moving from fixed cuts to linear cuts¹³: a triangle can be constructed with two linear functions as upper and lower bounds, and a vertical line of constant $m^*(\mu, E_{\text{T}}^{\text{miss}}) = 0$ GeV as the third (left) boundary, as follows: the upper edge has a slope of $m_{\text{vis}}(\mu, \tau_{\text{had}}) = -\frac{7}{9}m^*(\mu, E_{\text{T}}^{\text{miss}})$, with a y -intercept that varies from $m_{\text{vis}}(\mu, \tau_{\text{had}}) = 70$ GeV to $m_{\text{vis}}(\mu, \tau_{\text{had}}) = 140$ GeV. The lower edge has a slope of $m_{\text{vis}}(\mu, \tau_{\text{had}}) = \frac{7}{9}m^*(\mu, E_{\text{T}}^{\text{miss}})$, with a y -intercept that varies from $m_{\text{vis}}(\mu, \tau_{\text{had}}) = 70$ GeV to $m_{\text{vis}}(\mu, \tau_{\text{had}}) = 0$ GeV.

The method for obtaining triangular cuts is such that the rightmost angle is fixed and the apex position is varied. This allows for a complete description of the triangle with a single parameter, Δ , which is proportional to the height of the triangle, measured from the y -axis ($m^*(\mu, E_{\text{T}}^{\text{miss}}) = 0$ GeV) to the apex (in units of GeV). The triangle is positioned so that as much signal as possible and as little background as possible is selected at low triangle parameter, Δ . The number of events in the differential slices of Δ is shown in Figure 15. The evolution of the triangle as Δ is varied is illustrated in Figure 14. The construction of Δ provides significant improvement over simple rectangular that were shown in Figure 12 (see Figure A.3 for a full comparison).

One such triangle at maximum $\Delta = 90$ is shown in red in Figure 13. The black boundary is the output of a BDT (see Sections IV A and VII G) using the same two variables as input: entries to the left of the boundary are classified as signal-like, and events to the right are classified as background-like. The decision boundary approaches a rough triangular shape, which means we expect the method of selecting a triangular region based on 2D $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ to perform similarly to a BDT.

4. Γ in $\rho - \Delta$ space

Having established Δ as an efficient way of capturing the discriminating power of both $m_{\text{vis}}(\mu, \tau_{\text{had}})$ and $m^*(\mu, E_{\text{T}}^{\text{miss}})$ combined, we are left with finding another variable, which again allows for further discrimination in addition to Δ . The asymmetry ratio, given in Equation (9), is the most promising candidate: its correlations with Δ look most different between signal and background (see Figure A.2). In order to find the optimal regions to cut on in ρ - Δ space, we employ a similar strategy to that used in Section VII E 3 to derive Δ . The two-dimensional efficiency plot of $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ is shown for ρ and Δ in Figure 16. The red region of high signal-to-noise in the top left corner suggests cuts on ρ that vary linearly as a function of Δ . As the number of events in this region is very small, however, an extension of this scheme is needed: the slope of the triangular cut is increased as the cut region becomes larger.

This triangular cut scheme is denoted by a single parameter, Γ , which is constructed as follows: the lower edge of the triangle (illustrated for three possible values in red in Figure 16), is varied

¹³ They are linear in the sense that a given cut on one variable is determined by another variable at that point. Alternatively, they correspond to functions of the form $y = ax + b$ in the 2D plane spanned by the two variables.

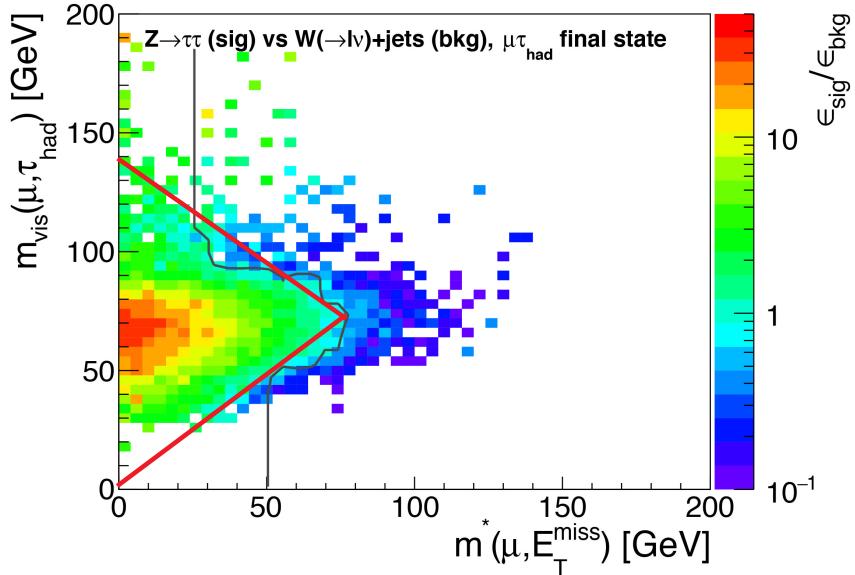


Figure 13: 2D $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ plot. The black line corresponds to a BDT decision boundary, which means events on either side of the boundary will be classified as signal-like (left) or background-like (right). Alternatively, it represents a line where it is impossible to assign an event as being more signal-like or background-like. The red triangle corresponds to the maximum extent of the Δ cut, as illustrated in Figure 14.

from $\rho = 0.01\Delta + 0.8$ to $\rho = 0.02\Delta - 1$. This variation occurs linearly in both the slope and the y -intercept. A rectangular region of $\Delta > 45$ and $\rho < 0.5$ is always excluded, as here $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ is mostly smaller than one. The number of events in differential slices of Γ is shown in Figure 17.

F. Linear correlations between variables

The linear correlations¹⁴ between all important individual variables considered are summarised in Figure 18 for signal and background, respectively. A linear correlation coefficient significantly away from zero for both signal and background means that a cut on one of the variables does not add any effective discriminating power. Figure 18 allow us to make a correspondence between the variables developed here and those used by ATLAS. p_T^μ is replaced by ρ , whereas $m^*(\mu, E_T^{\text{miss}})$ replaces $m_T(\mu, E_T^{\text{miss}})$. Figure 18 confirms that our initial choice for the first two variables to cut on is effective: $m^*(\mu, E_T^{\text{miss}})$ and $m_{\text{vis}}(\mu, \tau_{\text{had}})$ are not very correlated with one another, in both signal and background. Similarly, ρ as our third variable, is not very correlated with either $m^*(\mu, E_T^{\text{miss}})$ or $m_{\text{vis}}(\mu, \tau_{\text{had}})$.

G. BDT results

We trained various BDTs in order to compare the performance of the discrimination methods devised in this report with multivariate methods. Four BDTs were trained in total, corresponding to the set of cuts discussed above in the following fashion: BDT2 uses as input the same variables used in the construction of Δ . BDT3 uses in addition ρ , which enabled the construction of Γ . BDT4 additionally uses SumCos, corresponding to the next cut in the ATLAS selection (Table IV). The input variables of the different BDTs considered are summarised in Table V.

¹⁴ The linear correlation coefficient, ϱ , between two variables ranges from -1 to 1, where 0 means no correlation and +1(-1) means perfect correlation (anticorrelation). For two variables, X and Y , it is defined as $\varrho = \text{COV}(X, Y)/(\sigma_X \sigma_Y)$, where σ_X and σ_Y are the standard deviations of X and Y , respectively and $\text{COV}(X, Y)$ is the covariance of X and Y [25].

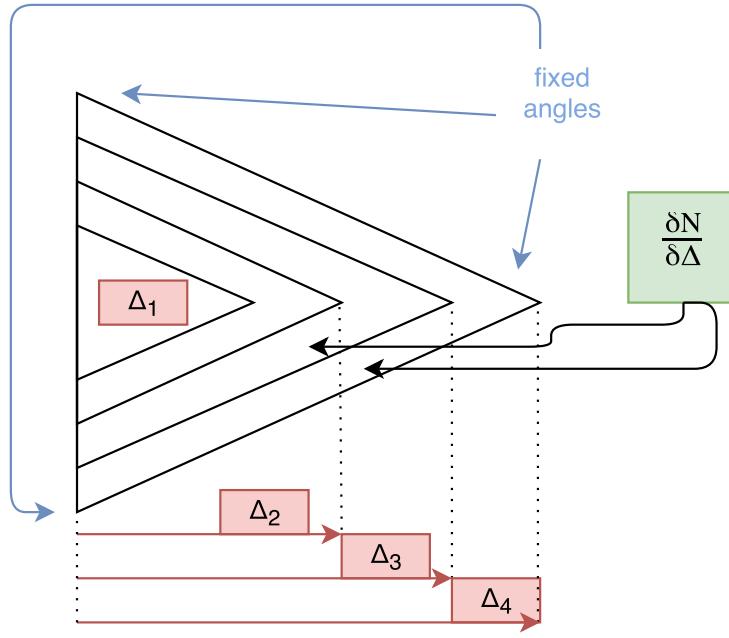


Figure 14: Schematic drawing of the change in size of the triangular cut region as the triangle parameter, Δ , is increased. The number of events captured by each slice of the triangle as Δ is increased incrementally is denoted by $\frac{\delta N}{\delta \Delta}$.

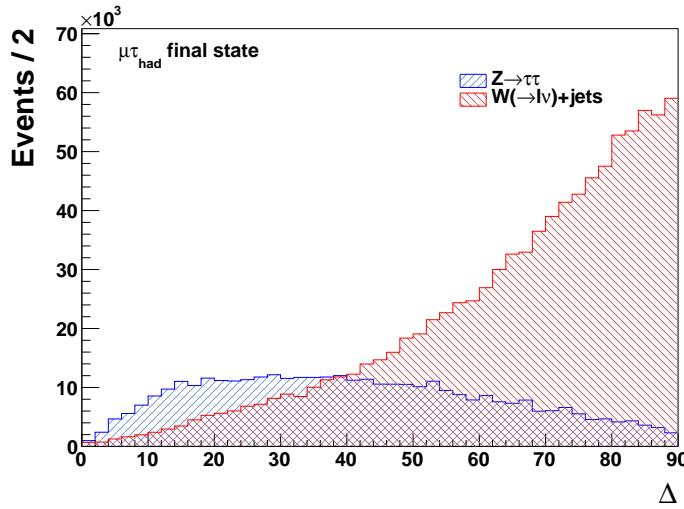


Figure 15: Histogram showing the change in the number of events captured by the triangular cut as the triangle parameter, Δ , is increased. This quantity is called $\frac{\delta N}{\delta \Delta}$ in Figure 14. As expected from Figure 13, the distribution is very signal-like for small values of Δ and becomes very background-like at high values.

More prone to training on statistical fluctuations, the most complicated BDT trained consisted of five input variables (BDT5) and was not affected by overtraining. Figure 19 shows the result of the K - S overtraining test [26, 27], showing that the number of events as a function of the BDT responses¹⁵ for the training and test samples in both signal and background agree very well with each other.

¹⁵ The BDT response is the combined classification of many decision trees of an event as signal-like (1) or background-like (-1).

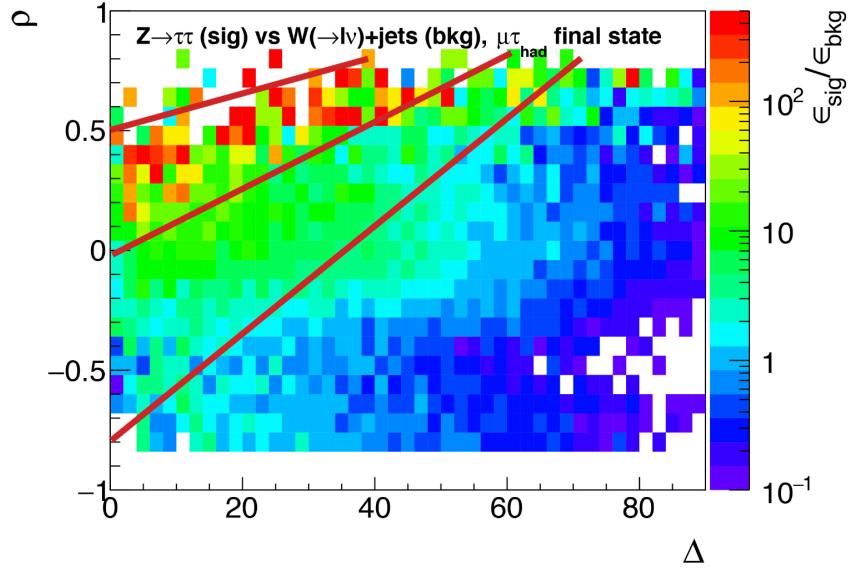


Figure 16: 2D histogram showing the ratio $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ for ρ vs Δ . The indicate three possible cuts described by the second triangle parameter, Γ .

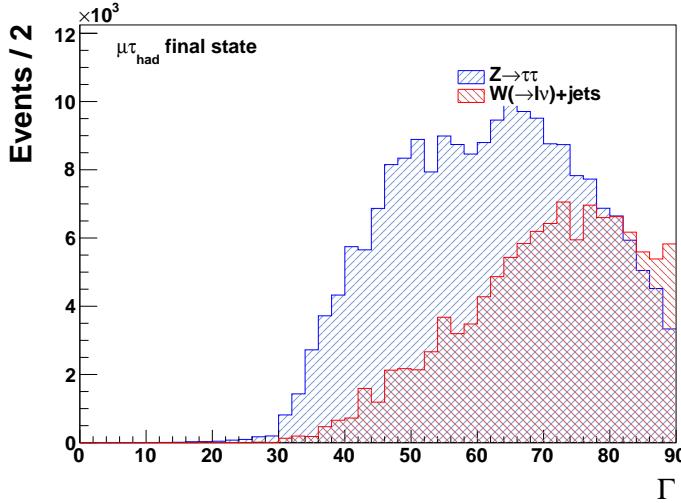


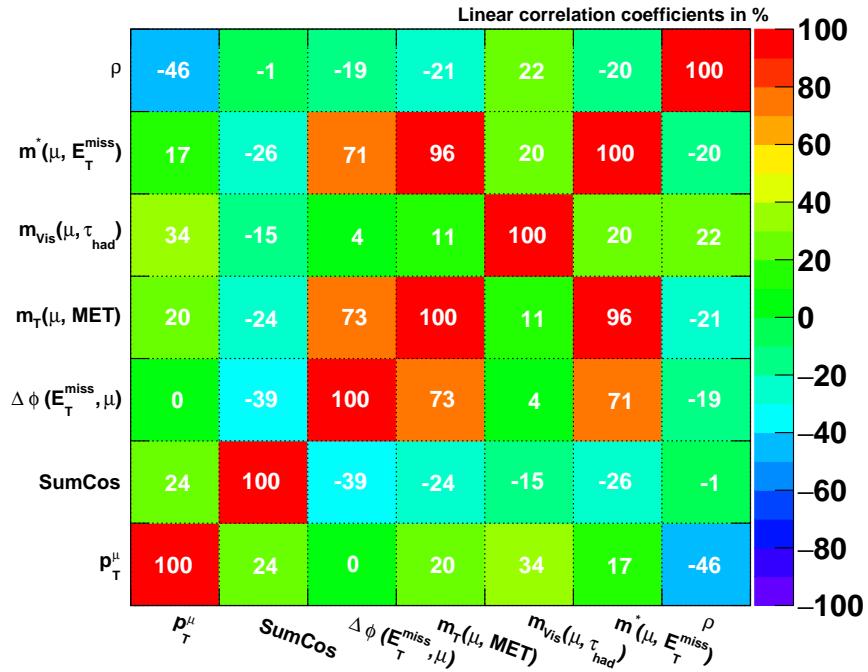
Figure 17: Evolution of the number of events as a function of the second triangle parameter, Γ . As was the case with Δ , small Γ corresponds to a high signal purity, whereas the region captured by large Γ contains roughly equal numbers of signal and background events.

H. Final results

The brown and grey curves in Figure 21 show the ROC curves obtained from cuts on Γ and Δ , respectively. While Γ uses cuts on only three variables ($m_{\text{vis}}(\mu, \tau_{\text{had}})$, $m^*(\mu, E_{\text{T}}^{\text{miss}})$, and ρ), it performs significantly better than the set of ATLAS cuts (consisting of fixed cuts on five variables). We select an operating point on this curve of $\Gamma \leq 74$, which corresponds to $\epsilon_{\text{sig}} = 38.32\%$ and $\epsilon_{\text{bkg}} = 1.73\%$, roughly matching in the signal efficiency of the ATLAS cuts. None of the variables used in the ATLAS cuts leave any additional discriminating power at this operating point. This is confirmed by the individual distributions shown in Figure 20.

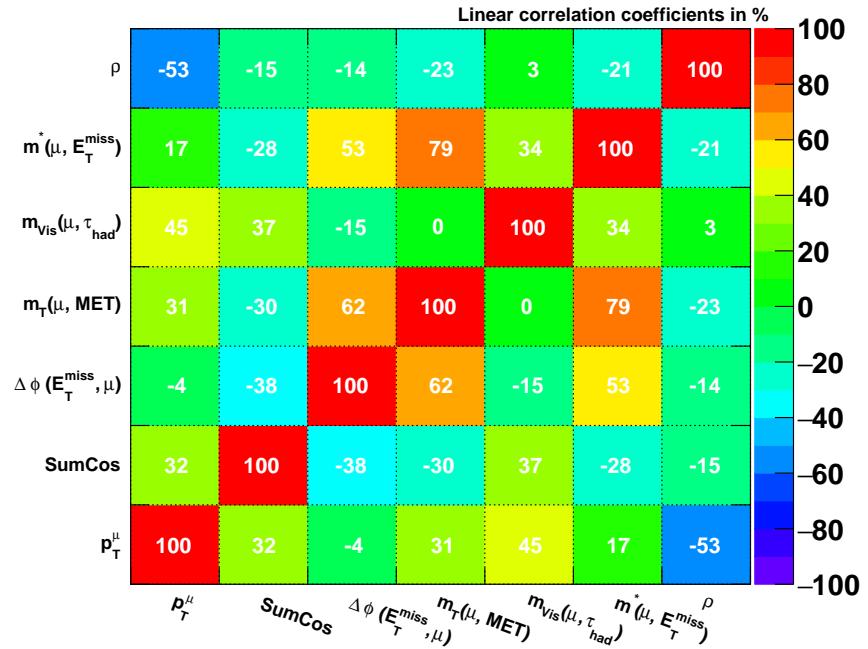
Figure 21 shows in addition the ROC curves obtained from the BDTs. BDT3 (cyan curve), which was trained using the same variables contained in Γ , performs marginally worse than Γ . The

Correlation Matrix (signal)



(a) Linear correlation between variables for signal events in percent.

Correlation Matrix (background)



(b) Linear correlation between variables for background events in percent.

Figure 18: Linear correlations between variables for signal and background events in percent. A high correlation for a pair of variables typically means that only one variable from the pair can be used, if the values are similarly high for both signal and background.

BDT2	BDT3	BDT4	BDT5
$m_{\text{vis}}(\mu, \tau_{\text{had}})$	$m_{\text{vis}}(\mu, \tau_{\text{had}})$	$m_{\text{vis}}(\mu, \tau_{\text{had}})$	$m_{\text{vis}}(\mu, \tau_{\text{had}})$
$m^*(\mu, E_{\text{T}}^{\text{miss}})$	$m^*(\mu, E_{\text{T}}^{\text{miss}})$	$m^*(\mu, E_{\text{T}}^{\text{miss}})$	$m_{\text{T}}(\mu, E_{\text{T}}^{\text{miss}})$
\times	ρ	ρ	p_{T}^{μ}
\times	\times	SumCos	SumCos
\times	\times	\times	$\Delta\phi(\mu, \tau_{\text{had}})$

Table V: Summary of input variables to the four different BDTs trained.

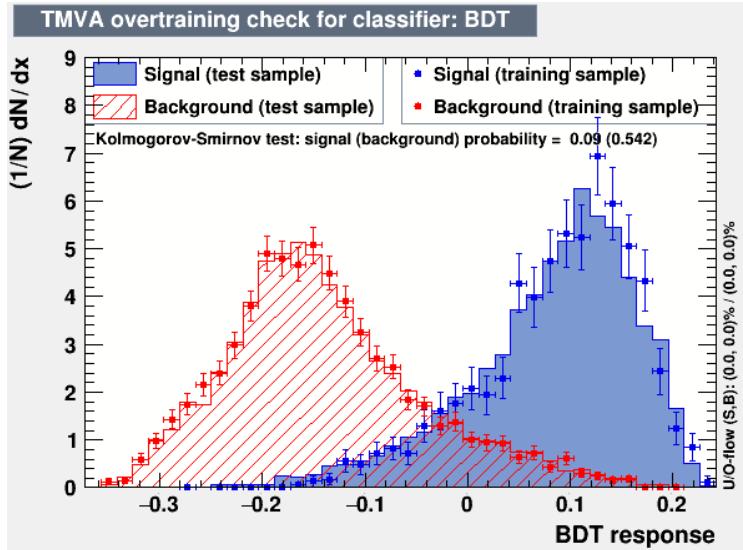


Figure 19: K-S test for BDT5. The BDT responses for the test samples match those for the training samples.

green curve of BDT4 shows that the addition of SumCos as a fourth variable provides a significant improvement in performance compared to BDT3. BDT5 (blue), which corresponds to the same set of cuts as ATLAS, performs worse than most other methods. It shows, however, that given the set of five cuts going into the $Z \rightarrow \tau\tau$ signal selection given in Table IV, the ATLAS cuts are almost ideally chosen: even a BDT algorithm with the same five input variables does not perform significantly better.

It looks as though the addition of SumCos to the three variables contained in Γ should add discriminating power. This possibility was explored, however, no obvious choice of cuts in the 2D plane spanned by SumCos and Γ was found. The underlying correlations for signal and background, as well as the the 2D $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ plot are given in Figures A.4 and A.5.

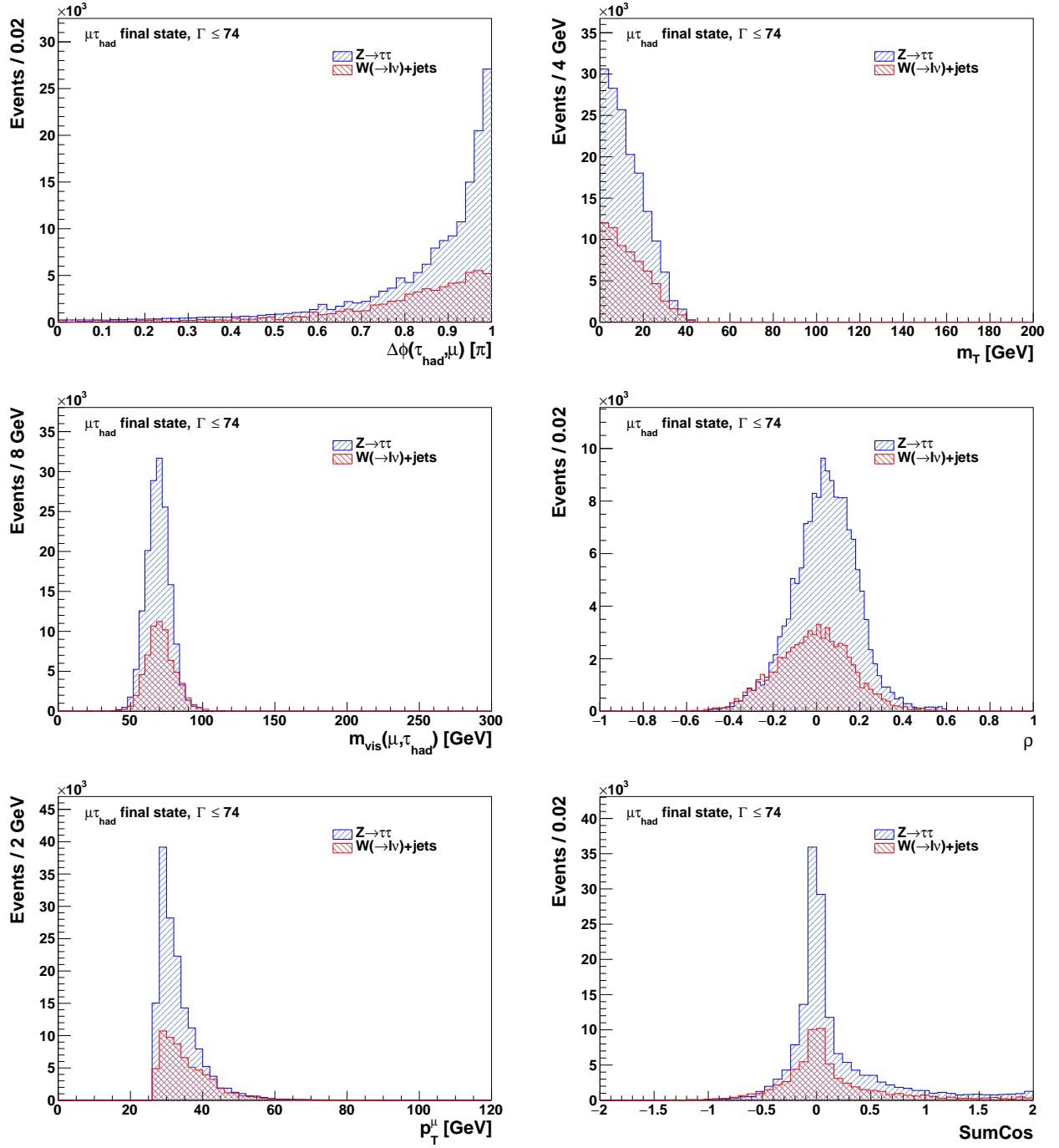


Figure 20: Distributions in the five variables used by ATLAS after having cut on Γ , as well as the distribution in ρ after the same cut (centre right). It is evident that none of the distributions show a region in which a linear cut would improve the performance. Additional distributions at this operating point can be found in Figures A.9 and A.10.

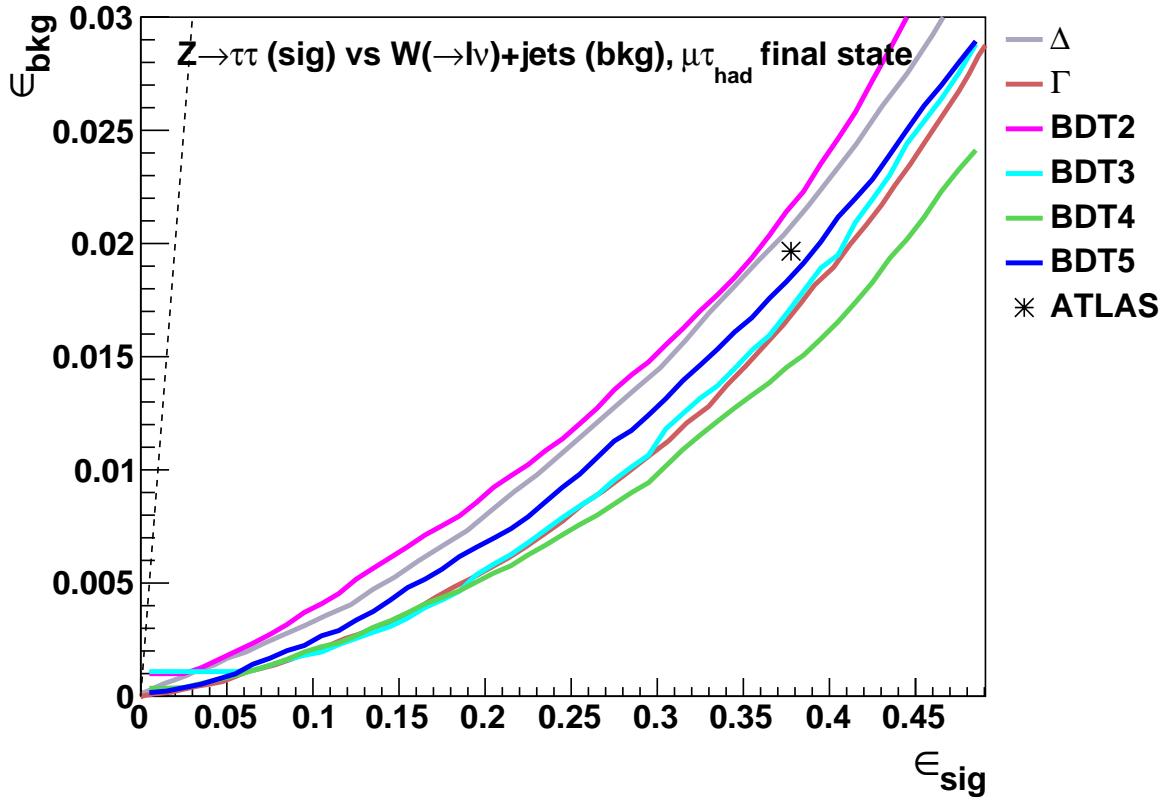


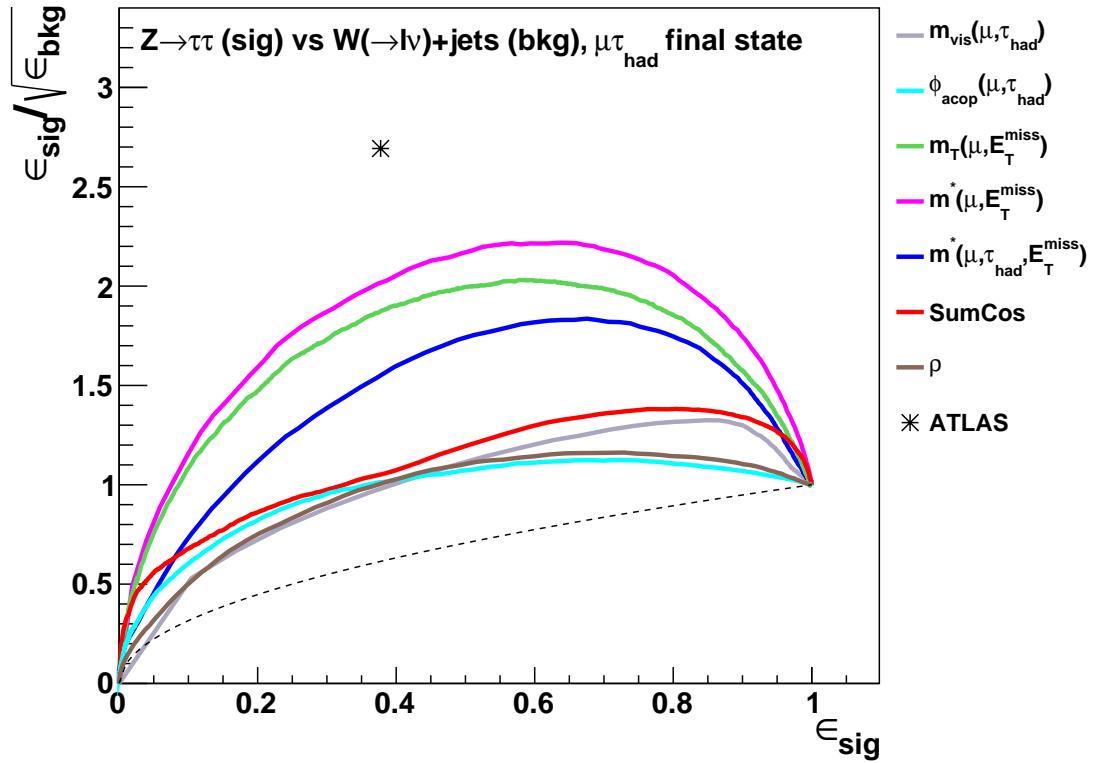
Figure 21: ROC curves obtained from cuts on Δ (grey), Γ (brown), as well as four versions of the BDT. Obtained from cuts on two variables ($m_{\text{vis}}(\mu, \tau_{\text{had}})$ and $m^*(\mu, E_T^{\text{miss}})$), Δ performs better than a BDT using the same two variables (BDT2, pink). Similarly, the three-variable Γ matches the performance of a BDT using the same three variables (BDT3, cyan).

1. The significance improvement characteristic, SIC

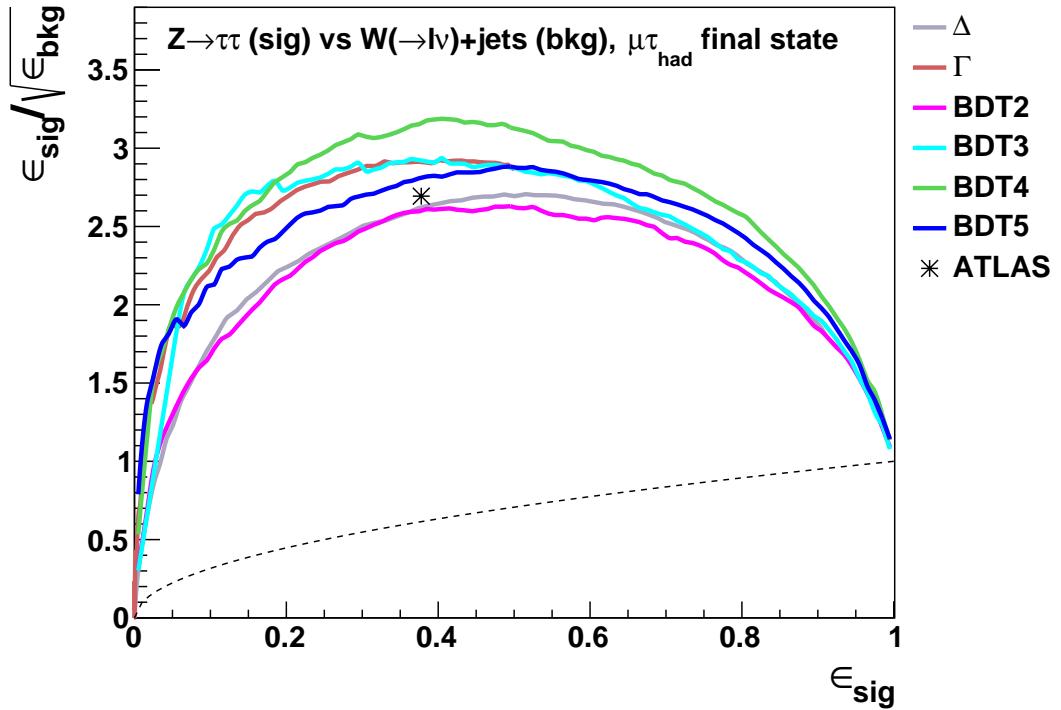
A useful way of visualising efficiencies and comparing in absolute terms the performance of different variables is the *significance improvement characteristic* (SIC) [28], which is defined as

$$\text{SIC} = \frac{\epsilon_{\text{sig}}}{\sqrt{\epsilon_{\text{bkg}}}}. \quad (20)$$

The maximum value of SIC as a function of ϵ_{sig} , $\overline{\text{SIC}}$, is helpful in ranking variables according to their optimal performance. Due to their exhibiting a clear maximum as a function of ϵ_{sig} , SIC curves are often easier to interpret than efficiency curves. The SIC curves corresponding to the primary variables of Figure 10 are shown in Section VII H 1, whereas those of the multivariate methods of Figure 21 are given in Section VII H 1.



(a) SIC curves obtained from the primary-variable ROC curves of Figure 10.



(b) SIC curves obtained from the multivariate ROC curves shown in Figure 21.

Figure 22: SIC curves obtained from ROC curves of primary variables (a) and multivariate methods (b). The SIC is defined as $SIC = \epsilon_{sig}/\sqrt{\epsilon_{bkg}}$.

As a measure of absolute performance, the $\overline{\text{SIC}}$ and its corresponding ϵ_{sig} for all single variables and multivariate methods considered up to now are presented in Table VI. As established before, $m^*(\mu, E_T^{\text{miss}})$ performs significantly better than all other variables out of the seven variables considered. In terms of multivariate methods, BDT4 ranks first as the best choice of algorithm, followed closely by BDT3 and Γ , whose performance can be considered identical. The ATLAS variables perform worse than Δ , and only slightly better than BDT2.

Primary variables				Multivariate methods			
rank	variable	SIC	ϵ_{sig}	rank	method	SIC	ϵ_{sig}
1	$m^*(\mu, E_T^{\text{miss}})$	2.220	0.635	1	BDT4	3.188	0.405
2	$m_T(\mu, E_T^{\text{miss}})$	2.032	0.578	2	BDT3	2.936	0.405
3	$m^*(\mu, E_T^{\text{miss}}, \tau_{\text{had}})$	1.836	0.675	3	Γ	2.923	0.402
4	SumCos	1.382	0.803	4	BDT5	2.882	0.495
5	$m_{\text{vis}}(\mu, \tau_{\text{had}})$	1.324	0.848	5	Δ	2.706	0.515
6	ρ	1.162	0.729	6	ATLAS	2.694	0.378
7	ϕ_{acop}	1.126	0.722	7	BDT2	2.630	0.495

Table VI: Ranking of single variables and multivariate methods according to their maximum SIC value, $\overline{\text{SIC}}$.

We have thus obtained an effective event variable, Γ , which exceeds in its discrimination power the set of ATLAS cuts listed in Table IV. A simpler triangular cut on the 2D plane spanned by $m_{\text{vis}}(\mu, \tau_{\text{had}})$ and $m^*(\mu, E_T^{\text{miss}})$, described by Δ , surpasses the ATLAS cuts in terms of $\overline{\text{SIC}}$. The BDT5 ROC curve in Figure 21 shows that the ATLAS cuts are sensibly chosen if exactly the five variables therein are selected.

VIII. DATA-MC COMPARISON

Having developed kinematic variables and selection cuts on the basis of physics simulations, it remains to be shown that the MC accurately reproduces the experimental data. To this end, six distributions in kinematic variables are shown in Figure 23, comparing the full ATLAS dataset of 2015 and 2016 with MC predictions. While all distributions suffer from a slight excess in the number of simulated events compared to data, there are no obvious systematic trends hinting at mismodelled variables. Due to a lack of access to recorded data that were used to create the distributions shown in Figure 23 at the time of writing, other variables considered such as Δ and Γ cannot yet be shown. It can, however, be reasonably assumed that these variables are modelled similarly accurately, as they are constructed out of the variables shown.

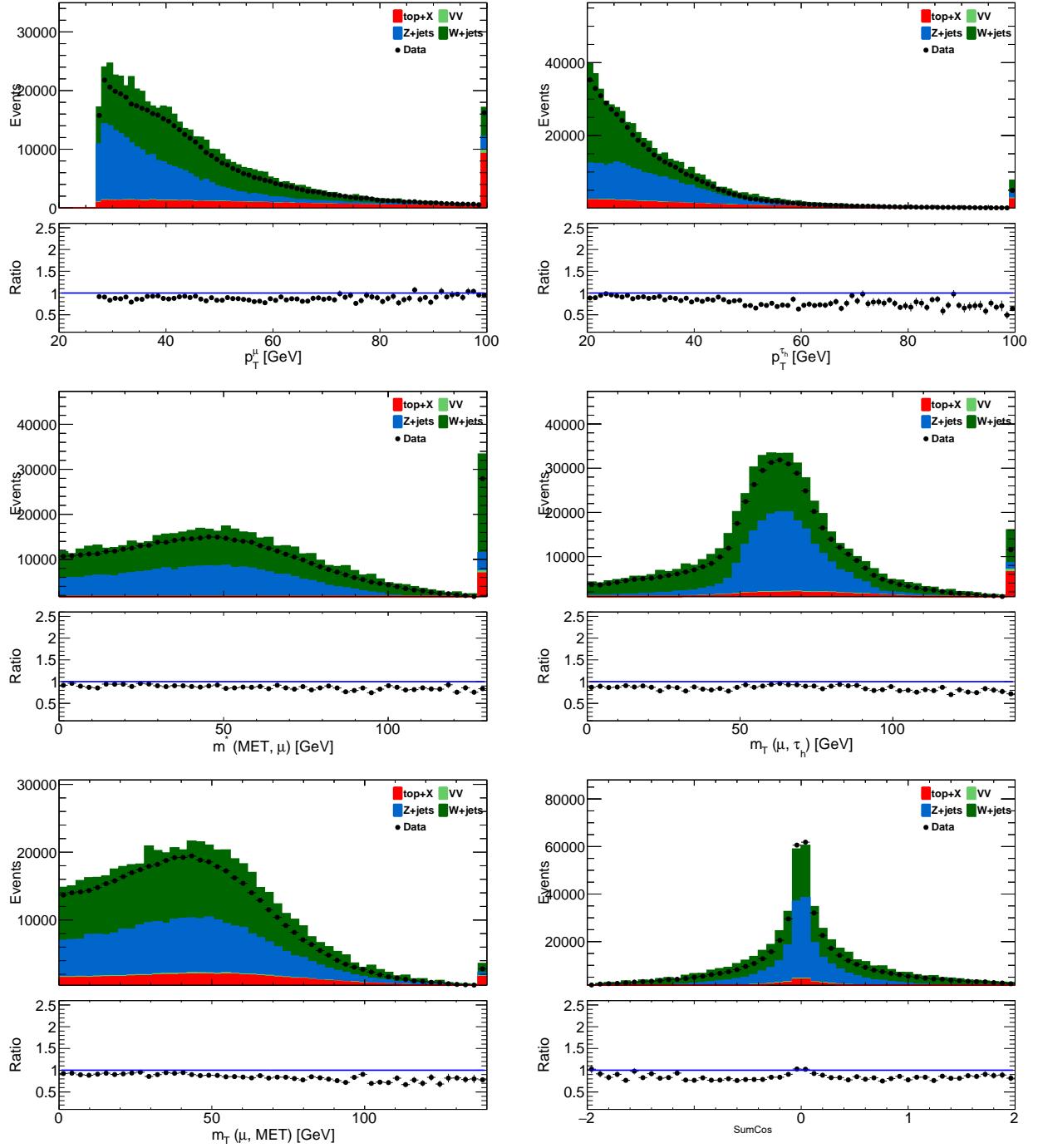


Figure 23: Data-MC comparison plots in six kinematic distributions. The data shown in black correspond to an integrated luminosity of \mathcal{L}_{int} . The overall excess of the MC is likely to be due to a flawed normalisation procedure, as there are no systematic trends.

IX. CONCLUSION AND FUTURE WORK

Obtaining pure samples of Z -boson decays into two tau leptons is important in the experimental verification of theories beyond the standard model. Heavier cousins of the Z boson (Z') could couple favourably to tau leptons in many models. In this work, the SM decay of $Z \rightarrow \tau\tau$ was investigated, with a τ_{had} and a muon as final-state particles. The main source of background in this decay comes from $W+\text{jets}$ events, in which the W decays into a μ - ν_μ pair, and one of the jets fakes a τ_{had} (jet $\rightarrow\tau_{\text{fake}}$). Various angular and mass-related variables were assessed with regard to their ability to discriminate between these two sets of events.

When constrained to the use of one variable, we propose the use of $m^*(\mu, E_T^{\text{miss}})$ as the most discriminating single variable. Constructed from a boost into the centre-of-mass frame of the decaying $\tau\tau$ system, it performs far better than its non-boosted analogue, $m_T(\mu, E_T^{\text{miss}})$. If two event variables are cut on simultaneously, Δ , which is constructed from triangular cuts in the $m_{\text{vis}}(\mu, \tau_{\text{had}})$ - $m^*(\mu, E_T^{\text{miss}})$ plane, provides much better discriminating power, and approaches in its effectiveness the set of five variables currently recommended by the ATLAS tau working group. After adding the p_T -asymmetry ratio between the muon and the hadronic tau called ρ , a third variable, Γ , was developed. It can be constructed from triangular cuts in the ρ - Δ plane, and exceeds in its effectiveness the ATLAS cuts and performs equally as well as a BDT algorithm with the same three input variables.

In future work, it will be important to verify using recorded collision data that the newly proposed variables, Δ , and Γ , are accurately described by the MC simulations. Another important avenue of work will be to perform a data-driven background estimation in order to assess the contamination of the signal from multijet events. The excess of simulated events compared with data remains to be cleared up.

It will be interesting to resolve up why the addition of SumCos as a fourth variable into a BDT algorithm creates a significant gain in performance, despite the fact that there does not seem to be any remaining discriminating information after applying a cut on Γ in this or any other of the variables considered.

ACKNOWLEDGMENTS

I am very grateful to my supervisor, Prof. Terry Wyatt, for his guidance and relentless support throughout this past year.

I would like to thank Dr. Rustem Ospanov and Rhys Roberts for providing their advice and the code base for this project, as well as Sam Dysch for sharing his insight into tau analyses. I would also like to thank the Manchester HEP group in general for welcoming me into the group. Finally, I am grateful to Vilius Čepaitis for being a fantastic project partner.

- [1] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. URL <http://arxiv.org/pdf/1207.7214.pdf>.
- [2] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. URL <https://arxiv.org/pdf/1207.7235v2.pdf>.
- [3] V Barger, Wai-Yee Keung, and Ernest Ma. Sequential W and Z bosons. *Physics Letters B*, 94(3):377–380, 1980.
- [4] LHCb Collaboration. Test of lepton universality using $B^+ \rightarrow K^+\ell^+\ell^-$ decays. *Physical review letters*, 113(15):151601, 2014. URL <https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.113.151601>.
- [5] ATLAS Collaboration. A search for high-mass resonances decaying to $\tau\tau$ in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 7:1–44, 2015. URL <https://arxiv.org/abs/1502.07177>.
- [6] ATLAS Collaboration. Search for Minimal Supersymmetric Standard Model Higgs bosons H/A and for a Z' boson in the $\tau\tau$ final state produced in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. *Eur. Phys. J.*, C76(11):585, 2016. doi:10.1140/epjc/s10052-016-4400-6. URL <https://arxiv.org/pdf/1608.00890.pdf>.
- [7] Pierre Fayet. Supergauge invariant extension of the higgs mechanism and a model for the electron and its neutrino. *Nuclear Physics B*, 90:104–124, 1975.
- [8] Pierre Fayet. Supersymmetry and weak, electromagnetic and strong interactions. *Physics Letters B*, 64(2):159–162, 1976.
- [9] Pierre Fayet. Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions. *Physics Letters B*, 69(4):489–494, 1977.
- [10] Rene Brun and Fons Rademakers. ROOT - an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1):81–86, 1997. URL <https://arxiv.org/pdf/1508.07749.pdf>.
- [11] ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, 2008. URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08003>.
- [12] Tanju Gleisberg, Stefan Höche, Frank Krauss, Andreas Schälicke, Steffen Schumann, and Jan-Christopher Winter. Sherpa 1. α , a proof-of-concept version. *Journal of High Energy Physics*, 2004(02):056, 2004. URL <https://arxiv.org/pdf/hep-ph/0311263.pdf>.
- [13] Sea Agostinelli, John Allison, K al Amako, J Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. URL <http://inspirehep.net/record/593382/files/fermilab-pub-03-339.pdf>.
- [14] Ricardo Wölker. MPPhys interim report. Submitted on 8/1/17. URL https://cernbox.cern.ch/index.php/apps/files/ajax/download.php?dir=%2FDocuments&files=W%C3%B6lker_MPPhys_interim_report_submitted.pdf.
- [15] Table of cross-sections and MC sample information. URL https://svnweb.cern.ch/trac/manc/browser/PhysicsNtuple/PhysicsTaus/trunk/data/Xsection13TeV_tth_bkg_v1.txt. Accessed on 13/5/17.
- [16] Pushpalatha C Bhat. Multivariate analysis methods in particle physics. *Annual Review of Nuclear and Particle Science*, 61:281–309, 2011. URL <http://lss.fnal.gov/archive/2010/pub/fermilab-pub-10-054-e.pdf>.
- [17] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.
- [18] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, Helge Voss, M Backes, T Carli, O Cohen, A Christov, et al. Tmva-toolkit for multivariate data analysis. *arXiv preprint physics/0703039*, 2007. URL <https://arxiv.org/pdf/physics/0703039.pdf>.
- [19] Particle Data Group. Review of particle physics. *Physical Review D*, 86(1):010001, 2012. URL <http://pdg.lbl.gov/2016/download/rpp-2016-booklet.pdf>.
- [20] Martin L Perl. The tau lepton. *Annual Review of Nuclear and Particle Science*, 30(1):299–335, 1980.
- [21] The ATLAS Collaboration. Identification and energy calibration of hadronically decaying tau leptons with the atlas experiment in pp collisions at $\sqrt{s} = 8$ TeV. *The European Physical Journal C*, 75(7):1–33, 2015. URL <https://arxiv.org/pdf/1412.7086.pdf>.
- [22] Zinonas Zinonos. Reconstruction and identification of hadronic decays of tau leptons in ATLAS. *arXiv preprint arXiv:1409.0343*, 2014. URL <https://arxiv.org/pdf/1409.0343.pdf>.
- [23] A Banfi, S Redford, M Vesterinen, P Waller, and TR Wyatt. Optimisation of variables for studying dilepton transverse momentum distributions at hadron colliders. *The European Physical Journal C*, 71(3):1–7, 2011. URL <https://arxiv.org/pdf/1009.1580.pdf>.
- [24] J Smith, WL Van Neerven, and Jos AM Vermaseren. Transverse mass and width of the W boson. *Physical Review Letters*, 50(22):1738, 1983. URL https://www.researchgate.net/profile/John_Smith427/publication/241324805_Transverse_Mass_and_Width_of_the_W_Boson/links/54872a420cf268d28f071d3a.pdf.
- [25] Allen L Edwards. An introduction to linear regression and correlation. pages 33–46, 1976.
- [26] Andrej N Kolmogorov. *Sulla determinazione empirica di una legge di distribuzione*. na, 1933.
- [27] Nickolay Smirnov. Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics*, 19(2):279–281, 1948.
- [28] Jason Gallicchio, John Huth, Michael Kagan, Matthew D Schwartz, Kevin Black, and Brock Tweedie. Multivariate

- discrimination and the Higgs+ W/Z search. *Journal of High Energy Physics*, 2011(4):1–41, 2011. URL <http://arxiv.org/pdf/1010.3698.pdf>.
- [29] Jason Gallicchio, John Huth, Michael Kagan, Matthew D. Schwartz, Kevin Black, and Brock Tweedie. Multivariate discrimination and the Higgs+W/Z search. *Journal of High Energy Physics*, 2011(4):69, 2011. ISSN 1029-8479. doi:10.1007/JHEP04(2011)069. URL [http://dx.doi.org/10.1007/JHEP04\(2011\)069](http://dx.doi.org/10.1007/JHEP04(2011)069).
- [30] Central MC15 Production List. URL <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CentralMC15ProductionList>. Accessed on 13/5/17.

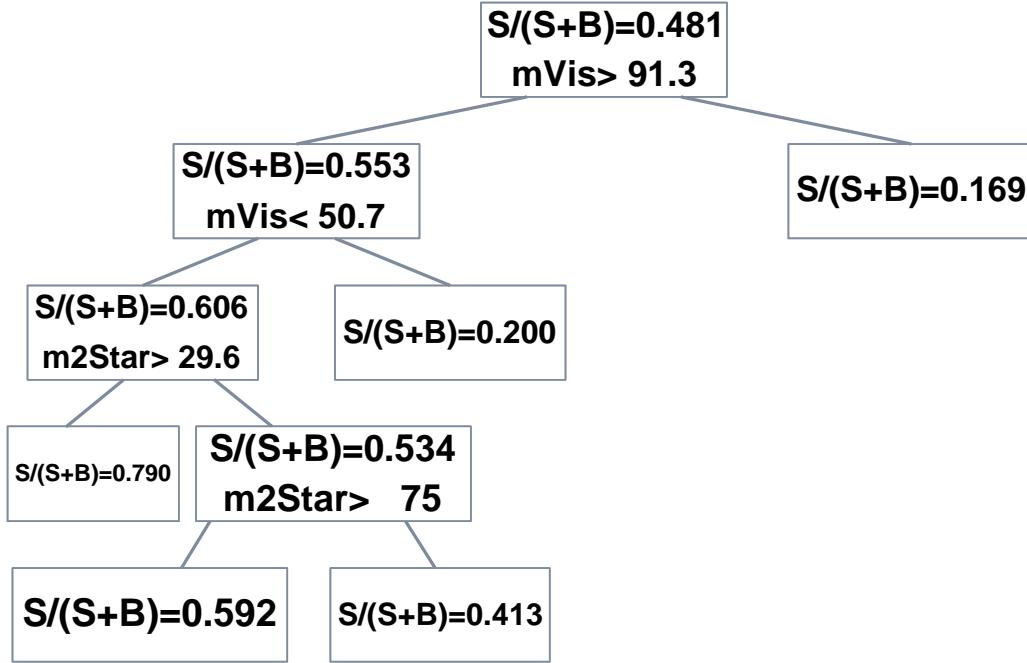
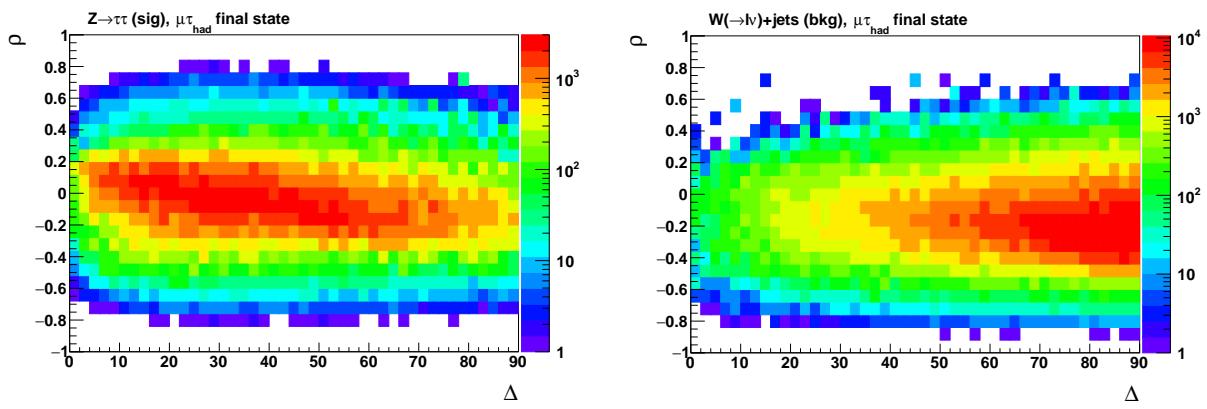
Appendix A: Supplementary figures


Figure A.1: Diagram exemplifying the output of a BDT. Events are classified into two subsets at the root node into those that pass the cut and those that fail it. At the next node, the second-most discriminating variable is selected and linearly cut on so as to achieve the best separation between the two samples, and so on. $S/(S + B)$ denotes the the number of signal events divided by all events, and is also called the *purity*.



(a) The correlation between ρ and Δ for signal events. (b) The same correlation in the background sample.

Figure A.2: 2D correlations between ρ and Δ for signal (a) and background (b). These plots are used as a basis for the construction of the second triangle parameter, Γ .

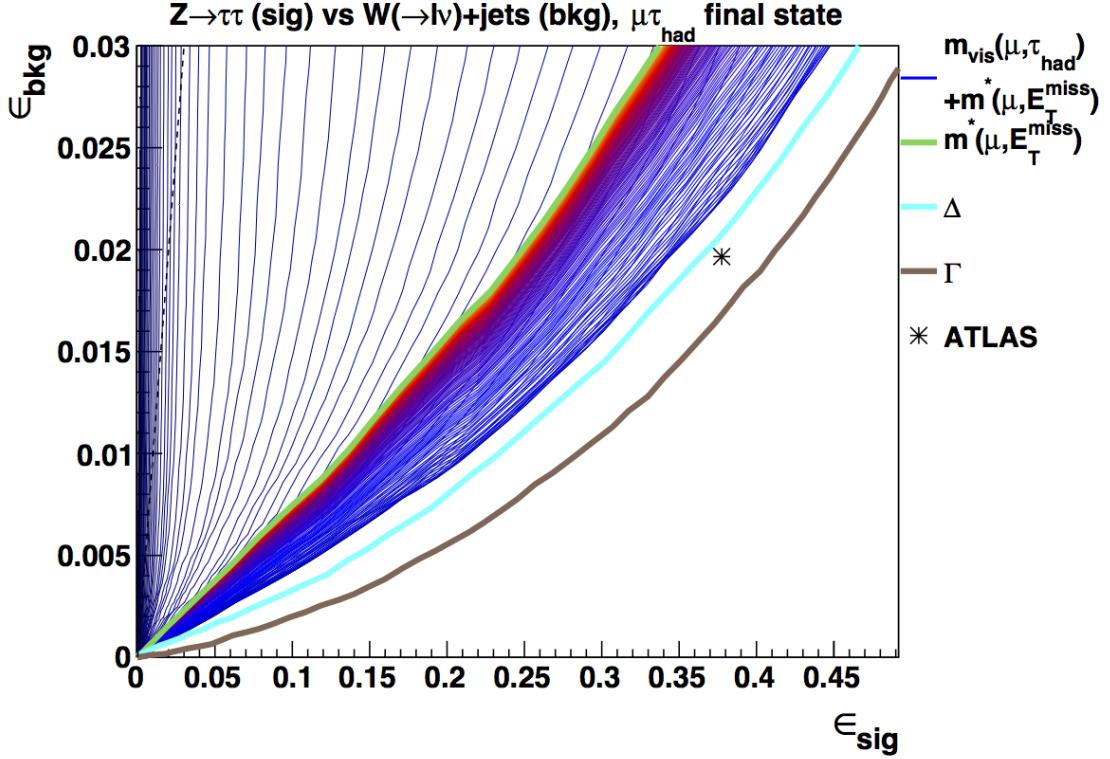
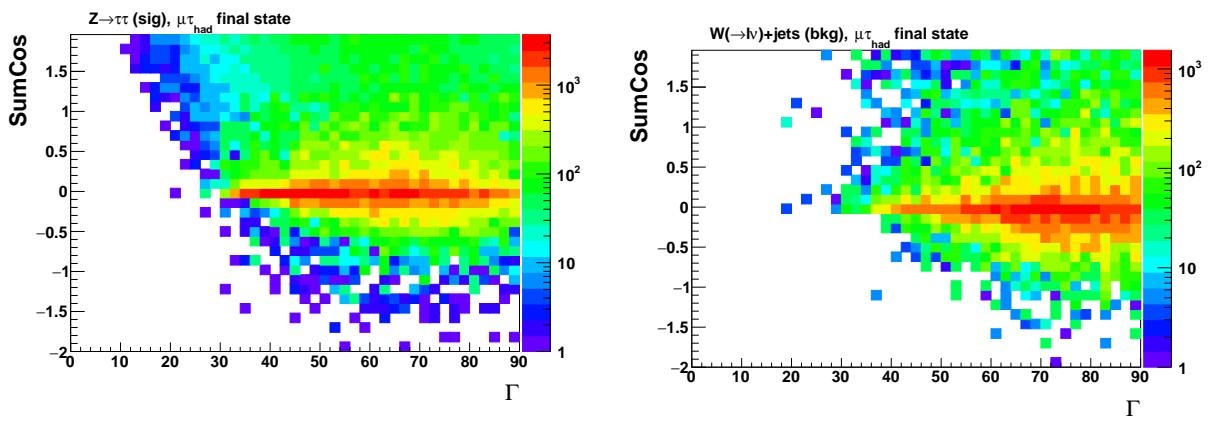


Figure A.3: Plot showing the gain in performance with each variable added to $m^*(\mu, E_{\text{T}}^{\text{miss}})$. The ROC curves shown in shades of blue to yellow are obtained from all possible fixed cuts on $m_{\text{vis}}(\mu, \tau_{\text{had}})$ and $m^*(\mu, E_{\text{T}}^{\text{miss}})$. The remaining ones show $m^*(\mu, E_{\text{T}}^{\text{miss}})$ alone (green), as well as Δ (cyan) and Γ (brown).



(a) The correlation between SumCos and Γ for signal events.

(b) The same correlation in the background sample.

Figure A.4: 2D correlations between SumCos and Γ for signal (a) and background (b). These plots are used as a basis for Figure A.5.

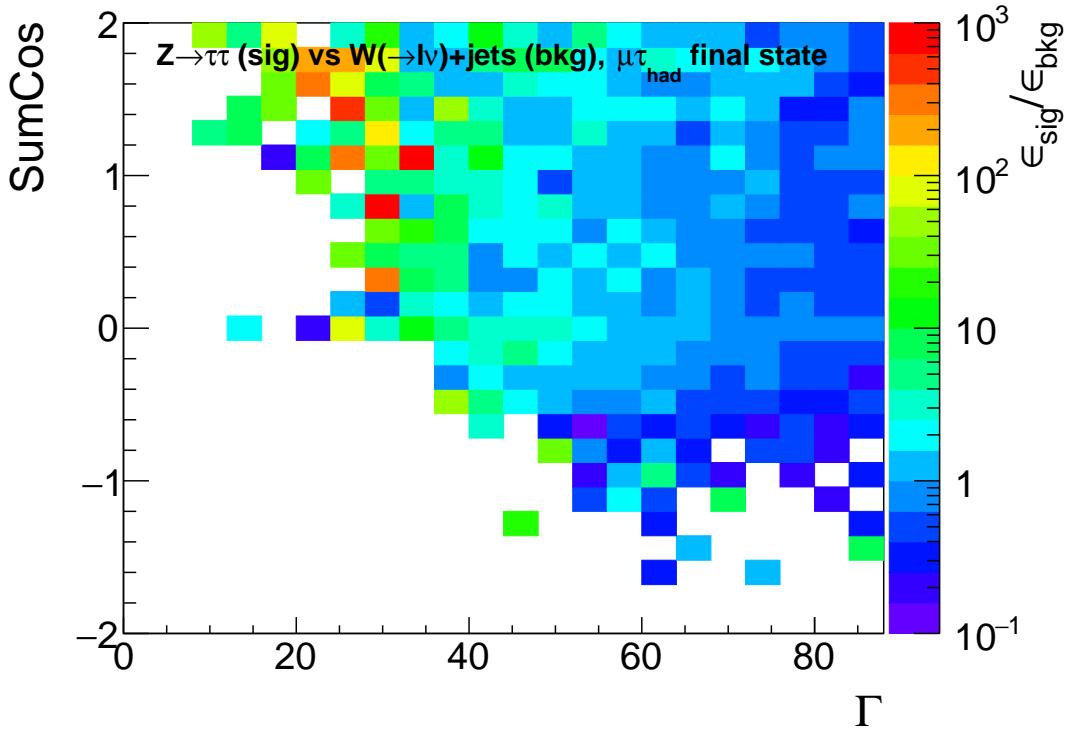


Figure A.5: 2D histogram showing the ratio $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ for SumCos vs Γ . This plot does not immediately motivate a certain shape for a cut in SumCos- Γ space.

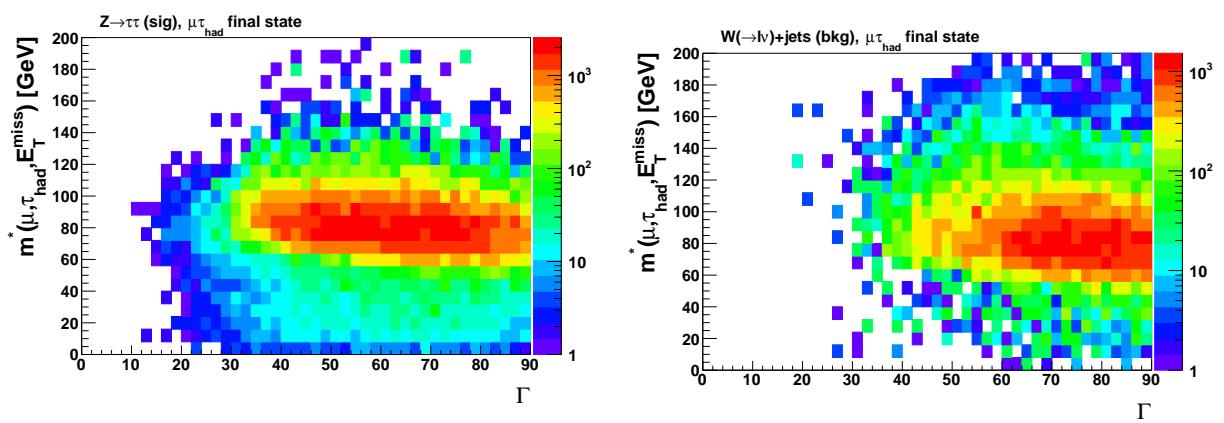


Figure A.6: 2D correlations between $m^*(\mu, \tau_{\text{had}}, E_T^{\text{miss}})$ and Γ for signal (a) and background (b). These plots are used as a basis for Figure A.7.

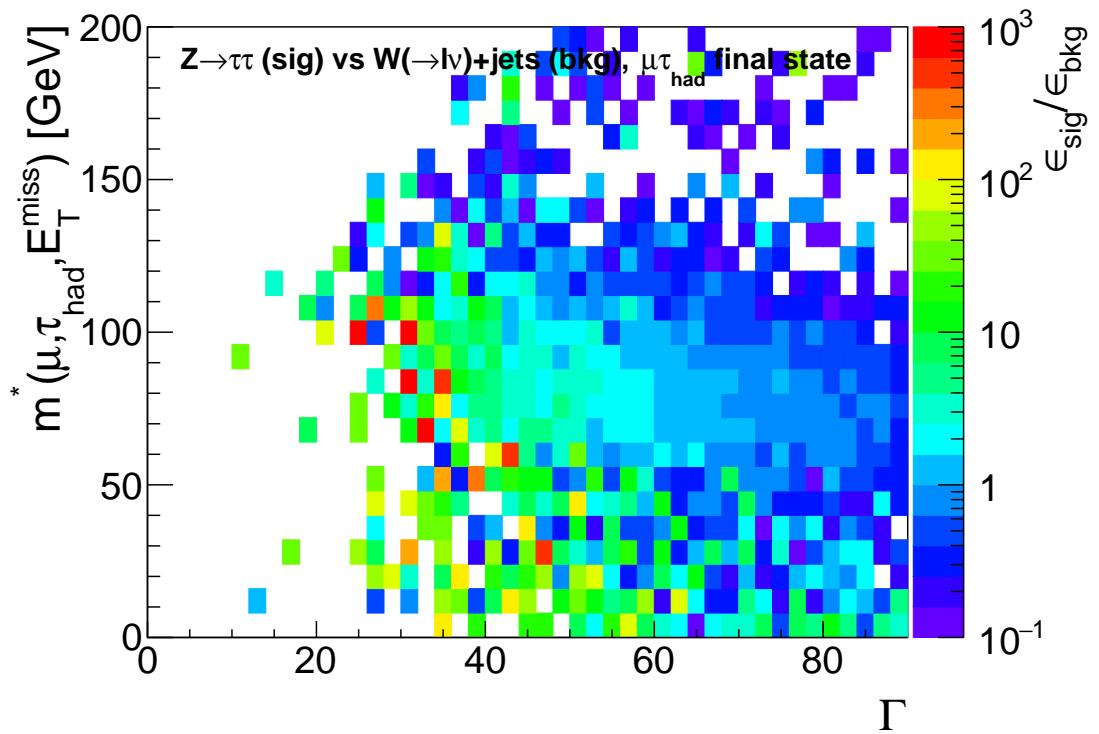
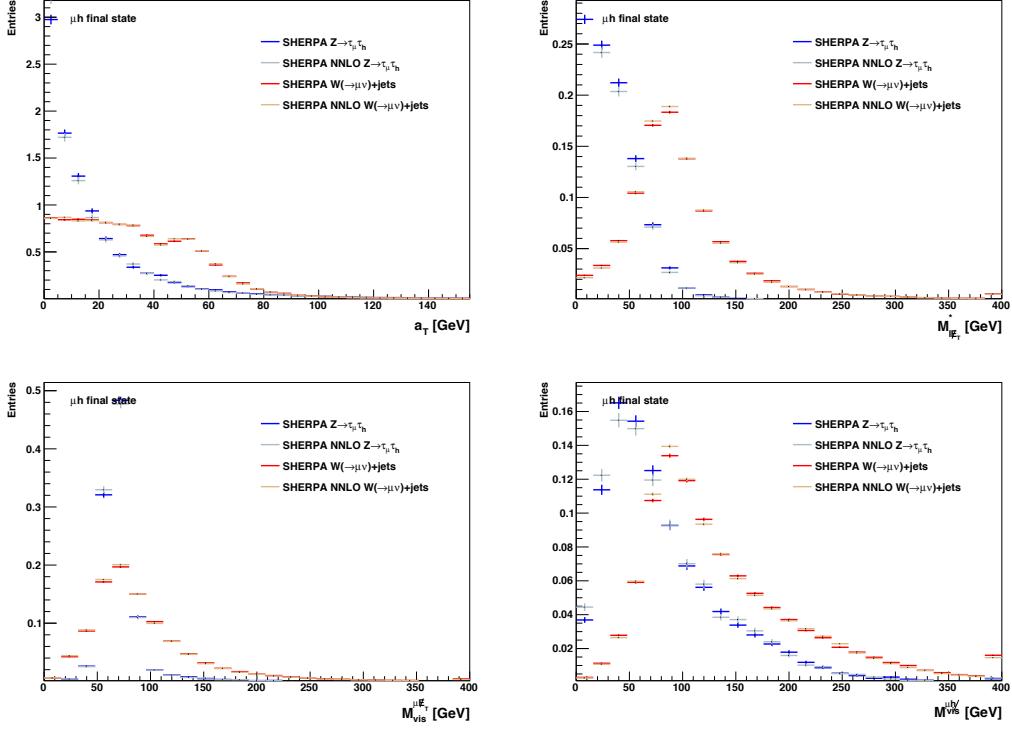
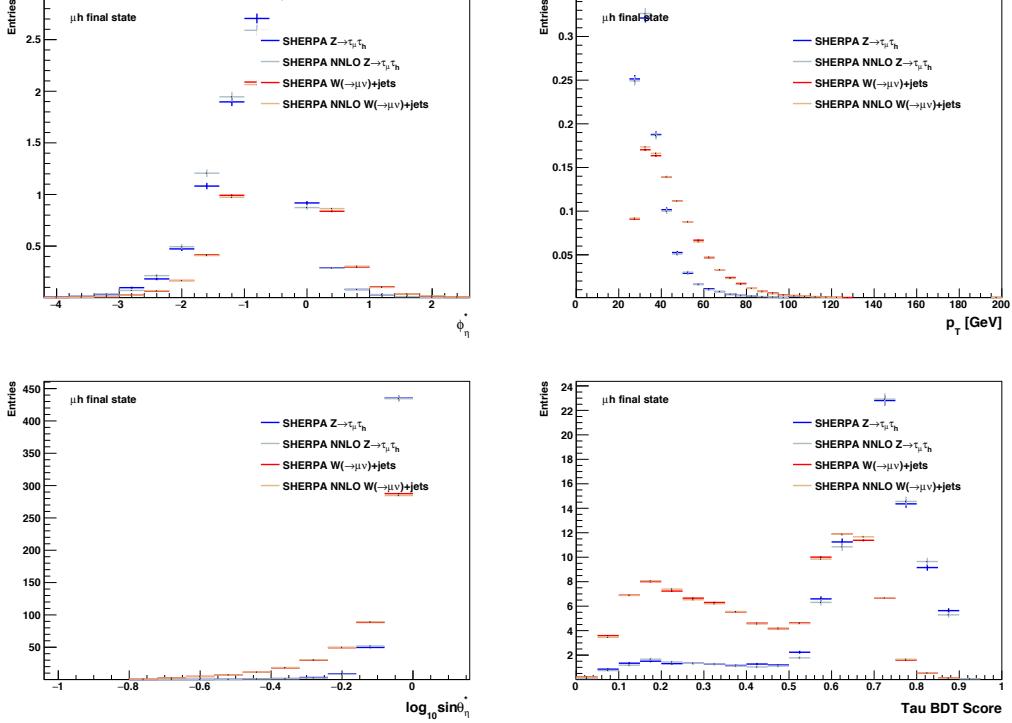


Figure A.7: 2D histogram showing the ratio $\epsilon_{\text{sig}}/\epsilon_{\text{bkg}}$ for $m^*(\mu, \tau_{\text{had}}, E_T^{\text{miss}})$ vs Γ . This plot does not immediately suggest a certain shape for a cut in the $m^*(\mu, \tau_{\text{had}}, E_T^{\text{miss}})$ - Γ plane.

A SUPPLEMENTARY FIGURES



(a) Distributions in a_T , $m^*(\mu, E_T^{\text{miss}})$, $m_{\text{vis}}(\mu, E_T^{\text{miss}})$, and $m_{\text{vis}}(\mu, \tau_{\text{had}})$ for $Z \rightarrow \tau\tau$ and $W+\text{jets}$, shown for two versions of SHERPA.



(b) Distributions in ϕ_η^* , p_T^μ , $\sin \theta_\eta^*$, and the tau BDT score.

Figure A.8: Comparison of the distributions in important kinematic variables between the two different versions of the SHERPA generator used. The level agreement in all distributions is very high.

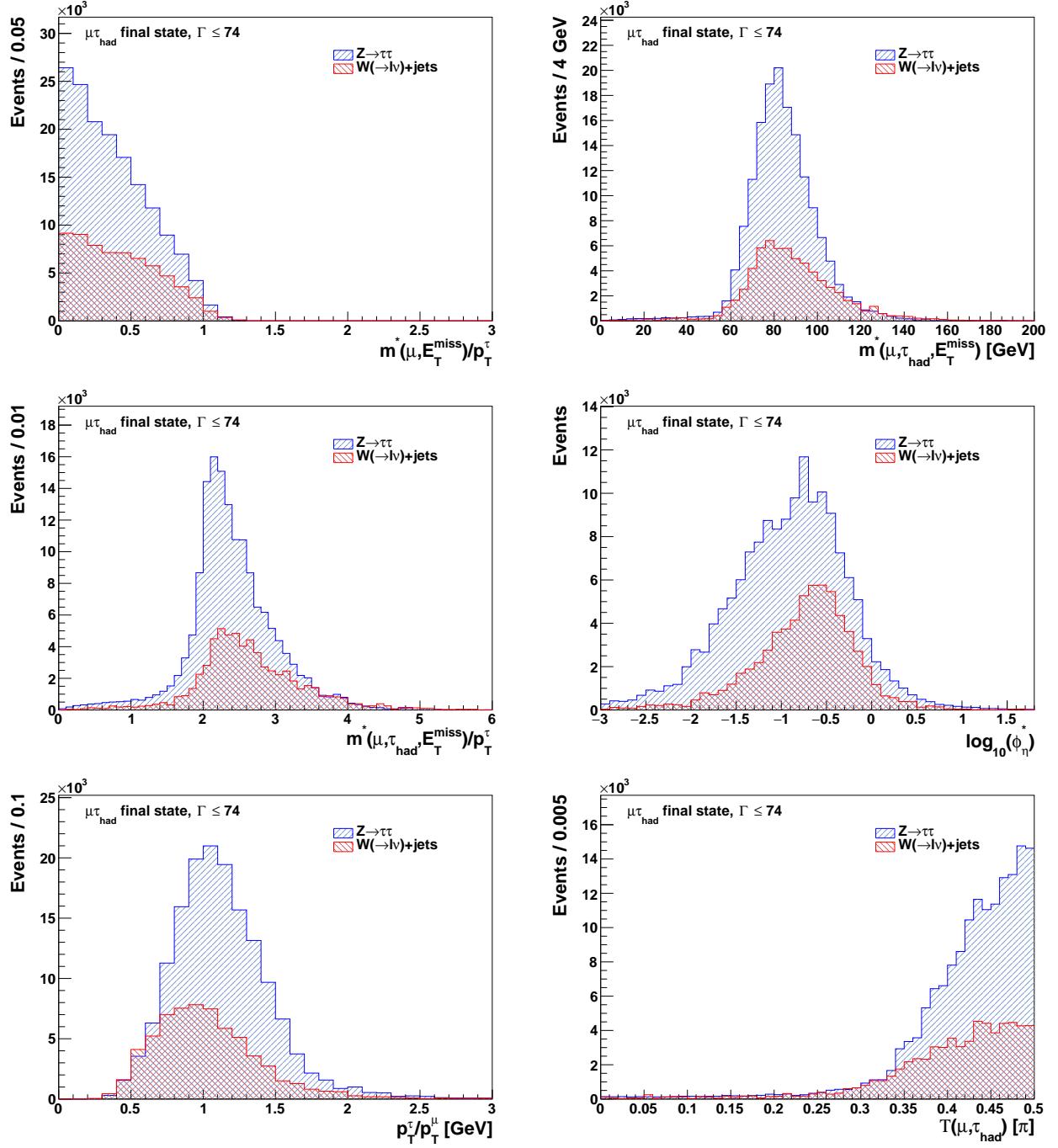


Figure A.9: Distributions in additional candidate variables at the operating point of $\Gamma \leq 74$. The *twist* (bottom right) is defined as $T(\mu, \tau_{\text{had}}) \equiv \arctan [\Delta\phi(\mu, \tau_{\text{had}})/\Delta\eta(\mu, \tau_{\text{had}})]$ [29]. It is clear that none of the distributions offer significant added discriminating power by applying fixed cuts.

A SUPPLEMENTARY FIGURES

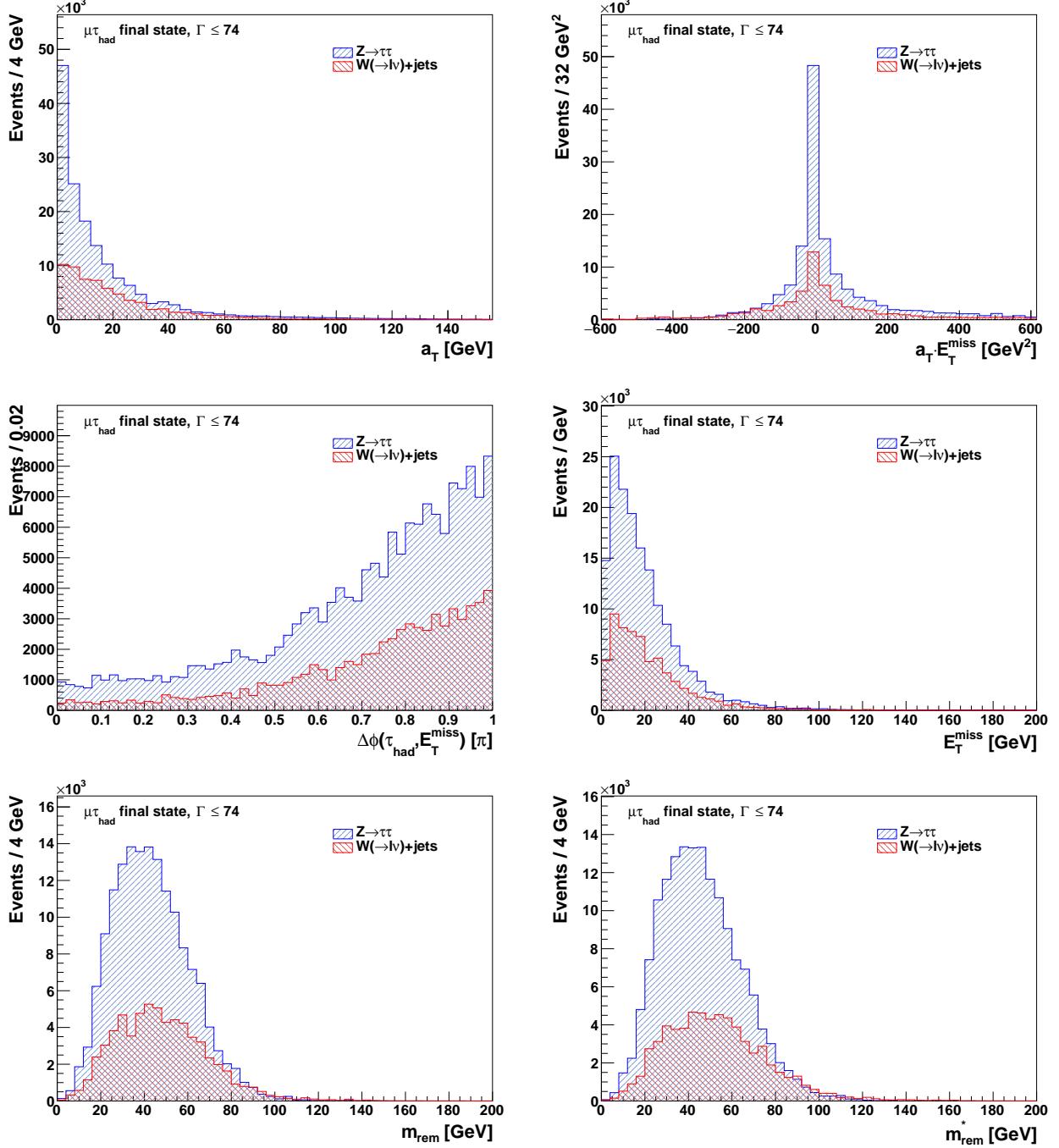


Figure A.10: Further distributions at the operating point $\Gamma \leq 74$. m_{rem} (bottom left) is defined as $m_{\text{rem}}^2 = 2 E_T^{\text{miss}} p_T^\mu (2 - \text{SumCos})$. The top-right plot shows the dot product of the vector versions of a_T and E_T^{miss} , i.e. $\vec{a}_T \cdot \vec{E}_T^{\text{miss}}$.

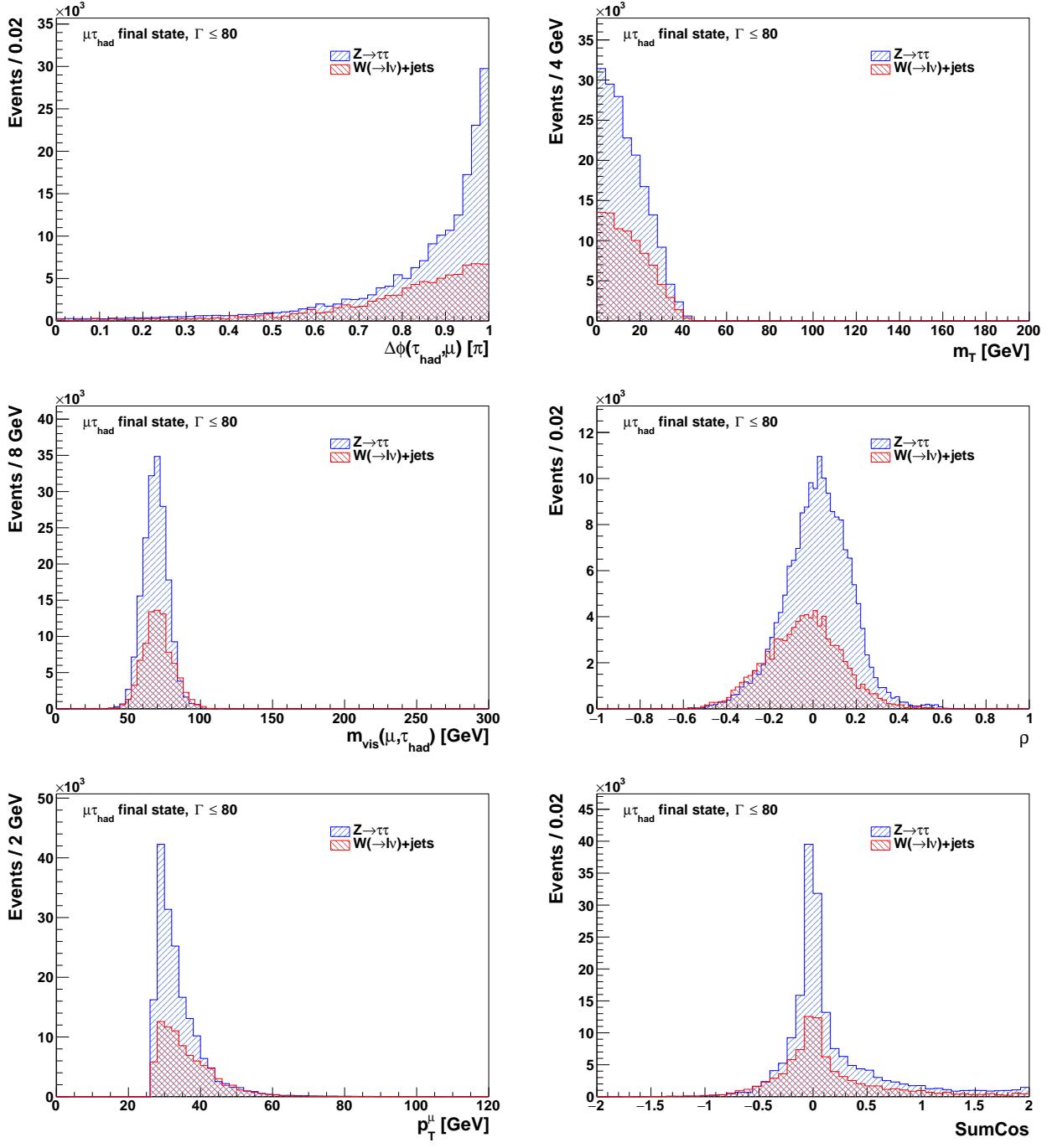


Figure A.11: Distributions in additional candidate variables at an alternative operating point of $\Gamma \leq 80$. Even with a looser Γ cut, none of the distributions become substantially more discriminating compared to $\Gamma \leq 74$ (cf. Figure 20).

A SUPPLEMENTARY FIGURES

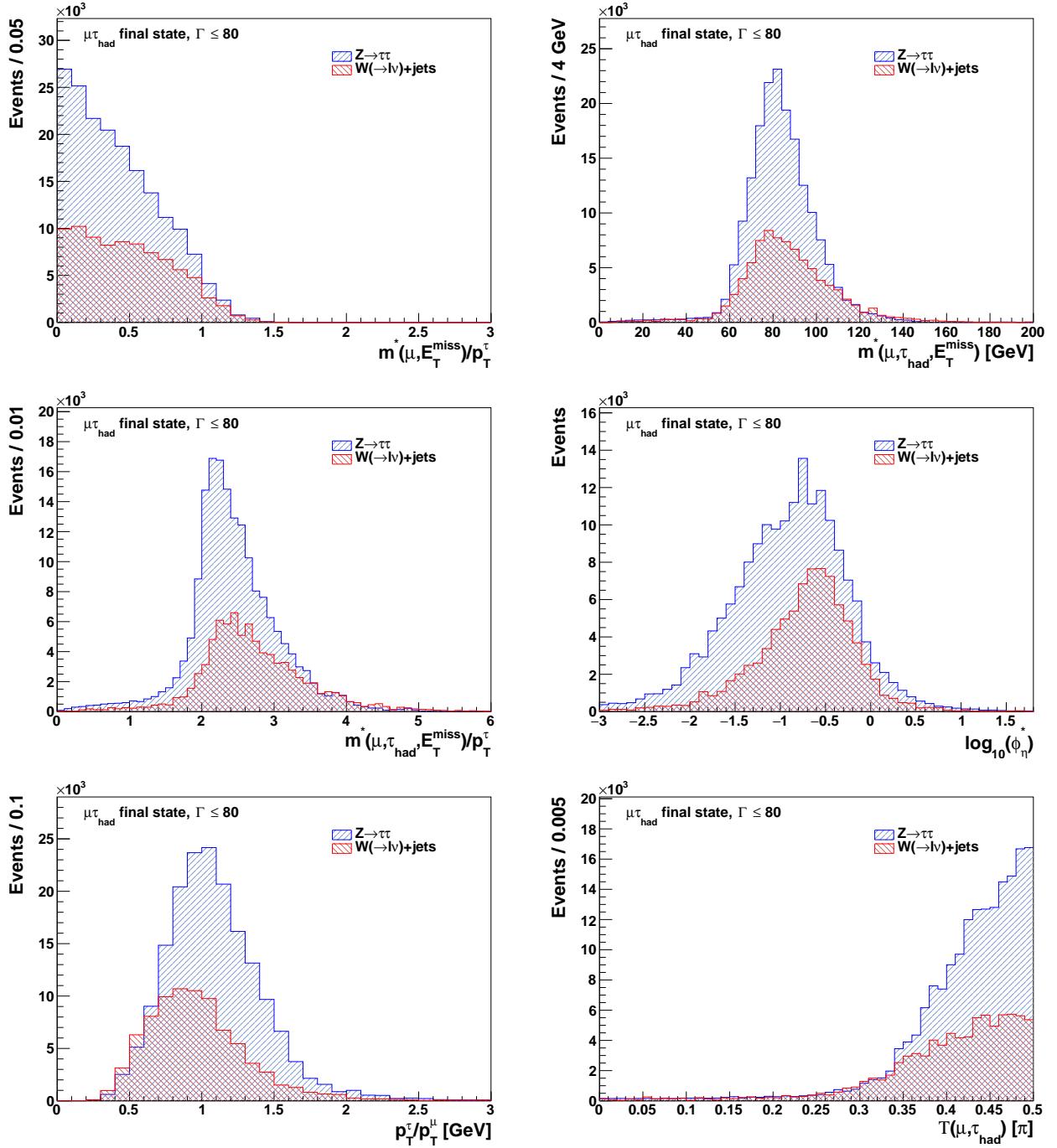


Figure A.12: Distributions in additional candidate variables at the operating point of $\Gamma \leq 80$.

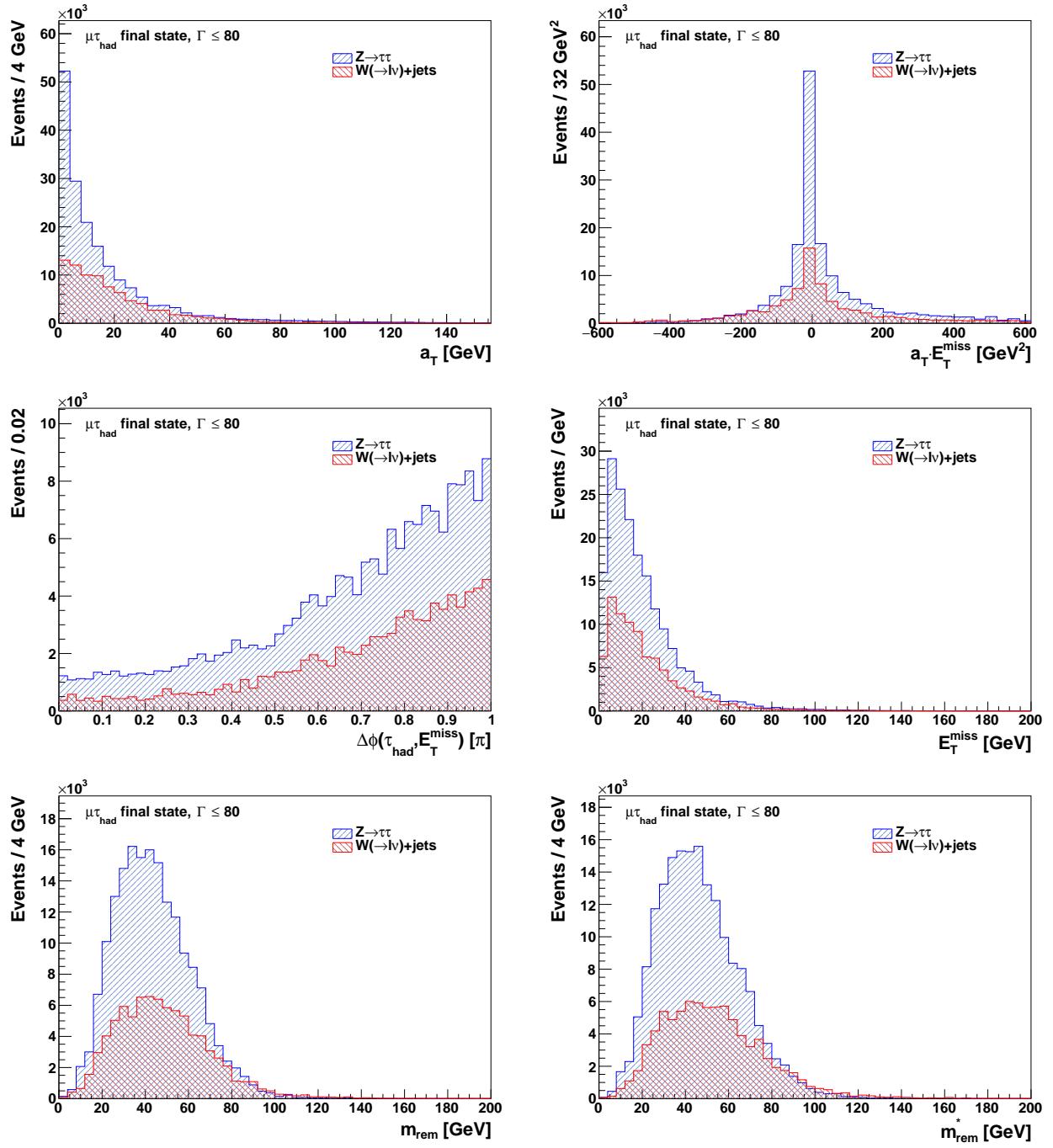


Figure A.13: Remaining distributions at the alternative operating point of $\Gamma \leq 80$.

Appendix B: Monte Carlo samples used

process	generator	sample DSIDs	order
$Z \rightarrow \tau\tau$	SHERPA	361420-361443	0,1,2j@NLO, 3,4j@LO
$Z \rightarrow \tau\tau$	SHERPA	36336(1-3), 3631(02-20)	NNLO
$W \rightarrow e\nu$	SHERPA	361300-361323	0,1,2j@NLO, 3,4j@LO
$W \rightarrow e\nu$	SHERPA	363460-363483	NNLO
$W \rightarrow \mu\nu$	SHERPA	361324-361347	0,1,2j@NLO, 3,4j@LO
$W \rightarrow \mu\nu$	SHERPA	363436-363459	NNLO
$W \rightarrow \tau\nu$	SHERPA	363361-363363	0,1,2j@NLO, 3,4j@LO
$W \rightarrow \tau\nu$	SHERPA	363331-363354	NNLO

Table A1: Monte Carlo samples used in this study. A comprehensive overview is available at [30]. The order is given for different numbers of jets (j), unless omitted, when it applies to all numbers of jets simulated. The dataset ID (DSID) uniquely identifies a simulated sample.

Appendix C: Muon triggers

muon triggers
HLT_mu20_iloose_L1MU15
HLT_mu24_imedium
HLT_mu24_ivarmedium
HLT_mu26_ivarmedium
HLT_mu40
HLT_mu50

Table A2: Summary of muon triggers applied. The triggers listed are combined with a logical OR.

Appendix D: Preselection cutflow

cut	$Z \rightarrow \tau\tau$	$W + \text{jets}$
input events	$1.28851 \cdot 10^7 \pm 18462$	$2.09905 \cdot 10^8 \pm 72528$
$1 \mu \&& 0 e$	$1.42454 \cdot 10^6 \pm 6003$	$5.90269 \cdot 10^7 \pm 38168$
$> 0 \tau$	556387 ± 3760	$6.89036 \cdot 10^6 \pm 12750$
$q_\tau \cdot q_\mu == -1$	514841 ± 3625	$4.61943 \cdot 10^6 \pm 10443$
muon trigger	408837 ± 3223	$3.73245 \cdot 10^6 \pm 9389$

Table A3: Summary table showing the cutflow of events for the preselection cuts.