
Measurement of the $H \rightarrow \tau^+\tau^-$ Production Cross-Section in Hadronic Final States in 13 TeV Proton-Proton Collisions with the ATLAS Detector

---

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsprogramm ProPhys
der Georg-August University School of Science (GAUSS)

vorgelegt von

Eric Drechsler

aus Greiz

Göttingen, 2018

Betreuungsausschuss

Prof. Dr. Arnulf Quadt
Dr.-habil. Ulrike Blumenschein
Dr. Zinonas Zinonos

Mitglieder der Prüfungskommission:

Referent:	Prof. Dr. Arnulf Quadt Institut für Kern- und Teilchenphysik, Georg-August-Universität Göttingen
Koreferent:	Prof. Dr. Stan Lai Institut für Kern- und Teilchenphysik, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Ariane Frey
Institut für Kern- und Teilchenphysik, Georg-August-Universität Göttingen

Prof. Dr. Karl-Henning Rehren
Institut für Theoretische Physik, Georg-August-Universität Göttingen

Prof. Dr. Steffen Schumann
Institut für Theoretische Physik, Georg-August-Universität Göttingen

Prof. Dr. Wolfram Kollatschny
Institut für Astrophysik, Georg-August-Universität Göttingen

II. Physik-UniGö-Diss-2018/02
Tag der mündlichen Prüfung: 14.03.2018

There is in all things a pattern that is part of our universe. It has symmetry, elegance and grace - those qualities you find always in that which the true artist captures.

- Frank Herbert, Dune.

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Measurement of the $H \to \tau^+\tau^-$ Production Cross-Section in Hadronic Final States in 13 TeV Proton-Proton Collisions with the ATLAS Detector

Abstract

The couplings of the Standard Model Higgs boson to fermions provide valuable insights into the nature of electroweak symmetry breaking in the Yukawa sector.

This thesis summarises the measurement of the Higgs boson production cross section in decays to two hadronically decaying τ-leptons. The measurement is based on $36.1\,\mathrm{fb}^{-1}$ proton-proton collision data at a centre-of-mass energy of 13 TeV recorded in 2015 and 2016 with the ATLAS detector at the Large Hadron Collider.

A cut-based and a multivariate analysis approach are investigated, based on the reconstructed $\tau\tau$-mass and boosted decision trees, respectively. The two dominant Higgs boson production modes, the gluon-gluon fusion and vector-boson fusion, are tageted for a mass hypothesis of $m_H = 125\,\mathrm{GeV}$. Their characteristic event topologies are exploited to separate Higgs decays from background processes.
The main background sources arise from $Z/\gamma^* \to \tau\tau$ and multijet events. The background estimation relies on both simulation and data-driven methods.

An excess of events over the predicted background was found, corresponding to an observed (expected) significance of 4.15 (3.56) standard deviations for the multivariate analysis. The signal strength for the multivariate analysis was measured as $\mu = 1.15 \pm 0.39$ and is compatible within total uncertainties with the cut-based result of $\mu = 0.73 \pm 0.34$.

The measured total cross section times branching ratio for the $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ process is $\sigma^{\mathrm{CBA}} = 1.06 \pm 0.49\,\mathrm{pb}$ for the cut-based analysis and $\sigma^{\mathrm{MVA}} = 1.67 \pm 0.57\,\mathrm{pb}$ for the multivariate approach, quoted with the respective total uncertainty.

This result constitutes a direct evidence of the Higgs boson couplings to fermions and confirms the Standard Model prediction for a Higgs boson mass of $m_H = 125\,\mathrm{GeV}$.

Contents

CHAPTER 1

Introduction

Since the early days of the great Greek philosophers, the search for answers to the fundamental questions of our existence incessantly drove mankind to reach unprecedented objectives. How did our universe emerge? What is our world made of? What are the principles behind the laws of nature?
A variety of different disciplines tackle these basic questions, ranging from philosophical and scientific reasoning to religious and spiritual enlightenment.
Modern elementary particle physics constitutes a scientific pinnacle of these disciplines. By investigating the properties and interactions of the fundamental constituents of matter, particle physics provides insights on the emergence, structure and evolution of our universe.

Formulated in the 1970s [1], the Standard Model of Particle Physics (SM) provides a quantum field theoretical description of our current knowledge of elementary particles and their interactions. The latter are mediated by three fundamental force particles called bosons, which give rise to the electromagnetic, weak and strong interactions between the constituents of matter, the fermions. Gravitation, the fourth fundamental force, is relevant only on energy-scales much higher than currently experimentally accessible [2].
Various experiments demonstrated the precise predictive power and consistency of the SM over the last decades. Experimental facilities like the Large Electron-Positron Collider LEP, TEVATRON and HERA successfully measured elementary particle phenomena, supporting many theoretical predictions [3–5].

In the early years of the 21st century, the last experimentally unconfirmed substantial ingredient of the current SM formulation was the Higgs boson. This boson arises as a consequence of a mass-generating mechanism, an essential part of the SM to incorporate particle masses into the theory. The couplings of this boson to the elementary particles is proportional to their mass and can be predicted by the SM for a given Higgs boson mass m_H. Searches for this boson and other new phenomena motivated the construction of the Large Hadron Collider (LHC) at CERN. This circular accelerator is designed to collide proton bunches at an unparalleled energy frontier of 14 TeV at four separate interaction points. Located at each of these are state-of-the-art particle detectors, namely ATLAS, CMS, ALICE and LHCb which measure the remnants of the proton-proton collisions [6].

The first research run of the LHC, labelled RUN I, took place between 2010-2013 at

centre-of-mass energies up to $\sqrt{s} = 8\,\mathrm{TeV}$.
In the summer of 2012, the ATLAS and the CMS collaborations announced the discovery of a new boson with properties consistent with the SM prediction for a Higgs boson with $m_H = 125\,\mathrm{GeV}$ [7, 8]. This marked a great success for the LHC research programme and the beginning of a new era in modern particle physics.
The discovery was mainly driven by searches in bosonic decay modes $H \to \gamma\gamma$, $H \to ZZ$ and $H \to W^{\pm}W^{\mp}$ which offer distinct, well-detectable event signatures. Since then, more precise measurements in bosonic decays confirmed the SM-like properties of the discovered boson. On top of the bosonic couplings, a measurement of the fermionic couplings is of utmost importance to validate the mass-generation mechanism in the SM. As these are introduced in an ad-hoc manner by adding Yukawa couplings between Higgs boson and fermions to the theory, experimental confirmation is essential to understand the exact nature of the Higgs boson. While the discovery of the Higgs boson in loop-induced $\gamma\gamma$ final states provides indirect evidence of the fermionic coupling to quarks, a direct measurement can be achieved by probing Higgs boson decays into two fermions. The $H \to \tau^+\tau^-$ process is the most sensitive channel among the fermionic decay modes at the LHC for an observation of Yukawa couplings and for the measurement of the Higgs boson production rate in vector-boson fusion. Three different final states are formed in $\tau\tau$ decays, depending on the respective leptonic or hadronic τ decay mode. The channel also allows for a determination of the quantum numbers of the Higgs boson in fermionic decays.
In RUN I, a search for the $H \to \tau^+\tau^-$ decay was performed with the ATLAS experiment [9], using $20.3 + 4.5\,\mathrm{fb}^{-1}$ collision data at $\sqrt{s} = 8 + 7\,\mathrm{TeV}$, respectively. Three analyses, each targeting a different final state, were combined. An excess with an observed (expected) significance of 4.5 (3.4) standard deviations (σ) was found over the background prediction, constituting an evidence for the Yukawa coupling.
In 2016 a measurement of the Higgs boson production and decay rates based on combined ATLAS and CMS data recorded in 2011 and 2012 was published [10]. The results, which include constraints on the Higgs boson couplings to vector bosons and fermions are based on approximately $20 + 5\,\mathrm{fb}^{-1}$ proton-proton collision data at $\sqrt{s} = 8 + 7\,\mathrm{TeV}$, respectively, per experiment and are consistent with the SM predictions. The vector boson fusion production process was observed with significance of $5.4\,\sigma$, while a $4.6\,\sigma$ deviation was expected for a SM Higgs boson with $m_H = 125\,\mathrm{GeV}$.
In this combined measurement, the direct observation of the $H \to \tau^+\tau^-$ process was measured (expected) with a significance of $5.5\,\sigma$ $(5.0\,\sigma)$.

Between 2013-2015, the LHC was upgraded to facilitate collisions at an increased $\sqrt{s}$ of $13\,\mathrm{TeV}$. During this phase, the ATLAS detector underwent significant modifications to cope with higher event rates and provide improved reconstruction efficiencies. The preparation of the second LHC research run began in mid-2014, with a commissioning phase successfully completed in April 2015, marking the start of RUN II.
In the years 2015 and 2016, the ATLAS experiment recorded $36.1\,\mathrm{fb}^{-1}$ collision data at $\sqrt{s} = 13\,\mathrm{TeV}$. This thesis is dedicated to the measurement of the Higgs boson coupling to hadronically decaying $\tau\tau$ pairs in this dataset. The analysis was performed in collaboration with the leptonic-hadronic and leptonic-leptonic decay channel analyses. In each channel, two distinct event topologies are targeted, motivated by the dominant Higgs boson

production modes at the Lhc: Vector-boson fusion provides the most distinct signal topology, with a characteristic signature of two jets close to the beampipe and a central τ-pair. The boosted topology is a result of a Higgs boson produced in gluon-gluon fusion with high transverse momentum and a recoiling jet.
Two separate experimental strategies are investigated for the hadronic channel, a cut-based and a multivariate analysis approach. The latter relies on a multivariate classification using the machine learning technique of boosted decision trees and provides the highest sensitivity amongst the two.
At the time of completion of this thesis, the preparation for the combination of the results of all three decay channels is ongoing.

This thesis is structured as follows:

After a brief summary of the theoretical aspects of the Standard Model in Chapter (2), details on Higgs boson specific properties and experimental results are provided in Chapter (3). Thereafter the experimental apparatus is presented together with related phenomenological characteristics of hadron colliders in Chapter (4). Subsequently, more analysis-specific topics are discussed in Chapter (5), starting with an overview of the event and object reconstruction at Atlas relevant for the studies presented here. The analysis strategy is outlined in Chapter (6), summarising the selection and categorisation steps. This is followed by a detailed overview of the event model in Chapter (7), specifically including discussions on the estimation of multijet and $Z/\gamma^{*}\rightarrow\tau\tau$ background contributions. Final discriminants used in the cut-based and multivariate analyses are presented in Chapter (8) together with the profile likelihood function based fit-model used for the measurement. The final results of the studies presented in this thesis are summarised in Chapter (9).

Personal Contributions

The Atlas experiment is a large collaboration with over 3000 members from all around the world. All Atlas measurements depend on the diligent work of the many collaborators, making the experiment one of the biggest scientific projects to date.
The following list highlights the main contributions done for the collaboration in course of this thesis in chronological order.

- studies for the optimisation and support of the Run I H $\rightarrow \tau_\ell \tau_\ell$ analysis, specifically investigating central jet veto requirements and providing analysis inputs [9]
- commissioning and validation of the Pixel detector data quality monitoring (DQM) in preparation of Run II, including on-call expert shift duties as well as general DQM shifts in the Atlas control room during data taking periods
- commissioning of the τ identification strategy in early Run II, in particular the electron rejection method, including a tag-and-probe efficiency measurement in 2015 data [11, 12]
- development and management of the central data reduction and preparation tool for τ-related analysis in Run II, the xTauFramework, in close collaboration with Dr. Zinonas Zinonos
- dedicated technical support for the data preparation of various analysis related to Higgs bosons and τ leptons, most noticeably the search for additional heavy neutral Higgs and gauge bosons in Run II [13]
- significant contributions to the development of the analysis framework used for the charge-parity measurements in $H \rightarrow \tau\tau$ decays
- the complete measurement of the Higgs boson couplings to hadronically decaying τ leptons, denoted as $H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$, subject to this thesis. This included
  - ⋆ the coordination of the analysis data preparation for all three channels
  - ⋆ design and validation of the final $\tau_{\text{had}}\tau_{\text{had}}$ background model
  - ⋆ implementation and optimisation of the cut-based analysis for Run II
  - ⋆ construction of the multivariate analysis, including implementation and training of dedicated boosted decision trees
  - ⋆ preparation and validation of the $\tau_{\text{had}}\tau_{\text{had}}$ standalone fit-model
  - ⋆ providing inputs for and developing the combined fit-model for cut-based and multivariate channel combinations.

CHAPTER 2

Standard Model and Higgs Boson

In the following chapter, a concise introduction to the Standard Model of Particle Physics is provided, outlining the theoretical framework with its remarkable experimental successes briefly.
The Higgs mechanism, a key ingredient of the Standard Model is presented together with its theoretical and experimental implications. Thereafter, recent experimental results in the search for the Higgs boson and measurement of its properties are summarised.

The nomenclature and theoretical discussion follows [1, 14, 15].

2.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) is a comprehensive theoretical model which describes elementary particles and their interactions. In its modern form, it is formulated as a gauge quantum field theory and contains the internal symmetries of a unitary group.

2.1.1 Historical Overview

In the early 1960s, Sheldon Glashow established the basis for the SM in its modern form by discovering a quantum field theory formulation to combine electromagnetic and weak interactions [16]. The lack of explanation for the origin of masses was answered in 1964 by three independent groups[1] with the introduction of spontaneous symmetry breaking. This theory implies massive vector bosons and is later on referred to as Higgs mechanism [17–19], detailed in Sec. 2.2. In the late 1960s, Steven Weinberg and Abdus Salam introduced the Higgs mechanism to the theory of electroweak interaction, expressing the SM in its current formulation [20].

The SM also includes the theory of Quantum Chromodynamics QCD which describes the interaction of quarks and gluons, constituents of hadrons like the proton.

---

[1] Robert Brout and Francois Englert; Peter Higgs; and Gerald Guralnik, Carl Richard Hagen, and Tom Kibble.

As a result of developing a more general form of quantum field theory in 1954, Chen Ning Yang and Robert Mills discovered a formulation which could combine both, electroweak and strong interactions [21].
In 1973, Harald Fritzsch, Heinrich Leutwyler and Murray Gell-Mann developed a formulation of QCD which introduced a new quantum number for particles partaking in the strong interaction, later on referred to as colour. Gluons, the force carrier particles of the strong interactions carry colour charge themselves, allowing for a self-interaction [22].
The concept of asymptotic freedom as proposed by David Gross, Frank Wilczek and Hugh Politzer in 1973 allowed for precise predictions of experimental measurements concerning QCD using perturbation theory [23, 24].

2.1.2 Theoretical Overview

The SM is formulated as a locally Lorentz-gauge invariant and renormalisable effective quantum field theory. The unitary gauge group

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

describes the internal symmetries intrinsic to the SM. These symmetries introduce interactions and properties of the SM particle fields. The indices in Eq. 2.1 depict different characteristics of particles undergoing the transformations related to each group:

- C is short for colour charge, the quantum number and conserved quantity of the strong interaction
- L for left-handedness, denoting the negative chirality of particles partaking in the weak interaction
- Y for the weak hypercharge, related[1] to the particles' electrical charge and third weak isospin component.

The SM can be expressed using the Lagrange formalism, by adapting principles of classical mechanics to relativistic field theory. The Lagrangian $\mathcal{L}$ is formulated as a function of the particle fields and their space and time derivatives. Using a generalised form of the Euler-Lagrange equations, field equations for non-interacting particles can be derived dependent on the particles' spin - in particular the Klein-Gordon equation for spin-0 (scalar), the Dirac equation for spin-$1/2$ and the Proca equation for spin-1 fields.
As a consequence of requiring invariance under local gauge transformations described by the group in Eq. 2.1, the force-mediating fields are introduced, thereby giving rise to the particle's interactions. Following Noether's Theorem each of the internal symmetries arising from the local gauge invariance results in a conserved quantity for the system [25].

In its current form, the SM describes the electromagnetic, the weak and the strong interaction of the particles represented in Fig. 2.1. The energies required to measure effects of gravity between the fundamental particles are not accessible with current collider

---

[1] $Y = 2(Q - I_3)$, where Q denotes the electrical charge and I_3 the third weak isospin component of a particle

experiments and can be neglected [2]. The SM is understood as an effective quantum field theory, an approximation to an underlying physical theory relevant at energies where gravitational effects become significant [26].

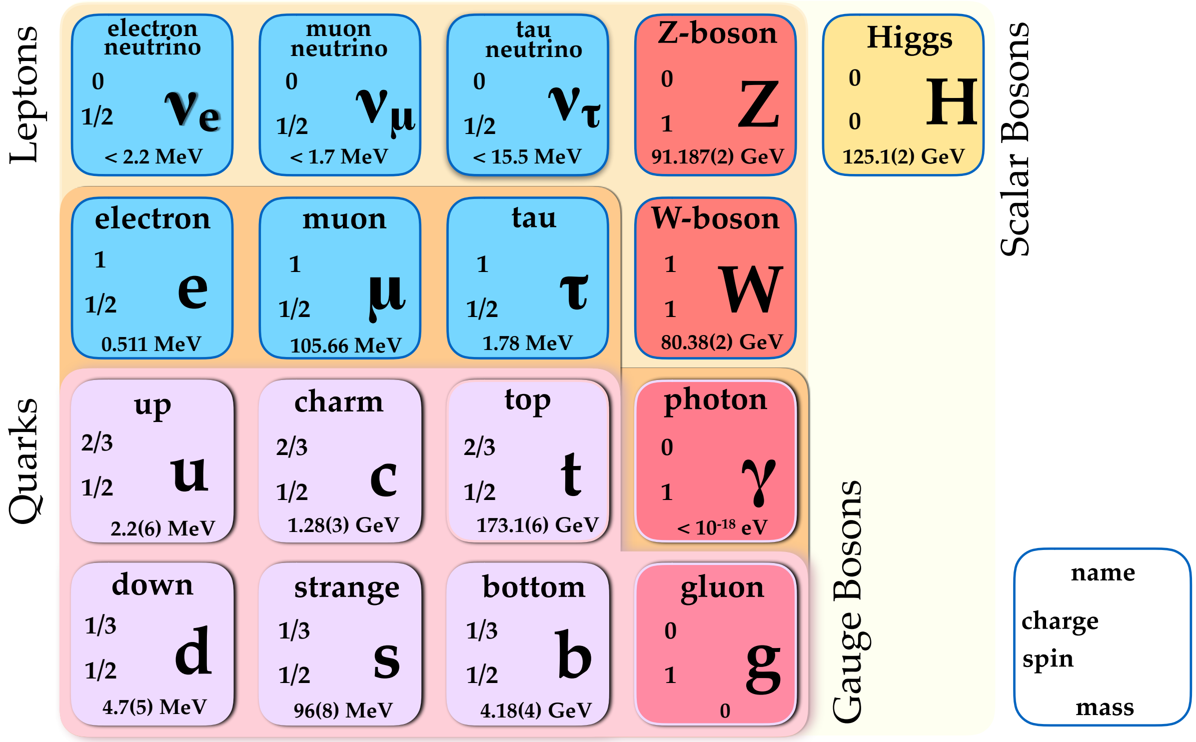


Figure 2.1: A representation of the fundamental particles described by the SM. All particles have been experimentally confirmed and their properties measured to high precision [27]. Particles are grouped in charge, spin and masses. The boxes around the fermions - quarks and leptons - indicate the interaction with the respective force particles.

The representation of SM's fundamental particles in Fig. 2.1 details quantum numbers and properties of each particle, determining their characteristics and behaviour.
Leptons and quarks are grouped as fermions, the building blocks of matter. They are spin-$^1/_2$ particles and arranged in three generations with successively increasing masses. Neutrinos ν_ℓ are electrically neutral while the massive leptons ℓ each carry a charge of 1, where $\ell = e$, μ, τ. Up-type quarks $q_u = u$, c, t have a non-integer charge of $+^2/_3$, while the down-type quarks $q_d = d$, s, b have a charge of $-^1/_3$.
For every fermion, there also exists an antiparticle with the same mass, lifetime and spin but opposite internal quantum numbers, like electrical charge or lepton number.
Vector-bosons are the force mediators in the SM with an integer spin of 1. The electromagnetic force is mediated by the neutral and massless photon γ, the weak force by the massive Z^0, $W^\pm$ bosons and the strong force by eight neutral, massless and colour-charged

gluons g. The special role of the massive, scalar Higgs boson H, the only boson with spin 0 in the SM, is detailed in Chapter (3).

2.1.3 Quantum Chromodynamics

The local gauge of the non-Abelian symmetry group $\mathrm{SU}(3)_C$ in Eq. 2.1 gives rise to the theoretical formulation of the strong force, QCD. Only quarks participate in the strong interaction which is mediated by the gluons.

The generators of the $\mathrm{SU}(3)_C$ transformations are represented by the eight Gell-Mann matrices λ_i. These transformations act on 3 component vectors ψ_q, built by the possible colour states of each quark. The colours c are the quantum numbers of $\mathrm{SU}(3)_C$ and commonly referred to as red, green and blue, $c = (r, g, b)$. As opposed to the electrical charge, every observable particle is colour neutral. Gluons carry colour and anti-colour charge $c\overline{c}'$, whereas the (anti-)quarks carry (anti-)color charge $c(\overline{c})$. Hence, gluons are able to interact with each other, characterising the non-Abelian nature of QCD.

The part of the SM Lagrangian describing QCD can be formulated as

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \psi_q^\dagger \left(i\gamma^\mu (\partial_\mu + 2\pi i\, \alpha_s \lambda_c \mathcal{A}_\mu^c) - m_q \right) \psi_q - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{2.2}$$

where $G^a_{\mu\nu}$ denotes the gluon field strength tensor, $\lambda_c \mathcal{A}^c_\mu$ the gluon fields contracted with the Gell-Mann matrices and α_s the strong coupling constant, which describes the strength of the energy-dependent coupling between quarks and gluons. The coupling constant is expressed as a function of an arbitrary energy scale $\Lambda_{\mathrm{QCD}} > 0$, which in context of the SM is commonly $\Lambda_{\mathrm{QCD}} \approx 200\,\mathrm{MeV}$. The coupling constant at a squared four-momentum $|q^2| \gg \Lambda^2_{\mathrm{QCD}}$ can be expressed as

$$\alpha_s^{-1}(|q^2|) = \frac{1}{12\pi} \left(33 - 2n_f\right) \ln \left(\frac{|q^2|}{\Lambda^2_{\mathrm{QCD}}} \right) \tag{2.3}$$

where n_f denotes the number of quark flavours available at the momentum scale q^2 with $n_f = 6$ for $q^2 > m^2_{\mathrm{top}}$.

Considering Eq. 2.3, it can be seen that the strength of the coupling decreases with increasing momentum. As large $|q^2|$ correspond to probing small distances, two characteristic effects of the strong interaction can be described as follows:

- For short distances, quarks act like free particles as the strong coupling constant is small. This effect is called asymptotic freedom. Theoretical predictions of QCD processes are possible when using perturbation theory techniques [28].
- For larger distances, quarks experience an increasing confining potential. Referred to as confinement, this is the reason why quarks only occur in bound states. Perturbation theory fails in this regime due to the increasing size of α_s. Instead, non-perturbative approaches like lattice QCD are used for theoretical predictions [29].

As the distance between two quarks gets larger, the confining potential increases up to a point where it is energetically favourable to combine the original quarks with spontaneously emerging quarks from vacuum. This process is called hadronisation and the bound states are hadrons. Different types of hadrons are

- Mesons with two bound quarks, e.g. pions π^0 and kaons K_s^0
- Baryons, such as the proton p, with three bound quarks
- Tetra- and pentaquarks, recently discovered states at the LHCb experiment, with four (five) bound quarks respectively [30, 31].

In Sec. 4.1, further theoretical and experimental implications of asymptotic freedom and confinement are outlined.

2.1.4 Theory of Electroweak Interaction

The theory of electroweak interaction unifies the electromagnetic and the weak forces and is described by the combination of symmetry groups $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ [16, 20, 32, 33].

The three Pauli matrices σ_i generate the non-Abelian unitary group $\mathrm{SU}(2)_L$. The requirement of local gauge invariance results in the emergence of three gauge fields W_μ^i with $i = 1, 2, 3$ and implies the conservation of the third weak isospin component I_3 according to Noether's theorem.

The fermions in Fig. 2.1 are comprised of components with negative and positive chirality, referred to as left- and right-handed particles, respectively. These behave fundamentally different with respect to the weak interaction: left-handed particles form doublets with weak isospin component $I_3 = \pm 1/2$, while right-handed particles constitute singlets with $I_3 = 0$. While left-handed doublets partake in the weak interaction, right-handed singlets remain invariant under transformations of $\mathrm{SU}(2)_L$. This is a characteristic of the parity-violating nature of the weak interaction [34].

In the SM, the electrically neutral neutrinos ν_ℓ form such doublets (ν_ℓ, ℓ) together with the charged leptons ℓ of the same generation. For quarks, these doublets are formed by the up-type quarks q_u and their respective down-type counterpart q_d of the same generation.

The conserved quantum number of the Abelian symmetry group $\mathrm{U}(1)_Y$, third group of Eq. 2.1, is the weak hypercharge Y. Local gauge invariance of this group implies the existence of a gauge field B^μ.

With $\psi_{L/R}$ denoting the left- or right-handed particle fields, the electroweak part of the SM Lagrangian can be written as

$$\begin{aligned} \mathcal{L}_{\mathrm{EW}} = {} & \psi_L^\dagger \gamma^\mu \left(i\,\partial_\mu - \tfrac{g}{2}\sigma_j W_\mu^j - \tfrac{g'}{2} Y B_\mu \right) \psi_L \\ & + \psi_R^\dagger \gamma^\mu \left(i\,\partial_\mu - \tfrac{g'}{2} Y B_\mu \right) \psi_R \\ & - \frac{1}{4} \left(W_{\mu\nu}^j W_j^{\mu\nu} + B_{\mu\nu} B^{\mu\nu} \right) \end{aligned} \tag{2.4}$$

with electroweak coupling constants g and g'. The terms in Eq. 2.4 describe the interaction of left-handed particles, of right-handed particles and the self-interaction of the gauge fields W_i^μ, B^μ with $i = 1, 2, 3$. The force mediators of the electroweak interaction are formed by a mixture of the gauge fields W_i^μ, B^μ as described in Sec. 2.2.

The weak interaction in the quark sector reveals another important aspect of the electroweak theory. The flavour and weak eigenstates of quarks, q and q' respectively, are not identical but related by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

with non-vanishing, experimentally measured, off-diagonal elements [27, 34, 35]. The weak eigenstates of quarks are hence a mixture of flavour eigenstates across different generations. This means that the weak interaction incorporates flavour-changing charged currents across different generations, mediated by $W^\pm$ bosons. So far, no flavour-changing neutral currents, which are suppressed in the SM by the GIM-mechanism[1], have been observed [36, 37]. The off-diagonal terms of the CKM matrix are complex numbers, carrying a phase factor which is the source of $\mathcal{CP}$-violation in the SM [34].
Analogously, the weak eigenstates of leptons are rotated by the PNMS-matrix[2], allowing for the oscillation of neutrino flavours [38]. As neutrinos are not measured directly by collider experiments, the effect is neglected in this thesis.

2.2 Electroweak Symmetry Breaking and the Higgs Mechanism

The aforementioned Proca-equation, describing a free spin-1 boson X, introduces a Lagrangian mass-term $m_X^2 X_\mu X^\mu$ with boson mass m_X. Adding such a term to the SM Lagrangian would spoil local gauge symmetry; the gauge bosons W_i^μ, B^μ in Eq. 2.4 must therefore be massless.

However, measurements from the UA1 and UA2 experiments in 1983 at CERN [39–41] revealed that the weak force mediators, the Z^0 and $W^\pm$ bosons are massive [27] with

$$\begin{aligned} m_W &= (80.385 \pm 0.015)\,\text{GeV} \\ m_Z &= (91.1876 \pm 0.0021)\,\text{GeV}. \end{aligned}$$

Hence the symmetry of the physical system described by the SM Lagrangian must be broken to introduce mass. To ensure that the theory is renormalisable, the Lagrangian itself is required to be symmetric under the gauge transformations in Eq. 2.1 [42]. By introducing a degenerate set of ground states to the system from which one is arbitrarily selected as the physical ground state realised in nature. The symmetry of the system is said to be spontaneously broken. In quantum field theory, a ground state is a field

---

[1] Named after S. Glashow, J. Iliopoulos and L.Maiani

[2] Named after B. Pontecorvo, Z. Maki, M. Nakagawa and S. Sakata

configuration with minimal energy - the vacuum. If the vacuum state is non-symmetric, then the system's symmetry is broken, while the Lagrangian itself may remain symmetric.

According to the Goldstone-theorem, every theory with an invariant Lagrangian under a set of transformations and a vacuum which is only symmetric for a subset of these introduces massless scalar particles, the Nambu-Goldstone bosons[1] [43, 44].
The Higgs mechanism combines the principle of local gauge invariance and spontaneous symmetry-breaking (SSB). By introducing a scalar field ϕ with a non-trivial, non-symmetric ground state, SSB occurs. The massless Nambu-Goldstone bosons that arise are eliminated from the theory by requiring a specific gauge - the unitary gauge - which only leaves the physical, observed particles in the Lagrangian. The additional degrees of freedom from the Nambu-Goldstone bosons are absorbed by the gauge bosons which become massive and acquire an additional longitudinal polarisation, a new degree of freedom.

2.2.1 Higgs Mechanism and Gauge Boson Masses

The complex, scalar Higgs field ϕ is a SU(2) doublet with a potential $V(|\phi|)$ with two parameters μ^2 and λ:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad \text{and} \quad V(|\phi|) = \mu^2 \left(\phi^\dagger \phi\right) + \lambda \left(\phi^\dagger \phi\right)^2 \tag{2.5}$$

To ensure a lower boundary for $V(|\phi|)$ λ must be positive. For $\mu^2 > 0$, the minimum of the potential is trivial at $\phi_0 = 0$, describing a scalar field with mass μ and quartic self-coupling constant λ [45]. If however $\mu^2 < 0$, the potential takes the form sketched in Fig. 2.2 with a non-zero global minimum

$$|\phi_0| = \frac{v}{\sqrt{2}} \ \text{ with } \ v = \sqrt{\frac{-\mu^2}{\lambda}} > 0 \ \text{ and } \ V(\phi_0) = -\frac{\lambda}{4} v^4 \quad . \tag{2.6}$$

Hence, the ground state of a system with such a scalar field is non-trivial, with a non-zero vacuum expectation value (VEV) v. As sketched in Fig. 2.2, this introduces SSB; the state at 0 is meta-stable, while the global minimum is a set of degenerated states $\phi_0 e^{i\alpha}$ with VEV v and arbitrary phase α. A particular choice for α breaks the symmetry of the system, but not the symmetry of the Lagrangian. Using its relation to the Fermi-coupling constant and experimental measurements from muon decays, the VEV is estimated to be $v \approx 246\,\text{GeV}$ [27].

The introduction of ϕ implies a new term in the SM Lagrangian

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} = & \left(\partial_\mu \phi + i\tfrac{g}{2}\sigma_j W_\mu^j \phi + i\tfrac{g'}{2} Y B_\mu \phi\right)^\dagger \left(\partial^\mu \phi + i\tfrac{g}{2}\sigma^j W_j^\mu \phi + i\tfrac{g'}{2} Y B^\mu \phi\right) \\ & - V(|\phi|) \qquad . \end{aligned} \tag{2.7}$$

including the kinetic terms, the potential and the interaction of ϕ with the weak gauge

---

[1] Named after Y. Nambu and J. Goldstone

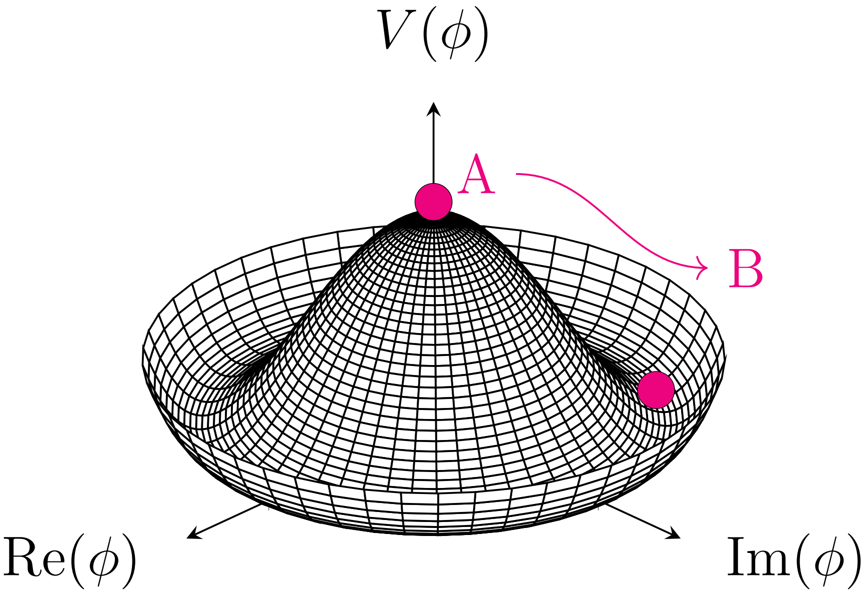


Figure 2.2: Potential $V(\phi)$ of the Higgs field ϕ. Instead of the trivial, symmetric ground state at A, the system's global minimum lies in the degenerate ground state B. This represents the spontaneous symmetry breaking of the system.

bosons W_i^μ, B^μ.
The Higgs field ϕ can be expressed as fluctuations around the minimum v

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \tag{2.8}$$

where $H = H(x)$ denotes the physical Higgs boson field. When expanding Eq. 2.7 in terms of Eq. 2.8, the physical mass eigenstates of the weak gauge bosons can be expressed as mixture of the weak eigenstates

$$\begin{aligned} A_\mu &= W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W && \text{(photon } \gamma) \\ Z_\mu &= W_\mu^3 \sin\theta_W - B_\mu \cos\theta_W && (Z\text{-boson } Z^0) \\ W_\mu^\pm &= \tfrac{1}{\sqrt{2}} \left(W_\mu^1 \mp i\, W_\mu^2 \right) && (W\text{-bosons } W^\pm) \end{aligned} \tag{2.9}$$

where θ_W is the weak mixing angle, called Weinberg-angle. This angle relates the electroweak coupling constants via

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \quad . \tag{2.10}$$

and has been experimentally measured to be $\sin^2\theta_W = 0.2223 \pm 0.0021$ [46].

The resulting expression for the part of the SM Lagrangian introducing the Higgs field

ϕ is then given as

$$\mathcal{L}_{\text{Higgs}} = \underbrace{\frac{1}{4}g^2v^2W_\mu^+W^{-,\mu} + \frac{1}{8}(g^2+g'^2)v^2Z_\mu Z^\mu}_{\text{gauge boson mass terms}} \\ \underbrace{+\mathcal{L}^H_{\text{kin}} + \mathcal{L}^H_{\text{mass}} + \mathcal{L}^H_{\text{Couplings}}}_{\text{Higgs physics terms}} \quad . \tag{2.11}$$

The Higgs mechanism hence generated mass terms for the weak bosons $W^\pm$ and Z^0, as well as terms describing the kinetic energy, mass and couplings of the Higgs boson H. The latter are further detailed in Sec. 3.1. The gauge boson masses are given as

$$\begin{aligned} \text{Photon } \gamma : m_A &= 0 = m_\gamma \\ Z^0 \text{ boson} : m_Z &= \frac{v}{2}\sqrt{g^2+g'^2} \\ W^\pm \text{ bosons} : m_W &= \frac{v}{2}g = m_Z\cos\theta_W \quad . \end{aligned}$$

resulting in a massless photon and massive weak gauge bosons. The masses of these depend on the weak coupling constants g, g' and the VEV v.

2.2.2 Masses in the Fermion Sector

The Dirac-equation of a free fermion field ψ implies a mass term

$$-m_f\bar{\psi}\psi = -m_f(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \ .$$

Adding such a term directly to the SM Lagrangian would spoil local gauge symmetry, as left- and right-handed components transform differently under $SU(2)$.
With the Higgs field ϕ in Eq. 2.5, a new coupling between fermions and ϕ can be introduced. The corresponding Lagrangian term for this Yukawa coupling[1] is

$$\mathcal{L}_{\text{Yuk}} = -\sum_f \lambda_f\left(\bar{\psi}^f_L\phi\psi^f_R + \bar{\psi}^f_R\phi\psi^f_L\right) \tag{2.12}$$

$$= -\sum_f\left(\underbrace{\frac{\lambda_f v}{\sqrt{2}}\bar{\psi}^f\psi^f}_{\text{Mass term}} + \underbrace{\frac{\lambda_f}{\sqrt{2}}H\bar{\psi}^f\psi^f}_{\text{Coupling term}}\right) \tag{2.13}$$

where the parameter λ_f describes the coupling strengths of the fermions f to the Higgs field. Expanding ϕ around its minimum v and using the unitary gauge, Eq. 2.12 transforms to Eq. 2.13, where explicit fermion mass terms and coupling terms between the Higgs boson H and the fermions ψ^f can be identified. The mass of a fermion and its couplings

---

[1] Named after H. Yukawa

to the Higgs boson are hence given as

$$\text{Fermion mass} : m_f = \frac{\lambda_f v}{\sqrt{2}} \tag{2.14}$$

$$\text{Yukawa coupling} : \frac{\lambda_f}{\sqrt{2}} = \frac{m_f}{v} \quad . \tag{2.15}$$

The coupling between Higgs boson and fermions in Eq. 2.15 is proportional to the fermion's mass. Comparing with the masses listed in Fig. 2.1, the top-quark has the largest Yukawa-couplings for all fermions, while the τ lepton has the largest coupling in the lepton-sector. The coupling parameter λ_f is not predicted but a free parameter in the SM.

In conclusion, the Higgs mechanism generated particle masses by introducing a scalar field with a non-trivial ground state. The SM Lagrangian remains symmetric under $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ transformations, but is spontaneously broken by the non-symmetric ground state. This results in masses for the weak gauge bosons $W^{\pm}, Z^0$ and fermions. The $\mathrm{SU}(3)_C$ colour and electromagnetic $\mathrm{U}(1)_Q$ symmetries remain unbroken and the SM Lagrangian invariant under the symmetry group in Eq. 2.1.

2.3 Open Questions in the Standard Model

The SM is a successful theoretical framework and predicts various physics quantities over several orders of magnitude at high precision. This is demonstrated in Fig. 2.3, summarising measured production cross sections and theoretical predictions at ATLAS for a variety of physics processes.

Despite these remarkable successes, there are several shortcomings and open questions which indicate the need for modifications or extensions of the SM.

Two important examples of experimental results which differ from SM predictions are:

- Measurements of the cosmic microwave background [47] and velocity of galaxy rotations [48] indicate that only ~ 5 % of the energy content observed in the universe is described by the particles of the SM, the remaining 95 % are composed of 27 % dark matter and 68 % dark energy.
- The anomalous magnetic moment of muons, which is sensitive to loop-corrections, shows a deviation of 3.4 standard deviations from the SM expectation [27].

Supersymmetric extensions of the SM give rise to additional particles, which can account for these observations among other candidate theories [49].

Furthermore, there are intrinsic, fundamental questions not addressed in the current formulation of the SM, examples being:

- Which fundamental mechanism causes quarks to carry a third of the electrons' charge?
- Why are there exactly three generations of matter?
- The group structure in Eq. 2.1 is expected to be unified at high energies, an effect

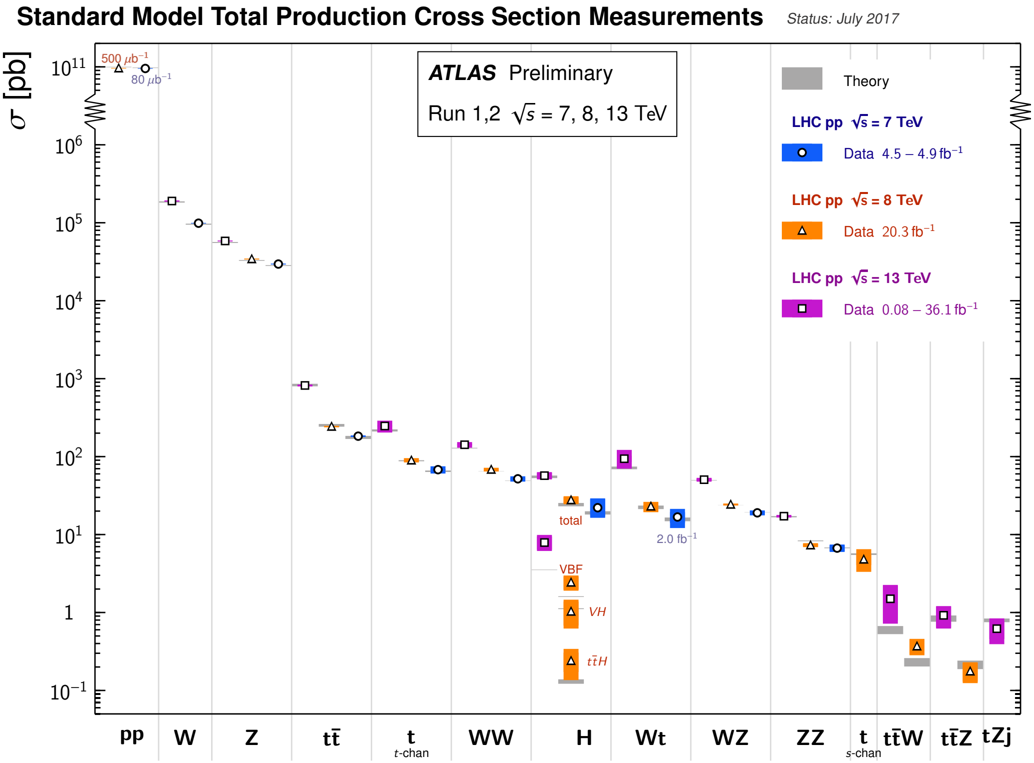


Figure 2.3: Summary of SM total production cross section measurements at Atlas, compared to theoretical predictions calculated at next-to-leading order or higher.

not described within the SM [49].

- Which mechanism caused the imbalance between matter and antimatter we observe today in our universe? The effect of $\mathcal{CP}$-violation as incorporated in the SM is too small to explain this asymmetry [50].

Specifically related to the Higgs boson sector of the SM is the so-called fine-tuning or hierarchy problem. Corrections to the Higgs boson mass from various loop diagrams result in a value several orders of magnitudes larger than 125 GeV. In the current formulation of the SM, this effect can only be cancelled by tuning μ^2 to a very precise value, which seems unnatural. Again, supersymmetric extensions to the SM provide an intrinsic solution through automatic cancellation of loop-diagrams [49].

One of the major goals of the Atlas-physics programme is the observation of physics beyond the SM, in particular supersymmetric processes. However, to date no experimental evidence for such processes has been found [51].

CHAPTER 3

Higgs Boson Phenomenology and Recent Experimental Results

The Higgs mechanism introduced in the previous chapter implies the existence of the Higgs boson, which gives an experimentally detectable signature. Various indirect, theoretical constraints on the Higgs boson mass m_H motivated large-scale experimental search campaigns.
As the Higgs field is a scalar particle field, the Higgs boson carries a spin of 0. The full expression for the Lagrangian in Eq. 2.11 for all terms containing the physical Higgs field H and gauge bosons $W^{\pm}, Z^0$ is

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} = & \underbrace{\tfrac{1}{2}\partial_\mu H \partial^\mu H}_{\text{kinetic term}} - \underbrace{\lambda v^2 H^2}_{\text{mass term}} - \underbrace{(\,\lambda v H^3 + \tfrac{1}{4} H^4\,)}_{\text{self coupling}} \\ & + \underbrace{\tfrac{1}{2}\, g^2 v H\, W^+_\mu W^{-,\mu} + \tfrac{1}{4}\,(g^2 + g'^2) v H\, Z_\mu Z^\mu}_{\text{triple gauge boson couplings}} \\ & + \underbrace{\tfrac{1}{4}\, g^2 H^2 W^+_\mu W^{-,\mu} + \tfrac{1}{8}\,(g^2 + g'^2) H^2 Z_\mu Z^\mu}_{\text{quartic gauge boson couplings}} \quad , \end{aligned} \tag{3.1}$$

encoding its properties and couplings. The mass of the Higgs boson is therefore given as

$$m_H = \sqrt{2\lambda}\, v = \sqrt{-2\mu^2} \tag{3.2}$$

with unknown parameters λ and μ.

The following section summarises constraints and phenomenological properties of Higgs boson production and decays at the LHC. A precise theoretical understanding of these is necessary to optimise analysis strategies and compare measurements with accurate predictions. Thereafter, a short overview of recent experimental results is given.

3.1 Constraints on the Higgs Boson Mass

The SM introduced in Chapter (2) has overall 19 free parameters. Constraining these with results from calculations and measurements allows to the prediction of experimental observables.
The main parameter determining the phenomenology of the Higgs boson is its mass, which by theoretical arguments is bound within a certain range as outlined in the following.

Unitarity Constraint The longitudinal degrees of freedom for weak vector bosons introduced by the Higgs mechanism dominate the tree-level contributions in scattering processes such as $W^+W^- \to W^+W^-$.
The amplitude of this process scales with the available energy in the scattering. Without a regulating mechanism, divergences would occur in its calculation causing unitarity violation. The Higgs mechanism introduces diagrams which cancel these divergences in dependence of Higgs boson mass. This allows to set an upper bound of 1 TeV on the mass as unitarity is required in order for the theoretical model to hold [52].

Vacuum Stability and Triviality Bound Lower limits are derived from a stability constraint on the vacuum expectation value. A low mass Higgs boson would lead to negative corrections to the parameter λ, which in turn would cause the quartic self coupling in Eq. 3.1 to become negative. This would imply an unstable vacuum. λ is therefore required to be > 0 which imposes a lower bound on the Higgs boson mass.
A further theoretical argument, the triviality bound, imposes upper limits for m_H in order for the SM to remain self-consistent and perturbative up to a scale Λ_{UV}.

Combining these arguments, limits from theoretical arguments on the Higgs boson mass are given as

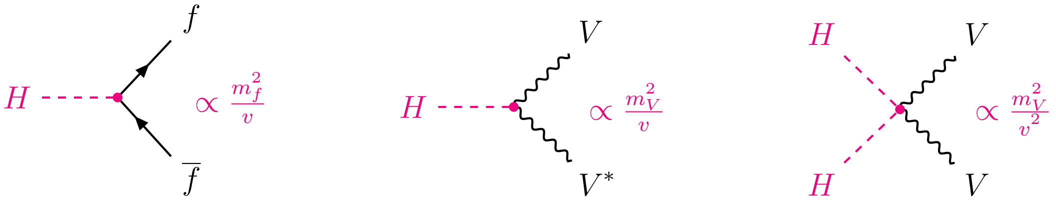
$$\begin{aligned} 50\,\text{GeV} \lesssim m_H \lesssim 800\,\text{GeV} \qquad &\text{for } \Lambda_{\text{UV}} \sim 10^4\,\text{GeV} \\ 130\,\text{GeV} \lesssim m_H \lesssim 180\,\text{GeV} \qquad &\text{for } \Lambda_{\text{UV}} \sim 10^{16}\,\text{GeV} \end{aligned}$$

depending on the cut-off scale Λ_{UV} [45].

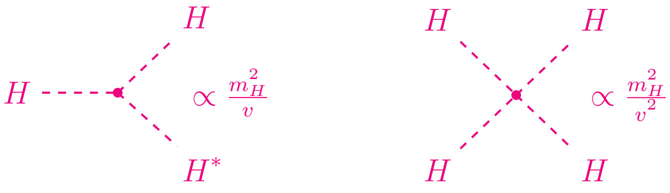
Global Electroweak Precision Measurements Another indirect constraint can be set by analysing combined electroweak precision measurements in context of the complete SM picture.
The global electroweak fits were performed in context of the LEP experiments. In particular direct measurements of the W boson mass, decay width and the top quark mass were used as input to the fit [3, 53]. The best-fit mass was estimated to $m_H = (94 \pm 29)$ GeV which poses a strong constraint on the :w Higgs boson mass range.

3.2 Higgs Boson Production and Decays at the LHC

The terms in Eq. 3.1 describe the Higgs boson interactions. These are sketched together with the Yukawa coupling from Eq. 2.13 in Fig. 3.1.



Yukawa Coupling Triple Gauge Coupling Quartic Gauge Coupling



Triple Self Coupling Quartic Self Coupling

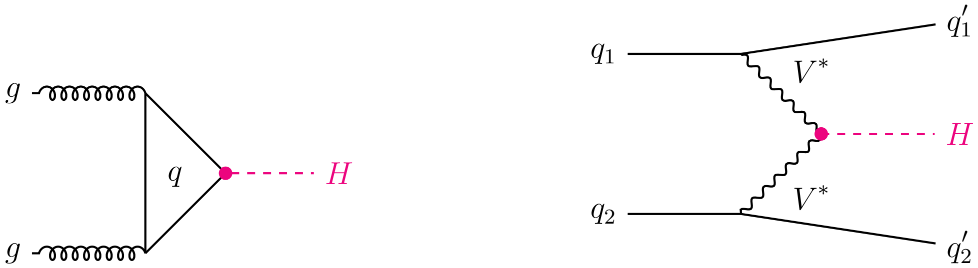
Figure 3.1: Feynman Diagrams of the dominant Higgs boson production channels at *pp*-colliders like the LHC.

3.2.1 Production Modes

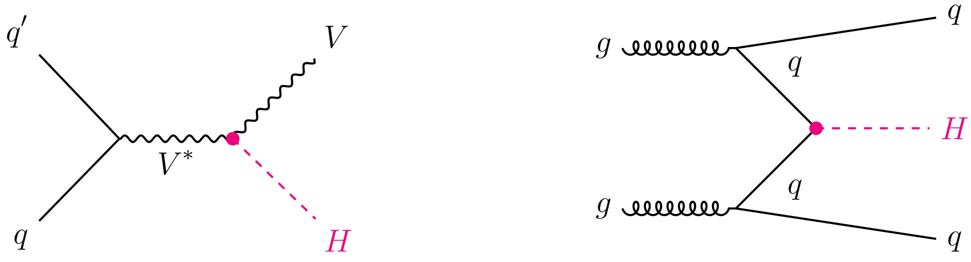
At hadron colliders like the LHC, Higgs bosons are predominantly produced by interactions involving gluons and quarks. Examples of such production processes are depicted in Fig. 3.2 and a consequence of the respective couplings.
The numerical values quoted in the following are based on calculations for a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and a Higgs boson mass of $m_H = 125\,\text{GeV}$ [54]. Further analysis-specific details are given in Sec. 7.2.

The gluon-gluon fusion ggH in Fig. 3.2(a) is the dominant production mode at LHC. with a cross section depending on the gluon momentum fraction in the proton and on radiative corrections. This process is mediated by triangular loops of heavy quarks, primarily the top-quark in the SM due to its large Yukawa coupling. Cross sections are calculated at the N^{3}LO with additional use of NLO electroweak corrections [55–58] to $\sigma(pp \to H) = 48.6^{+4.6\,\%}_{-6.7\,\%}\,\text{pb}$. Higher-order corrections play an important role for this process, caused by a large factorisation and renormalisation scale dependence. The quoted uncertainties arise from parton density function and α_s uncertainties as well as from the missing higher order corrections and are sizeable compared to the following production modes. Events are simulated with at least one additional jet, which recoils against the Higgs boson and therefore gives a distinct signal topology with a high energetic jet and boosted Higgs boson decay products.

The second most dominant production mode originates from vector-boson fusion VBFH sketched in Fig. 3.2(b). Two quarks radiate weak bosons at tree level, which form a Higgs boson. Both quarks propagate in opposite directions with wide separation and create two well-separated jets with large transverse momenta in the forward and backward detector region. This signature is used to efficiently differentiate signal from brackground contributions. Given the lack of colour flow in the central detector region for this process, only small additional jet activity is expected in this direction.
Higher order effects are taken into account up to an approximate-NNLO QCD order for the cross-section calculation. For the latter, NLO electroweak corrections are applied [59–61]. The cross section is estimated to $\sigma(pp \rightarrow qqH) = 3.782^{+0.4\,\%}_{-0.3\,\%}$ pb, i.e. an order of magnitude smaller than for the ggH process.



(a) gg-Fusion (ggH) **(b)** Vectorboson-Fusion (VBFH)



(c) Higgs-strahlung (VH) **(d)** Associated Production (ttH)

Figure 3.2: Feynman Diagrams of the dominant Higgs boson production channels at pp-colliders like the Lhc.

In Higgs-strahlung processes shown in Fig. 3.2(c), a weak boson $V = W, Z$ radiates a Higgs boson. The cross section is calculated at NNLO QCD precision with corrections from radiative electroweak effects at NLO applied [62–64]. For $V = W(Z)$ the values are $\sigma(pp \rightarrow VH) = 1.373^{+0.5\,\%}_{-0.7\,\%}$ pb and $0.8839^{+3.8\,\%}_{-3.1\,\%}$ pb, respectively. In case of leptonic decays of the weak bosons, efficient triggers are used which allow to distinguish between signal and background events.

Finally, the fourth most dominant production channel at the Lhc is the associated production of a Higgs boson with pairs of top-quarks. The corresponding sketch in Fig. 3.2(d) reveals that this channel can be used to measure the Yukawa coupling to top-quarks directly. Its cross section has been calculated to $\sigma(pp \rightarrow t\bar{t}H) = 0.5071^{+5.8\,\%}_{-9.2\,\%}$ pb at QCD NLO accuracy. The phase space limitations due to the two heavy top quarks are reason for the comparable small cross section. Events from this process are not expected to contribute significantly to the final states in this analysis.

The above values are summarised in Fig. 3.3 (left), showing the dependence of the cross section on the Higgs boson mass in the relevant regions.

3.2.2 Decays

The branching ratios (BR) of Higgs boson decays are determined by the decay products' masses, given their proportionality to the coupling strength.
For the calculation of the BR at the highest available order, the programs HDECAY and PROPHECY4F are used [65, 66]. Uncertainties originate from missing higher order corrections and parameters such as the decay products masses and the strong coupling constant.
For the $H \to \tau\tau$ decay, the BR is estimated with a NNNLO QCD accuracy and NLO EW corrections at $m_H = 125\,\text{GeV}$ to

$$BR(H \to \tau\tau) = 0.06272^{+1.17\%}_{-1.16\%}. \tag{3.3}$$

In Appendix A, a summary of the simulated samples and values used in this thesis is provided. The BR for all possible decay modes are depicted in Fig. 3.3 (right).

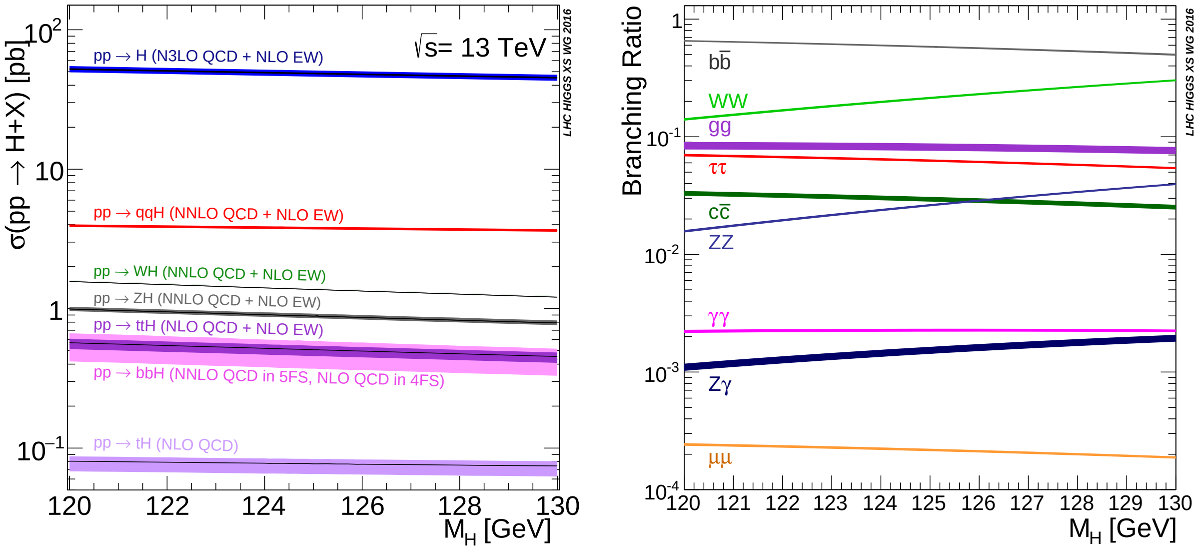


Figure 3.3: Production cross sections (left) and branching ratios (right) for a Higgs boson with a mass around 125 GeV calculated at $\sqrt{s} = 13\,\text{TeV}$ [54].

3.3 Direct Measurements and Recent Experimental Results

The Higgs boson coupling strengths are proportional to the masses of the fermions and gauge bosons, as noted in Fig. 3.1. Combining this with experimental limitations such as reconstruction efficiencies, background contamination and available energy ranges, only certain processes provide experimental sensitivity.

Prior to the LHC era, particle physics experiments investigated various energy ranges and restricted the relevant search range significantly. The results of these experiments are summarised in Sec. 3.3.1.
On July 4th 2012, both the ATLAS and CMS collaborations independently announced the discovery of a new boson [8] with a background fluctuation probability of $p_0 = 1.7 \times 10^{-9}$. This result was mainly driven by bosonic decays of the Higgs boson to pairs of photons γ and Z^0 bosons as detailed in Sec. 3.3.2. The discovery is based on the combined datasets recorded in 2011 and 2012. The various studies performed since then, on both RUN I and RUN II datasets, confirmed the SM-like nature of this resonance, in particular its scalar spin and preference of the $JP = 0^+$ quantum number hypothesis [67].
Measurements of the Higgs boson triple and quartic couplings depicted in Fig. 3.1 are challenging at the LHC given the large background contributions and small event rate. Future collider projects, like the International Linear Collider (ILC), will be able to provide complementary measurements [68].

3.3.1 Pre-LHC Results

The aforementioned electroweak precision results of LEP launched a large campaign of Higgs boson searches prior to the LHC. The preferred light Higgs boson mass motivated direct searches at existing collider experiments, as summarised in the following.

Large Electron-Positron Collider

Direct Higgs boson searches were conducted by the ALEPH, DELPHI, L3 and OPAL experiments at the Large Electron-Positron Collider LEP at CERN. With a total amount of $2.4\,\text{fb}^{-1}$ data at a centre-of-mass energy between 189 GeV to 209 GeV, a lower bound on the Higgs boson mass was estimated to $m_H > 114.4\,\text{GeV}$ [69].
These measurements focused on model-independent searches in Higgs-strahlung processes involving a Z boson, in particular $ZH \to q\bar{q}b\bar{b}$. Given the well-defined initial state at lepton colliders, events were completely reconstructable and signatures with b-quarks were efficiently exploited.

Tevatron

The CDF and DØ experiments at the proton-antiproton collider TEVATRON at FERMILAB recorded in combination $10\,\text{fb}^{-1}$ data at $\sqrt{s} = 1.96\,\text{TeV}$. Measurements focussing on

Higgs-strahlung processes and final states with b-quarks resulted in an exclusion of the mass region $147 < m_H < 180\,\text{GeV}$. Additionally, an excess with a global significance of 2.7σ was observed in the range between $115 < m_H < 140\,\text{GeV}$ [4, 70].

In Fig. 3.4 a summary of the combined exclusion limits from the experiments at LEP and TEVATRON is shown.

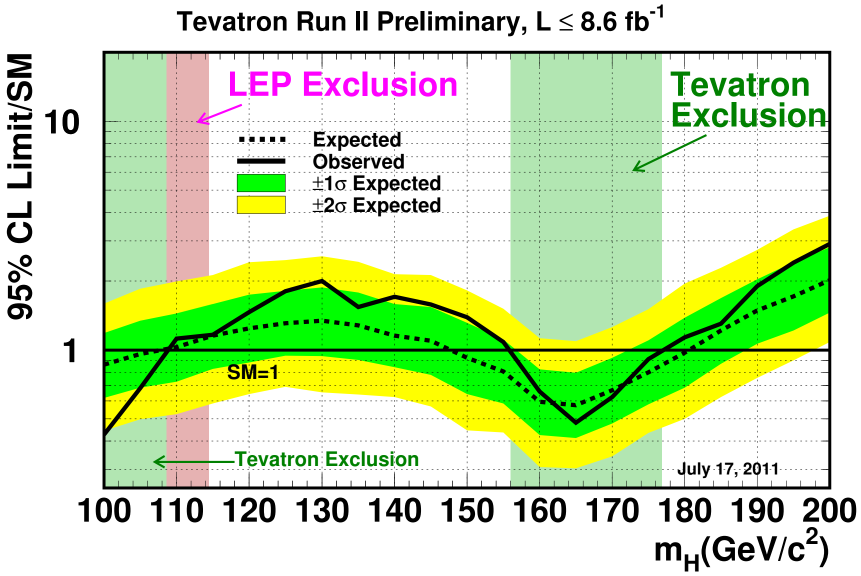


Figure 3.4: Shown are the observed and expected upper limits on the cross section ratio to the SM prediction at a 95% confidence level for the combination of CDF and DØ analyses. Additionally, the combined LEP exclusion range is indicated [4].

3.3.2 Discovery and Measurements in Bosonic Decays

The higher energy and event rates at the LHC offer a significantly improved sensitivity for Higgs boson related measurements. However, this comes at the cost of large multijet background contribution, which imposes experimental challenges in reconstruction and analysis.

Higgs boson decays to bosons offer relatively clean signatures and thereby constitute important measurement channels. The decay into two photons $H \to \gamma\gamma$, initiated through quark loops similar to Fig. 3.2(a), has a low branching at the per-mill level. Nevertheless, it plays an essential role given its accurate energy resolution achieved in the experiments at the LHC. Highly segmented calorimeter systems are able to differentiate prompt photons from neutral pion decays, providing an effective handle against backgrounds. Decays into pairs of Z^0 bosons[1] subsequently decaying to four leptons $\ell = e, \mu$ offer a very clean

---

[1] One of them being virtual.

signature and the possibility to fully reconstruct the Higgs boson mass from the four decay products.

The bosonic decay mode with the highest BR, the $H \rightarrow W^{\pm}W^{\mp}$ process, provides a sensitive experimental channel. In particular, W pair decays into an $e^{\pm}, \mu^{\mp}$ final state with a large momentum imbalance from the escaping neutrinos offer a characteristic signature. The combination of these three channels, as published in Summer 2012 [7] resulted in the observation of a new bosonic particle, consistent with the SM Higgs boson predictions at $m_H = 125\,\text{GeV}$. Fig. 3.5 shows the measured local background fluctuation probability for the combination. The bosonic decay channels $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ$ were subsequently used for a precise estimation of the Higgs boson's properties. All results are in agreement with SM predictions. In Fig. 3.6, the mass measurements in Run II are shown and compared to the combined Run I result. While the precision of the measurement in the $H \rightarrow \gamma\gamma$ channel is limited by systematic uncertainties, $H \rightarrow ZZ$ is restricted by the available dataset size.

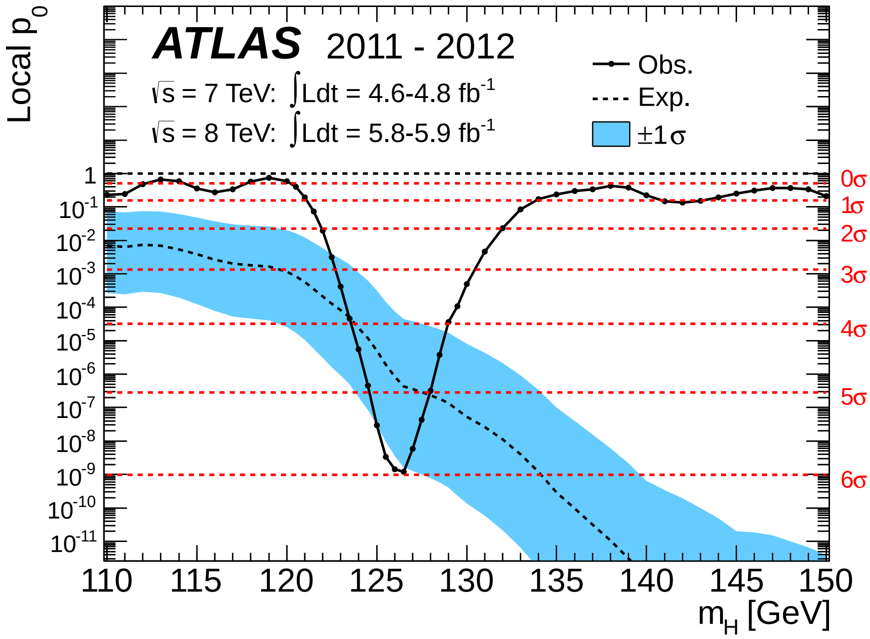


Figure 3.5: Local background fluctuation probability (black) and corresponding significance (red) for the combination of the searches for $H \rightarrow \gamma\gamma$, $H \rightarrow ZZ$ and $H \rightarrow W^{\pm}W^{\mp}$ processes in Summer 2012 with the Atlas detector [7].

3.3.3 Measurements in the Yukawa Sector

Channels which allow to probe for the Yukawa coupling of Higgs bosons are difficult to distinguish from the background sources arising at the Lhc.

In particular, the $H \rightarrow b\bar{b}$ decay offers the highest branching ratio in the fermionic sector but is experimentally difficult to distinguish from the large multijet background. Measurements

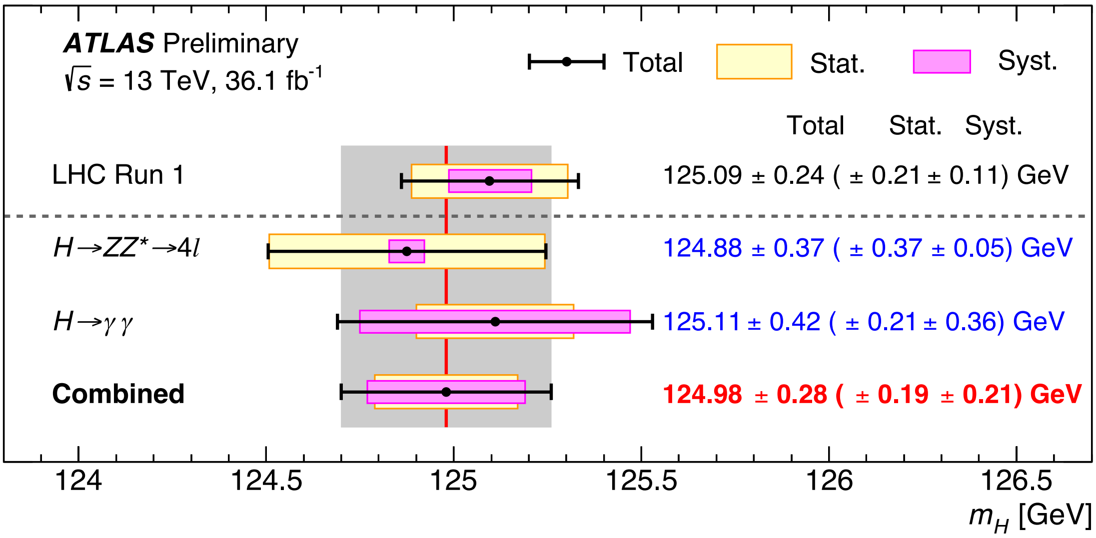


Figure 3.6: Summary of Higgs boson mass measurements in $H \to \gamma\gamma$ and $H \to ZZ$ decays. The RUN II results based on the 2015 and 2016 dataset are compared to the RUN I combined CMS and ATLAS RUN I measurement [71].

are hence performed in association with weak bosons or top-quarks to ensure efficient triggering and possibilities for background suppression.

The Higgs boson decay into a pair of oppositely charged τ leptons $H \to \tau^+\tau^-$ on the other hand provides distinguishable event topologies. having a sufficiently large branching ratio to reach sensitivities. In particular, the distinct topology of the VBFH production mode allows for efficient background suppression. Further details will be discussed in depth in Chapter (6) of this thesis.
The ATLAS collaboration performed a search for $H \to \tau\tau$ decays with data recorded in the first research run of the LHC. Overall, $20.3 + 4.5\,\text{fb}^{-1}$ data with centre-of-mass energies of $8 + 7$ TeV were analysed in combination of all possible $\tau\tau$ final states [9]. An observed (expected) excess of $4.5\,\sigma$ $(3.4\,\sigma)$ gave a clear evidence of this process. The signal strength was measured to

$$\mu_{H\tau\tau}^{\text{RUN I}} = 1.43\ ^{+0.3}_{-0.3}\ (\text{stat})\ ^{+0.3}_{-0.2}\ (\text{syst})\ ^{+0.1}_{-0.1}\ (\text{theo})$$

and is consistent with the SM prediction. In Fig. 3.7, the results for μ for all channels and investigated topologies are summarised. The analysis strategies were chosen similar to the ones presented in Chapter (6). Significant differences to the analysis presented in this thesis arise from changes in the RUN II data taking conditions:

- higher integrated luminosity of the analysed dataset with $36.1\,\text{fb}^{-1}$, implying a larger number of recorded events
- higher centre-of-mass energy of 13 TeV resulting in increased cross sections, in particular for rare decays
- upgraded ATLAS experiment with improved tracking system, reconstruction software

and particle identification methods

- increased activity in the detector from pileup, underlying event and other multiple partonic interactions.

Another major difference arises from the estimation of $Z/\gamma^* \to \tau\tau$ background contribution. In the RUN I analysis, the embedding method [72] is used to model this process. A $Z/\gamma^* \to \mu\mu$ dataset with high-purity is selected and both μ are replaced with simulated τ leptons, assuming lepton universality. The so constructed events are then processed with the ATLAS reconstruction software. This creates a hybrid dataset with jet activity taken from actual collision data, with injected τ decays from simulation, corrected for τ specific topology and kinematics. The RUN II analysis described in this thesis relies on pure simulation to modell this process.

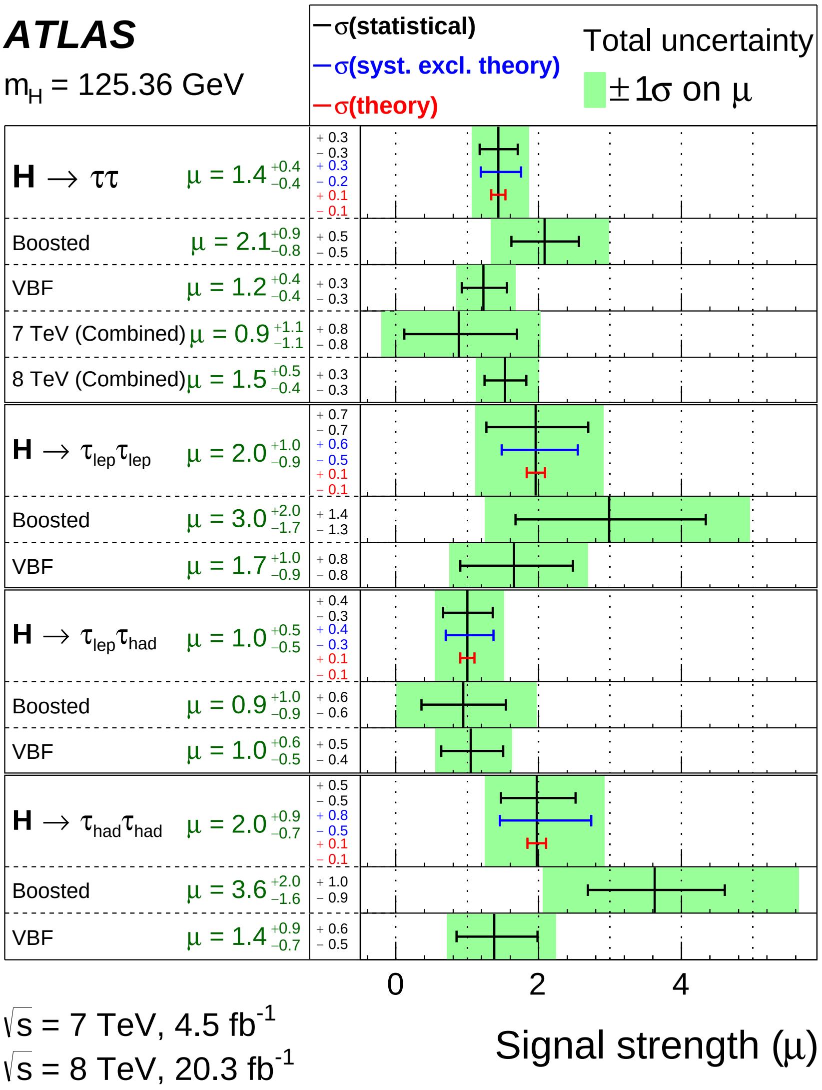


Figure 3.7: Signal strength for the combined RUN I $H \to \tau\tau$ measurement [9]. The analysis strategies similar to the RUN II approach described in Chapter (6).

CHAPTER 4

The LHC and the ATLAS Experiment

In this chapter, the experimental apparatus used to record the analysed dataset is presented. After a brief overview of quantities and strategies defining the baseline of measurements in hadron colliders, a technical overview of the Large Hadron Collider (LHC) is provided in Sec. 4.2. At the LHC, bunches of protons p are collided and the resulting collision products are measured in various experiments. The studies in this thesis were performed in context of the ATLAS[1] experiment, which is detailed in Sec. 4.3.

4.1 Phenomenology of *pp* Collisions

The protons colliding in the LHC constitute quarks and gluons, the partons. In the SM, these partons partake in interactions as described in Chapter (2). Any quarks or gluons formed in these hard interactions generate showers of additional partons undergoing radiation or splitting processes. Due to the confinement of colour-charged particles discussed in Chapter (2), the parton shower results in bound, colour-neutral hadronic particles. This hadronisation finally yields measurable physics objects: the hadrons interact with the detector material and deposit energy in the various detector systems described in Sec. 4.3. The complexity of these processes is visualised in Fig. 4.1, depicting the propagation of particles from the initial, hard interaction to the final, collimated particle showers, called jets. From the characteristics of these jets, conclusions about the initial particles can be drawn.

The prediction of cross sections, kinematics and other observables of processes should account for all possible contributions to a final state. Complexity and divergencies in the analytical calculations of a complete final state require a plethora of theoretical and experimental strategies specific to hadron colliders. The following sections will give a brief overview and explain each stage of Fig. 4.1 in more detail.

---

[1] **A** **T**oroidal **LHC** **A**pparatu**S**

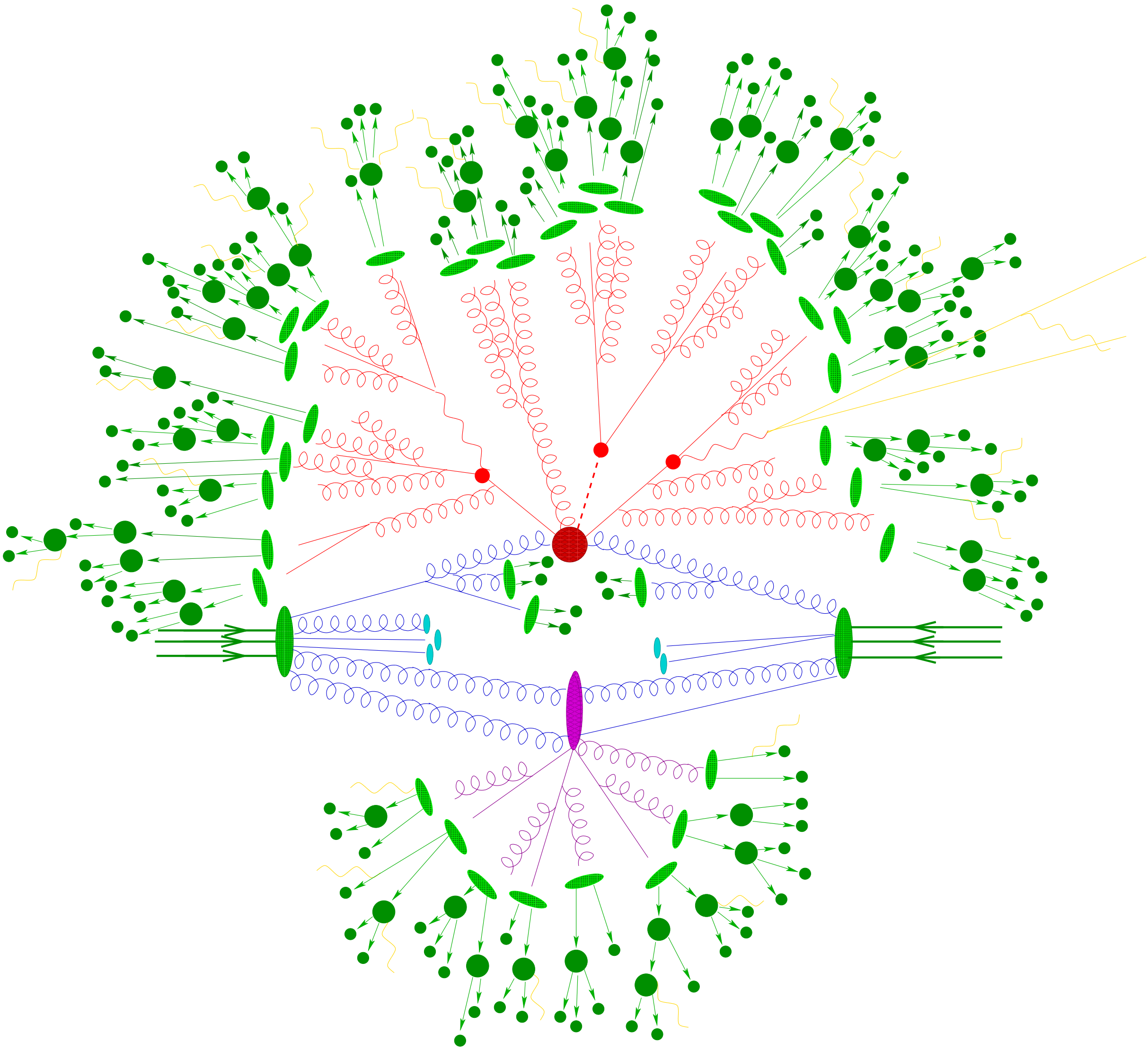


Figure 4.1: Representation of a *pp*-collision and the propagation of initial particles from the hard interaction (dark red circle) to the jet constituents (dark green circles) via the parton shower of quarks and gluons (light red) and the hadronisation process (light green). Further contributions to a final state can arise from secondary interactions (violet circle) and photon radiation (yellow lines) [73]. Contributions from these secondary interactions are called pile-up and arise from additional *pp*-collisions within a bunch or from previous (following) bunch crossings.

4.1.1 Parton Density Functions

It is essential to correctly describe the energy-dependent structure of a proton for any final state prediction in hadron colliders. The content of low-energetic protons can be approximated by three valence-quarks $p = (uud)$, whereas for high energy ranges at the LHC, contributions from gluons and sea-quarks become important. A visualisation of this energy dependent proton content distribution is shown in Fig. 4.2. The energy scale at which the proton structure is probed is called factorisation scale $Q = \mu_F^2$.

The fraction of the protons' total momentum x_i that a parton i carries[1] can be described by a probability (or parton) density function PDF $f_i(x_i, \mu_F^2)$.A perturbative description of these PDFs is not possible, due to divergences caused by the large QCD coupling constant in the proton-internal energy regime. The PDFs are measured instead and extrapolated as explained in the following.

Deep inelastic scattering experiments, like the electron-proton collider HERA, fixed-target experiments or hadron colliders such as TEVATRON and the LHC provide PDF measurements at certain energy scales [5, 75]. The DGLAP[2] QCD evolution equations [76] offer the possibility to extrapolate the PDF measurement to different energy regimes. This extrapolation is based on a parametrisation of the PDF, for which a variety of approaches exists. In this thesis, parametrisations from the NNPDF [77] and CTEQ [78] collaborations are used. These PDFs introduce an uncertainty based on measurement and extrapolation.

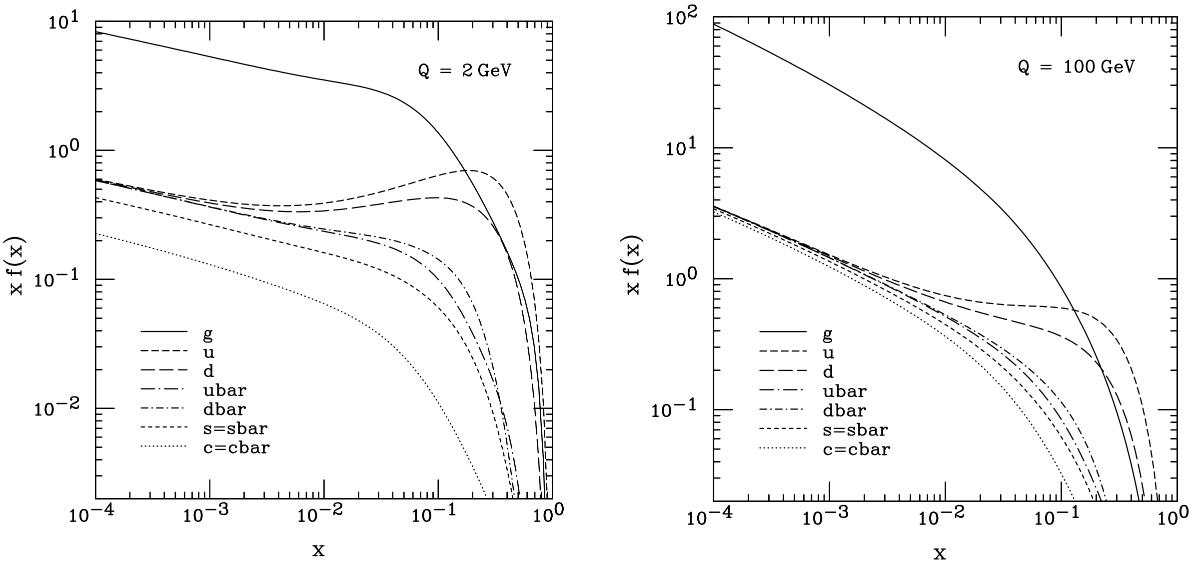


Figure 4.2: As an example, the standard set of CTEQ6M parton distribution functions are shown [78]. Two different energy scales $\mu_F^2 = Q = 2\,\text{GeV}$ and $Q = 100\,\text{GeV}$ were used to probe the proton structure. The solid and hatched lines indicate the probability to find a respective parton at a fraction x of the proton's momentum Q.

The chosen PDF set influences the modelling of simulated observables. By comparing these for varied sets, the impact of the PDF set uncertainty on the measurement and final result can be studied.

---

[1] Referred to as the Bjorken-x [74].

[2] Named after Y. Dokshitzer, V. Gribov, L. Lipatov, G. Altarelli and G. Parisi.

4.1.2 Cross Sections and Parton Showers

The rate dN at which events in a collider occur can be expressed as

$$dN = L\ \mathrm{d}\sigma \tag{4.1}$$

with the collider-dependent instantaneous luminosity L quantifying the rate of possible interactions over a fixed time and the differential cross section $\mathrm{d}\sigma$. The latter describes the probability for the occurrence of a specific process when two particles interact. While L is defined by beam parameter settings in the collider, $\mathrm{d}\sigma$ is a calculable quantity based on the transition matrix element $|\mathcal{M}|^2$ and a phase space factor $\mathrm{d}Q$, describing all possible kinematic configurations for a given final state. The matrix element depends on the physics process in question and describes the probability for a transition of an initial parton state to a specific final state. While the matrix element can in principle be calculated from the Lagrangian presented in Chapter (2) using perturbation theory, the calculation of a final state as depicted in Fig. 4.1 requires further, hadron-collider specific methods described in the following.
The calculation of the total differential cross section $\mathrm{d}\sigma(pp \to X)$ for two colliding protons to produce a final state X is derived from the $\mathrm{d}\sigma_{ij\to X}$ for each individual parton i and j in the two protons. These partonic $\mathrm{d}\sigma$ can be predicted, as the matrix element for interactions of individual partons can be calculated with perturbative techniques. The total $\mathrm{d}\sigma(pp \to X)$ is then given as sum over all possible parton combinations interacting in the colliding protons which can result in the final state X via

$$\mathrm{d}\sigma(p_1p_2 \to X) = \sum_{i,j} \int \mathrm{d}x_i \int \mathrm{d}x_j\ f_i^1(x_i, \mu_F^2) f_j^2(x_j, \mu_F^2) \mathrm{d}\sigma_{ij\to X}(x_i, x_j, \alpha_s, s_{i,j})\ . \tag{4.2}$$

Here, $f_{i,j}$ denotes the PDFs introduced in the previous section. $s_{i,j} = x_{i,j}s$ is the centre-of-mass energy of the interacting partons. The expression in Eq. 4.2 is a convolution of the partonic $\mathrm{d}\sigma$ with the PDF and known as QCD factorisation theorem.

Perturbative calculations of the matrix elements for the complete parton shower depicted in red in Fig. 4.1 are necessary to account for their sizeable contributions due to the large QCD coupling constant α_s. The perturbative precision to which a matrix element has been calculated depends on the order of α_s. Contributions from higher order processes are estimated using phenomenological models known as parton shower algorithms. These introduce additional partons for example from gluon splitting $g \to q\bar{q}$ or radiation processes $q \to qg$, called initial and final state radiation (ISR/FSR).

4.1.3 Hadronisation

Colour-confinement dictates that no unbound, propagating quarks and gluons can be observed. Instead, they form bound hadronic states sketched by the light green circles in Fig. 4.1. This hadronisation is not covered by the perturbative calculations nor by the parton shower algorithms. To combine the quarks and gluons from the parton shower to hadrons, several models exist.

The Pythia generator for example uses the string fragmentation [79], which utilises colour strings connecting close-by partons. Breaking these strings up results in the production of a new quark pair in this model. The colour of the strings determines how hadrons are formed from the initial partons.
Hadronisation modelling in Sherpa, the main generator for background events in this thesis, relies on the cluster fragmentation [73]. Each gluon from the parton shower is split into a quark-pair. The resulting quarks are clustered into colour singlet states with hadrons arising from the decay of these clusters. Both methods rely on parameter tuning to dedicated measurements.
The particles that are detected are baryons and mesons, which can be simulated to model, for example, detector responses.

4.1.4 Underlying Event and Pile-Up

The remnants of the proton which did not partake in the initial hard interaction depicted as red circle in Fig. 4.1 also contribute to the final state in hadron colliders. These contributions are referred to as underlying event and are estimated using a model of multiple partonic interactions in the Pythia generator. The additional interactions at smaller energy scales than the hard processes, are modelled in a probabilistic way and tuned to dedicated measurements [80].

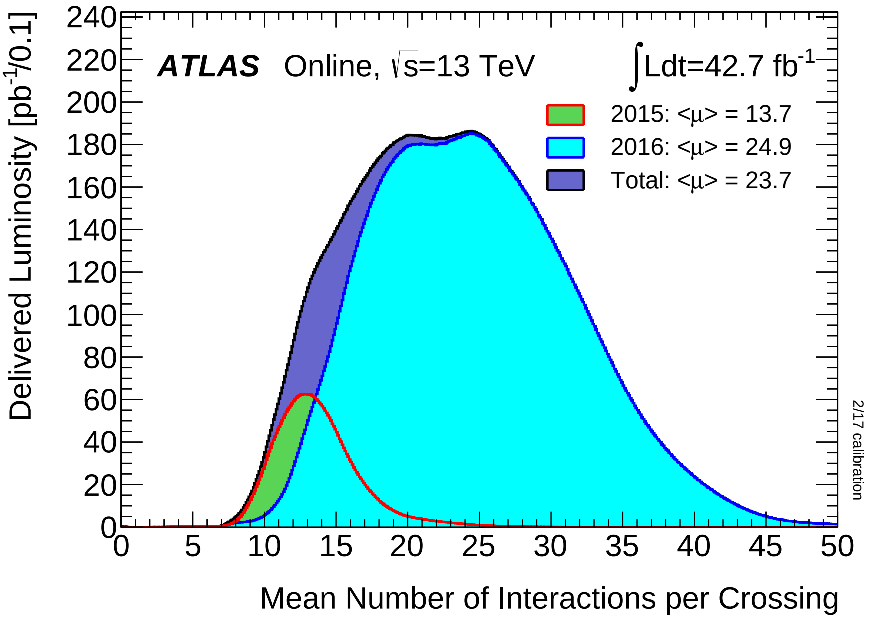


Figure 4.3: Average number of interactions per bunch crossing $\langle\mu\rangle$ in the datasets recorded with Atlas in 2015 and 2016.

At the Lhc, bunches of protons are collided which typically leads to multiple hard proton interactions per event, called in-time pile-up. The spacing between two colliding bunches at the Lhc for the datasets analysed in this thesis is 25 ns. This interval is shorter

than the limited read-out time of the ATLAS detector detailed in Sec. 4.3.Hence, final states from different bunch-crossings can overlap and be recorded in the same moment. These final contributions are called out-of-time pile-up. The recorded profile of the average interactions per bunch crossing $\langle\mu\rangle$ in ATLAS for the 2015 and 2016 data taking period of the LHC is presented in Fig. 4.3. An average of 23.7 interactions is observed which imposes various experimental difficulties as detailed in Chapter (5).
The contributions from underlying event and pile-up are estimated in simulation with a process-independent, minimum-bias selection. These simulated events are overlaid on top of the simulation of hard interaction, parton shower and hadronisation.

4.2 Technical Overview of the LHC

The Large Hadron Collider (LHC) located south of Lac Léman on the Franco-Swiss border is the world's largest and most powerful particle collider. At a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, the LHC collided trillions of proton-bunches - in 2016 alone. Its installment took place in the tunnel of the retired Large Electron Positron (LEP) collider [81], 45 m-150 m below ground and with a circumference of 26.7 km [82]. With two separate beam pipes, the LHC is designed to allow for a second operation mode, the collision of heavy ions, at a design centre-of-mass energy of 2.76 TeV/ Nucleus. Fig. 4.4 shows a schematic overview of the LHC accelerator complex and its various experiments. The four main experiments, located on the main ring are:

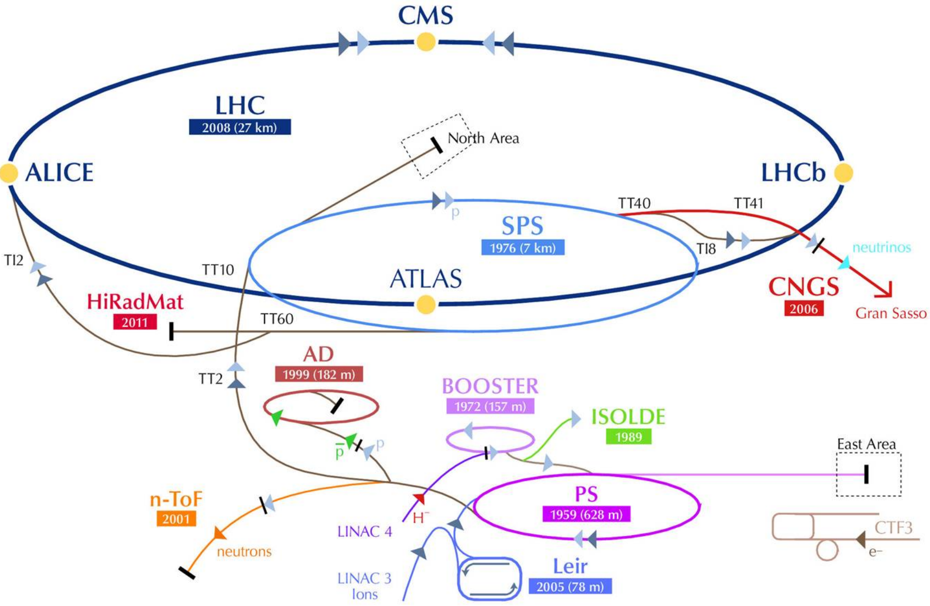


Figure 4.4: The LHC accelerator complex including the various pre-accelerator systems. After initial acceleration in LINAC2, BOOSTER, PS and SPS the proton bunches enter the main LHC ring. The beams are crossed at each of the four main experiments, resulting in collisions of the proton bunches.

- the multi-purpose detectors ATLAS and CMS, designed to measure established SM processes, detect the Higgs boson and potential physics beyond the SM [83, 84].
- the LHcb detector, a specialised experiment for b-physics, in particular studying the nature of $\mathcal{CP}$-violating processes [85].
- the ALICE experiment, specifically designed to cope with lead ion collisions, studying quark-gluon plasma states [86].

The starting point of any pp collision lies in a canister of hydrogen gas, filled with protons produced by ionisation and injected into the Linear Accelerator LINAC2 [87]. With an energy up to 50 MeV, bunches of protons then enter the Proton Synchrotron Booster (BOOSTER) which accelerates them to 1.4 GeV and injects the beam in the Proton Synchrotron (PS). They subsequently enter the Super Proton Synchrotron SPS with an energy around 25 GeV. This is the final pre-acceleration step; the bunches are thereafter injected to the LHC main ring with an energy of 450 GeV.

In the LHC, the proton bunches are accelerated counter- and clockwise to a collision energy of 6.5 TeV per beam in 2015 and 2016. The acceleration involves radio-frequency cavities, which are tuned such that the bunches are collimated and spaced with a 25 ns distance. If protons in a bunch carry an energy different from the bulk, they are accelerated or decelerated due to the oscillating electromagnetic fields in the cavities until they are synchronized with the radio-frequency. With a field strength of up to 8.33 T, superconducting dipole magnets keep the proton bunches in a circular orbit. The lateral expansion is regulated with quadrupole and higher multipole magnet system, to keep the beam focussed. The superconducting properties are realised by cooling the magnet systems with liquid helium to 1.9 K. The LHC is designed to collide beams with a maximum centre-of-mass energy of 14 TeV.

The event rate in Eq. 4.1 is defined by the collider-specific instantaneous luminosity, which for the LHC can be expressed as

$$L = \frac{n_b N_1 N_2 f_{\text{rev}} R}{\pi \sqrt{\sigma_{x,1}^2 + \sigma_{x,2}^2} \sqrt{\sigma_{y,1}^2 + \sigma_{y,2}^2}} \tag{4.3}$$

with the number of proton bunches n_b, the number of protons per bunch for both beams $N_{1/2}$ and the bunch spacing $1/f_{rev}$ [88]. A loss factor R accounts for residues in the beam pipe and other, luminosity-reducing effects. In the denominator of Eq. 4.3, the horizontal and vertical (x and y) sizes of the colliding bunches (labelled 1 and 2) is encoded. Therefore, a smaller beam width corresponds to a higher L.

Integrating the luminosity over a period of time gives a measure of the number of collision events N, quantifying the size of the collected data

$$N = \mathcal{L}\sigma(pp \to X) \quad \text{with} \quad \mathcal{L} = \int L dt \, . \tag{4.4}$$

Performance of the LHC in 2015 and 2016 Since the very first beam circulation on the September 10th 2008, the LHC was subject to countless upgrades, slowly approaching the design machine parameters. The first collisions in November 2009 at 450 GeV beam

energy mark a milestone in the history of particle physics and were successfully detected by all four experiments.
After the first successful operation period between 2010-2012, named Run I, the LHC underwent a shut-down phase in preparation of Run II. A total amount of 28 fb^{-1} collisions were delivered by the LHC to ATLAS at 7 TeV and 8 TeV centre-of-mass energy in Run I. The analyses presented in this thesis are based on collisions recorded with the ATLAS detector in 2015 and 2016. In Fig. 4.5, the delivered and recorded integrated luminosity per day for the ATLAS detector in this period is shown. The peak instantaneous luminosity reached in 2015 is $L = 5 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$, which was more than doubled in 2016 to $L = 13.8 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$ with $\sim$ 2200 bunches per beam, reflecting the outstanding performance of the machine. The delivered integrated luminosity is a measure of the total number of collisions in the ATLAS detector. For these data taking periods, these are

- 2015: $\mathcal{L} = 4.2\,\text{fb}^{-1}$ with $\langle\mu\rangle$ =13.7
- 2016: $\mathcal{L} = 38.5\,\text{fb}^{-1}$ with $\langle\mu\rangle$ =24.9

with the average number of interactions per bunch crossing $\langle\mu\rangle$ as presented in Fig. 4.3. The actual recorded number of events ready for analysis depend on various detector- and analysis-specific factors and is discussed in Sec. 7.1.

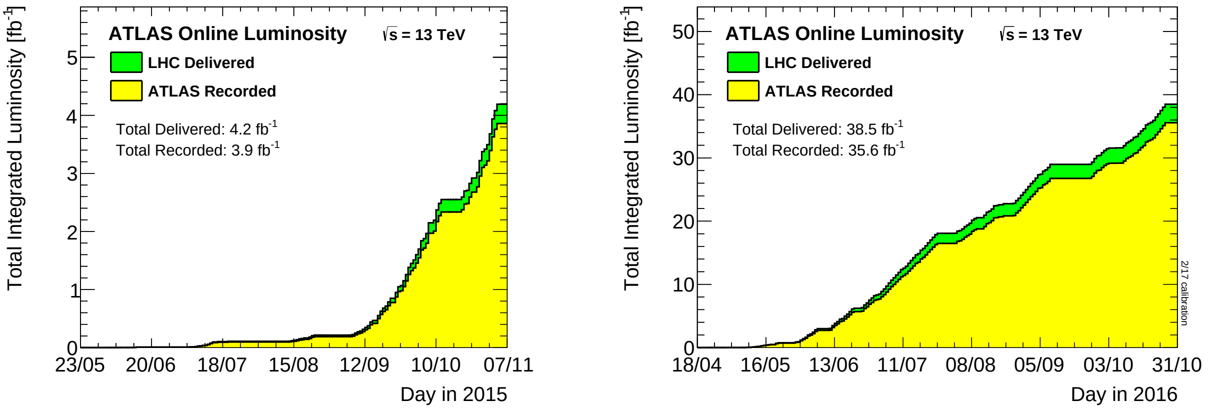


Figure 4.5: The delivered and recorded integrated luminosity per day for 2015 (left) and 2016 (right) at the ATLAS experiment.

4.3 The ATLAS Experiment

The ATLAS experiment is one of the largest physics collaborations to date, connecting 3000 active scientists from 38 countries. It covers a wide variety of scientific disciplines, from accelerator and detector physics, to data analysis and computing. Measuring 44 m in length with a radius of $\sim$ 12.5 m and a weight of 7 000 t, the ATLAS detector is the largest of the four experiments at the LHC. Its multipurpose design and targeted sensitivity to new physics imposes various experimental challenges, amongst them the high event rate and the resulting large background contributions from pile-up and underlying events. To cope with these challenges, a very precise particle identification and radiation-safe detector components are necessary to handle high occupancies.

A schematic layout of the Atlas detector's onion-structure is shown in Fig. 4.6. The proton beams circulate in the beam-pipes at the very centre of the detector. This pipe is surrounded by the inner detectors which are immersed in a solenoidal magnetic field and measure signals from charged tracks. The subsequent layer is formed by the electromagnetic and hadronic calorimeter systems, where electromagnetic or hadronic showers are induced by traversing particles. These particles deposit their energy in the calorimeter cells, a fact in particular exploited in the reconstruction of jets and hadronically decaying τ leptons. The outer layer is formed by the muon spectrometers and the eponymous unique toroidal magnet systems. After a description of the coordinate system used in Atlas, each detector component is described in more detail in the following sections.

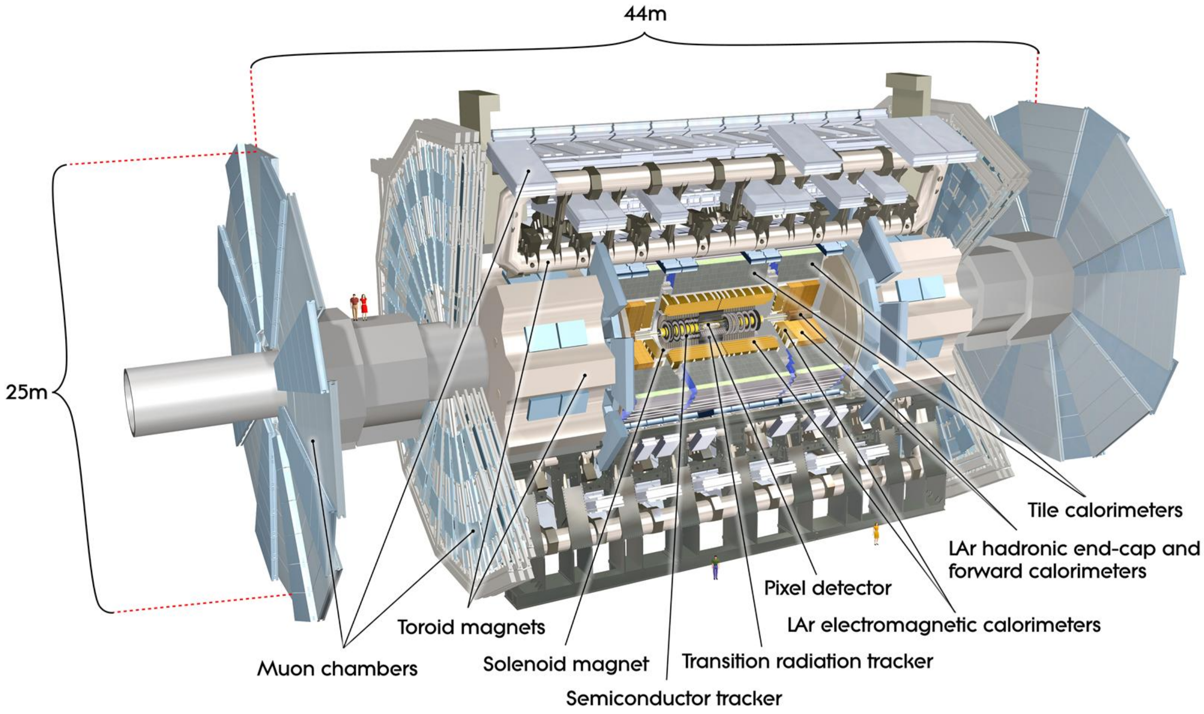


Figure 4.6: Schematic view of the Atlas detector, highlighting the main inner detector, calorimeter and muon system components [83].

4.3.1 Coordinate System

The position of signals in the Atlas detector is described in a right-handed cylindrical coordinate system, a convenient choice given the detector's shape. The origin of this system is set to the middle point of the detector, with the x-axis pointing to the centre of the Lhc-ring and the y-axis upwards. The z-axis points along the beam pipe. The azimuthal angle ϕ and the polar angle θ completely define the position of a signal in the

x-y plane.

An important characteristic of hadron colliders is the choice of reference frame to express kinematic quantities. While the collision energy E_p per proton refers to the centre-of-mass frame of the two colliding protons, the actual collision partners are the partons i, j in the proton with unknown momentum fraction $x_{i/j}$. This leads to an longitudinal momentum due to a possible imbalance between x_i and x_j. The final objects from such a collision will receive a boost along the z-axis, an intrinsic characteristic of hadron colliders. For this reason, the polar direction is commonly expressed as a function of the energy and momentum in z-direction, called rapidity

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad .$$

An advantage of this approach is, that intervals Δy are invariant under Lorentz transformations along the beam direction. This implies that the unknown longitudinal momentum in the collision does not affect Δy. The rapidity can be approximated in the ultra-relativistic limit, i.e. if the mass of a given particle is significantly smaller than its momentum, which is a valid approximation for most particles traversing the ATLAS detector. This approximation yields the pseudorapidity η, defined as

$$\eta = -\log \left[\tan \left(\frac{\theta}{2} \right) \right] \quad .$$

The differential cross sections for production of jets is approximately constant in η [2]. This implies, that the number of jets is the same in equal pseudorapidity intervals $\Delta\eta$. With this definition, the angular distances between objects can be described by the variable

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

As discussed in Sec. 4.1, protons contain various partons each carrying a fraction of the proton's total energy. The exact momentum of interacting partons is not known, however, due to energy and momentum conservation the sum of their momenta in the x-y plane are 0. Expressing energy and momenta in this plane therefore allows for an indirect measurement of undetectable particles, which cause an imbalance in the momentum or energy sum in this transverse plane. The transverse momentum p_t and the transverse energy E_t are defined as

$$p_t = \sqrt{p_x^2 + p_y^2} \quad \text{and} \quad E_t = \sqrt{E_x^2 + E_y^2} \quad .$$

An imbalance is called missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ and discussed in detail in Sec. 5.5.

4.3.2 Technical Detector Design

In the following sections each detector component is briefly introduced, based on the more detailed explanations in [89, 90].

4.3.2.1 Magnets

One outstanding, unique characteristic of the ATLAS detector is the design of the magnet system. Two separate fields are used; a solenoid field surrounding the inner detector and a toroidal field interspersing the muon spectrometer. A charged particle traversing these fields is subject to the Lorentz force, resulting in a measurable curvature of its track. This curvature can be used to measure the momenta p_t of the particle. The uncertainty of the measurement increases with increasing p_t, as the measurable curvature gets smaller and the resulting resolution decreases.
The superconducting inner magnetic solenoid coils offer a uniform field strength of 2 T. A non-uniform field is created by the toroid systems, with an average strength of 0.5 T over a large volume.

4.3.2.2 Inner Detector

The main purpose of the first detector surrounding the beam pipe, the inner detector (ID), is the precise measurement of tracks formed by charged particles and the related measurement of interaction vertices.

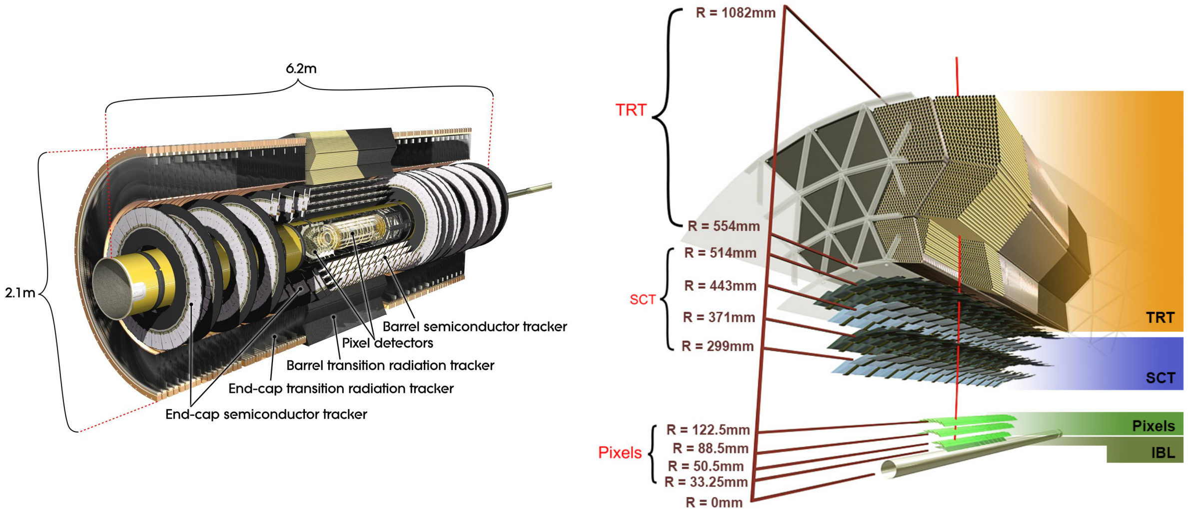


Figure 4.7: Schematic cut-away views of the ATLAS inner detector. The left-hand plots depicts the different layers of the inner detector, while the right-hand side shows different detector regions along the beampipe [83, 91].

A cut-away of the ID and its components is shown in Fig. 4.7 (left). The ID is comprised of three components:

- the pixel detector with the insertable b-layer installed after the first LHC run [92]
- the semiconductor tracker (SCT)
- and the transition radiation tracker (TRT).

Along the z-axis the ID regions are divided into barrel (or fiducial) detector region and end-cap disks for measurement in forward regions as sketched in Fig. 4.7 (right). The coverage of the ID extends to regions $|\eta| < 2.5$, covers the full ϕ-range and has an increasing granularity towards the beampipe.

IBL and Pixel Detector The innermost detector layer around the beampipe is the IBL, installed in May 2014 to cope with the expected high occupancies in RUN II. This layer improves the track resolution and vertex reconstruction efficiency, quantified by the impact parameters d_0 and z_0, which describe the closest distance of a track to the primary vertex in (η, ϕ) and z, respectively. The three subsequent pixel layers are divided in barrel region and three disks in the detector end-caps. The pixel detector alone consists of a ~ 80 million independent read-out channels, pixels, offering a hit resolution $14 \times 115\,\mu\text{m}^2$ for $(R\text{-}\phi) \times (z)$-directions. All of the semiconducting pixels are read out separately by front-end chips which are bump-bonded to each pixel. The pixel detector and the SCT are operated at -10 to $-5\,^\circ\text{C}$ to decrease electronic noise.

Semiconductor Tracker In the SCT semiconducting silicon strips are used for track measurements complementary to the Pixel information. Four barrel layers are each comprised of single sided silicon microstrip detectors, installed with a small stereo angle allowing for the measurement of the azimuthal angle of hit positions. Barrel region and the 9 SCT end-caps offer a maximal hit resolution of 17 µm in the $(R\text{-}\phi)$ plane and 580 µm along the z-axis.

Transition Radiation Tracker The outermost component of the ID is the TRT, which comprises around 400 000 Xenon-gas filled straw tubes with a 4 mm diameter each. Within the tubes, a tungsten wire is suspended and an electrical field forces free electrons to drift towards the wire, which results in a measurable signal. These free electrons emerge from transition radiation of charged particles traversing foils with different dielectric constants around the tubes. The magnitude of the transition radiation depends on the particle type, i.e. a traversing electron results in a larger signal than a pion. This offers an additional handle on particle identification within a range of $|\eta| < 2.0$. The TRT has a resolution of 130 µm in the $(R\text{-}\phi)$ plane and registers on average 36 hits per traversing particle. It is operated at room temperature and has ~ 351000 separate read-out channels.

Momentum Resolution Any physics analysis involving high-energetic particles benefits from a precise measurement of their momentum. The relative momentum resolution of the inner detector was measured [93] after initial calibration to

$$\frac{\sigma_{p_T}}{p_T} = (4.83 \pm 0.16)\, p_T \times 10^{-4}\ \mathrm{GeV}^{-1}\ .$$

The resolution was observed to be independent of the η range but shows a degradation for higher transverse momenta, as the TRT contribution to the track measurement becomes more important.

4.3.2.3 Calorimeter Systems

For energy measurements in ATLAS, a sampling calorimeter system is used. Alternating layers of dense and active material induce particle cascades and measure the resulting energy deposits. These cascades, also called showers, are a result of electromagnetic or strong interactions between traversing particles and the dense, inactive calorimeter material. The system comprises two independent calorimeters - the electromagnetic (ECAL) and the hadronic calorimeter (HCAL), discussed in more detail in the following.

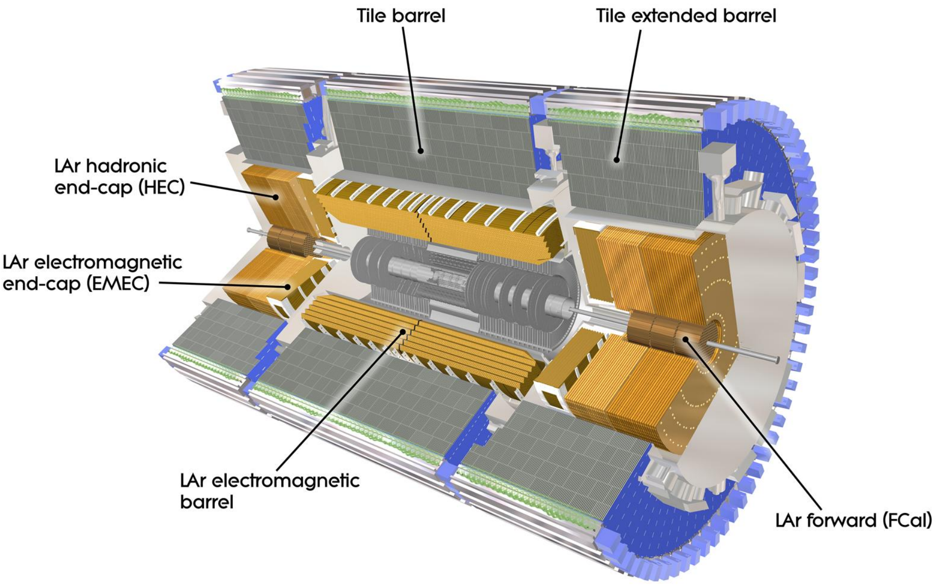


Figure 4.8: A labelled scheme of the ATLAS detector's calorimeter system showing its sub-detector systems [83].

In Fig. 4.8, the calorimeter systems are sketched, their combined coverage extends up to $|\eta| < 4.9$, including the full ϕ-range. The path length required to reduce the number of particles in a cascade to $1/e$ of the original amount in the electromagnetic part of the calorimeter is called interaction length X_0. At $\eta = 0$, the ATLAS calorimeter system has a total interaction length of $11\,X_0$, ensuring that the majority of electromagnetic and hadronically interacting particles deposit all their energy within the respective system. This is important for an accurate energy measurement and ensures a minimum amount of escaping shower particles to the muon spectrometer, referred to as punch-through. The calorimeter's energy resolution can generally be parameterised by three terms

$$\frac{\sigma_E}{E} = \underbrace{\frac{a}{\sqrt{E}}}_{stochastic} \oplus \underbrace{\frac{b}{E}}_{\text{noise}} \oplus \underbrace{\quad c \quad}_{systematic}$$

where $\oplus$ denotes a quadratic sum between the terms. The stochastic term is intrinsic to any sampling calorimeter; it describes the Poisson distributed number of particles in a shower. Noise effects from electronics and pile-up contributions are taken into account in the second term. Finally, systematic effects, e.g. from dead detector material are quantified by the third term. For high energies, the latter is the dominant of the three.

Electromagnetic Calorimeter High energetic electrons or photons passing the inner detector are interacting electromagnetically with the ECAL through bremsstrahlung and e^+e^--pair production. The resulting secondary particles themselves continue to interact with the material, forming the aforementioned particle cascades. Once the energy of a particle falls below a critical value (around 10 MeV), the cascade stops and the initial energy is completely deposited in the calorimeter. The absorber material in the ECAL is composed of lead and copper, while the active cells are filled with liquid argon. Following an accordion-like structure, the material layers are arranged to ensure a full coverage in ϕ. Any charged particle traversing the active cells, ionize the contained argon and the resulting charges are collected in electrodes. This gives an electrical pulse which is proportional to the deposited energy. As sketched in Fig. 4.8, the ECAL is divided in barrel region and two end-caps and covers a range of $|\eta| < 3.2$. The first ECAL layer is finely granulated to facilitate identification of pion-decays into a collimated photon pair. The ECAL is designed to stop the majority of electrons and photons by depositing all of their energy in the calorimeter. The energy resolution of the ECAL was measured [93] to

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus \frac{170\,\text{MeV}}{E} \oplus 0.7\% \; .$$

Hadronic Calorimeter Hadronic particles are typically of minimal ionising nature and hence need denser material to deposit their complete energy. The showers are initiated through electromagnetic or strong interactions of the hadrons with the absorber material, Two parts compose the HCAL:

- The tile calorimeter covering the barrel region with $|\eta| \lesssim 1.7$. Steel is used as absorber, scintillators and wavelength-shifting fibres as active material [94]. The shower particles excite the plastic of the scintillator tiles which emit photons subsequently detected by photo-multipliers in return.
- The region $1.5 < |\eta| < 3.2$ is covered by the hadronic end-caps (HEC), which use liquid argon for detection and tungsten and copper as absorber material given the higher occupancy in the forward region.

The tile calorimeter is designed [95] to yield an energy resolution of

$$\frac{\sigma}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \, .$$

An additional calorimeter system in the forward region of the detector , the forward calorimeter (FCAL), uses liquid argon and tungsten and copper to measure energy deposits in $3.1 < |\eta| < 4.9$. This part of the calorimeter system is particularly important for measuring forward jets.

4.3.2.4 Muon Spectrometer

The outermost detector layer is formed by the muon spectrometer (MS), which provides measurements of the muons' transverse momenta. Muons with an energy of $\mathcal{O}(100\text{ GeV})$ are minimum ionising particles, described by the Bethe-Bloch-formula[1] [96]. They are hence able to pass the calorimeter and penetrate the MS, which is depicted in Fig. 4.9 and covers a range of $|\eta| < 2.7$. For the detection and measurement of muons the MS comprises different subdetector components, providing efficient tracking and fast triggering:

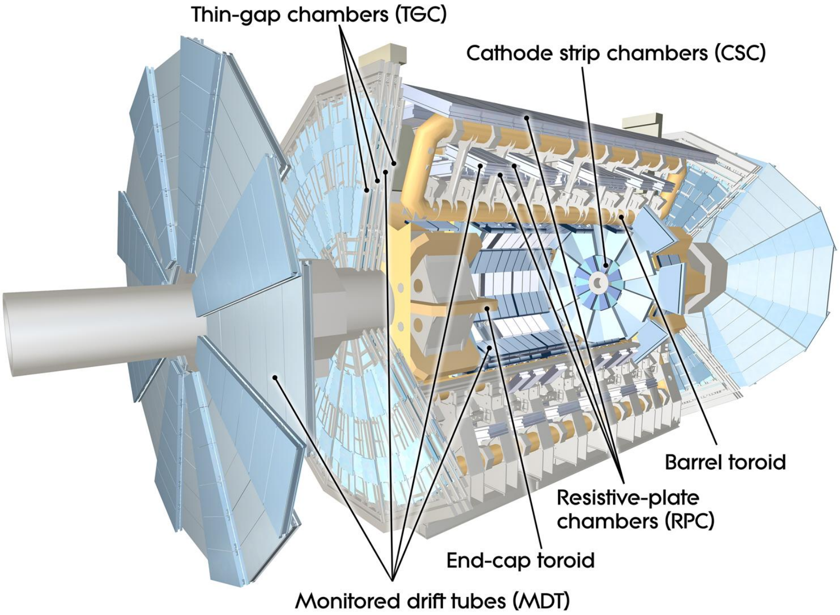


Figure 4.9: A labelled scheme of the muon system. The tracking and trigger components are shown [83].

- The Monitored Drift Tubes (MDT) are filled with gas, which ionises if a muon passes through. Resulting charges are collected at an anode wire, with signals allowing for a precise position measurement along the bending direction of the magnetic field. With modules in the barrel region, the MDTs cover a range of $|\eta| < 2.0$. The modules are arranged such that a measurement of both η and ϕ coordinates is facilitated.

---

[1] Named after H. Bethe and F. Bloch.

- The Cathode Strip Chambers (CSC) provide precise tracking information in the forward regions $2.0 < |\eta| < 2.7$. These multi-wire proportional chambers handle the higher occupancies in this region with a 50 µm resolution.
- In the barrel region $|\eta| < 1.05$ Resistive Plate Chambers (RPC) are used for triggering and position measurements. A traversing muon creates an avalanche between two resistive plates at 2 mm, with a large electric field interspersing the gap. Due to the small time resolution of only 1.5 ns, the RPC provide efficient triggering information.
- Finally, Thin Gap Chambers (TGC) are similar to the CSC but provide a smaller time resolution, capable of providing signals with 25 ns. They are therefore used in the forward region $1.05 < |\eta| < 2.4$ to

The MS extends from 4.25 m out to the full detector radius of 11 m and therefore provides a large enough volume to allow for accurate tracking measurement of muons.

4.3.3 Data Acquisition and Trigger System

The small bunch spacing of the LHC implies a high collision rate of 40 MHz, which imposes technical limitations for the recording of events. In addition, only a fraction of the collision events provide signatures relevant for the ATLAS physics programme, the majority of collisions result in low-p_T multijet events [97]. To cope with these issues, ATLAS deploys a two-level trigger system in RUN II, which reduces the maximum rate of recorded events down to 1 kHz [98, 99]. In order of processing, these trigger sytems are:

- The hardware-based **Level-1** (L1) trigger reduces the event rate to 100 kHz by using trigger information from muon chambers and the calorimeters only. Using FPGA[1]-technology, fast-triggering algorithms search for region-of-interest (RoI) formed by high-p_T particles traversing the calorimeter or MS trigger components and allow for a decision within 2.5 µm.
- For events which pass the L1 trigger, data is sent through a read-out system to the software-based high-level-trigger (HLT) which filters the incoming events to a rate of 1 kHz. The HLT reconstructs objects using precise tracking information, the full calorimeter granularity and employs the topological clustering algorithm for hadronically decaying τ leptons described in Sec. 5.1. Here, object reconstruction steps are harmonised to the techniques used in the full event reconstruction.

Each event passing the L1 trigger is buffered until the HLT is ready for processing and is permanently stored if it passes the latter [97]. The complete trigger chain is presented in Fig. 4.10, depicting in particular the information flow between detector and trigger system components. In ATLAS, so called trigger menus are used to define selections of different trigger signatures depending on the data-taking period. This ensures an optimal usage of the available bandwidth.

---

[1] Field-programmable gate array, an integrated circuit programmed specifically for L1-purposes.

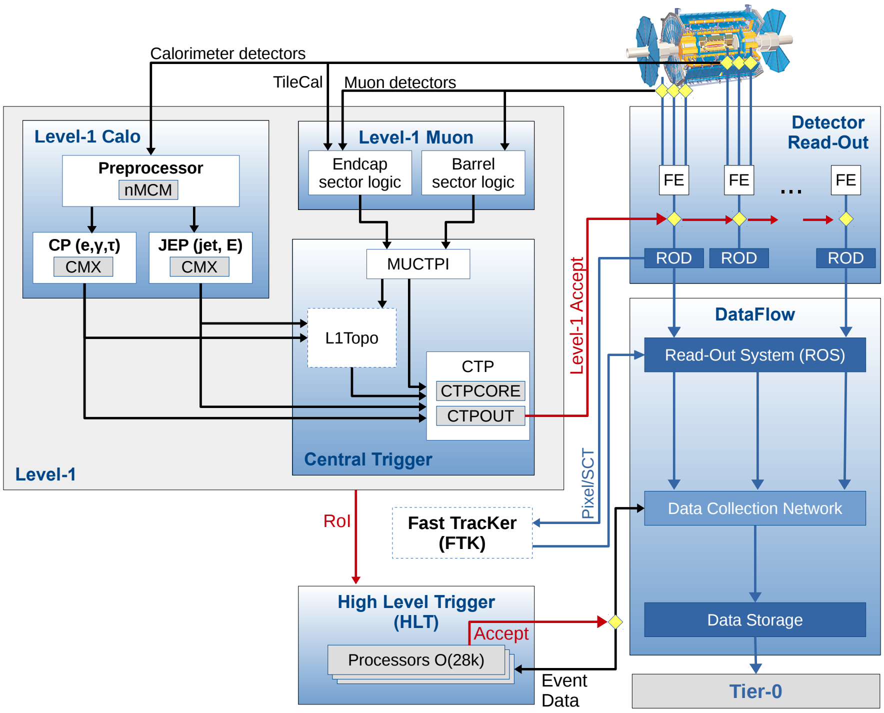


Figure 4.10: Schematic view of the data flow in the trigger system deployed in ATLAS Run II operations. Detector signals from calorimeters and muon spectrometer are analysed in L1 for potential ROIs. If such a ROI is present, the event is processed by the HLT and stored, if the respective trigger conditions are met. In Run II, ATLAS introduced the fast tracker (FTK) and L1topo systems, which significantly improve the trigger efficiency [99].

4.3.4 Data Quality Monitoring

In order to ensure stable operations of the ATLAS detector and well-understood data taking conditions, a complex data quality monitoring (DQM) system is deployed [100]. During operation, online DQM provides ATLAS operation shifters with live information about low-level quantities, such as number of hits for a given detector component. As problematic LHC beam conditions or malfunctioning detector components may be present anytime during data-taking, automated and manual checks are necessary for quick responses. This prevents unrecoverable data corruption. The work on this thesis included the commissioning of the online DQM system for Run II, incorporating various updates to the DQM facilities and ensuring stable operations after various updates during the shutdown phase of the LHC.

The online DQM system comprises three facilities which enable the communication between the detector output streams, the databases and the shift crew:

- the Information Service (IS),
- the Online Histogramming Service (OHS),
- and the DQM framework (DQMF).

The information flow is sketched in Fig. 4.11. The ATLAS monitoring system is a collection of distributed software applications. In this environment, the IS facilitates information

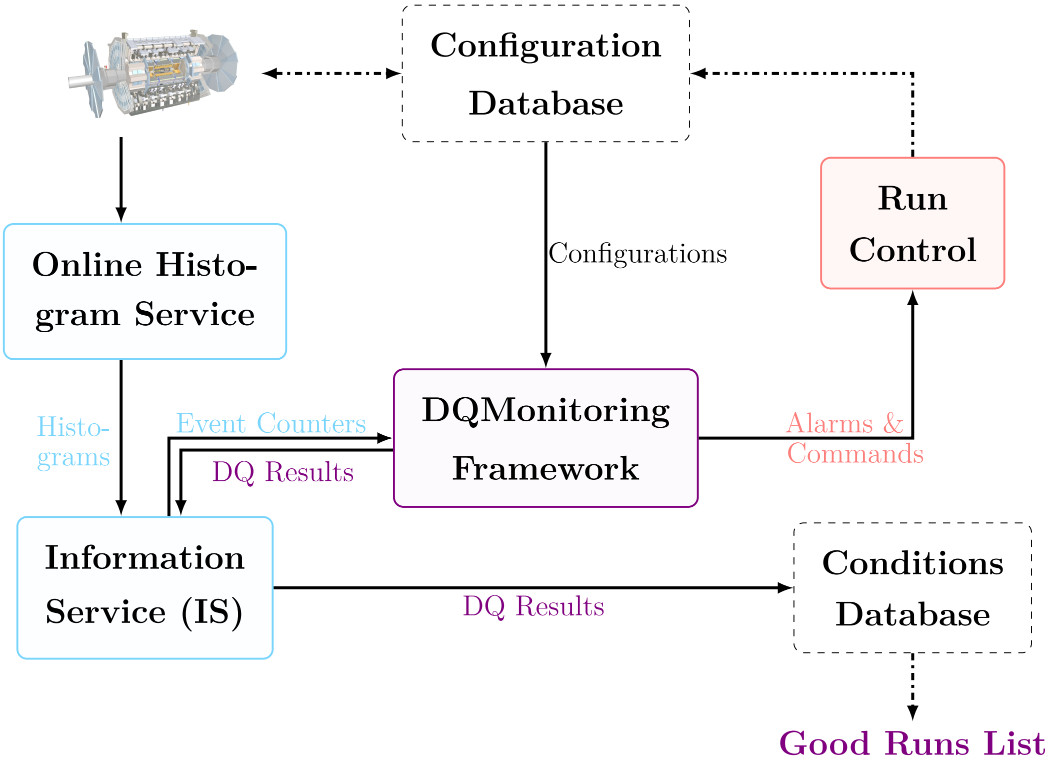


Figure 4.11: Sketch of the information flow in the ATLAS data quality (DQ) monitoring system during operations. The good runs lists is created after additional post-processing data quality evaluation. Figure adapted from [101].

sharing between different applications. An application is allowed to update, add and remove information from the IS repository. Applications can subscribe to the IS if notifications about information changes are needed. On top of that, the IS itself can send commands to the applications to retrieve information. These functionalities enable communication between different DQM components.

The OHS has a similar structure albeit with a specific focus on acting as a transient storage between histogram providers and monitoring display applications. Any histogram published to the OHS server is available to the whole DQM system and can be routed between different applications.

The DQM framework (DQMF) is a software package performing user-defined analysis of monitoring data and summarises the result in a convenient way. Subsequently, the analysed monitoring data are stored in a database, displayed within a graphical interface used by shifters or automatically assessed to perform actions such as alarms or stopping runs.

Each subdetector and physics object is represented as a DQ-region in this framework, which can be further split in sub-regions resulting in a tree-like structure, which facilitates the shifter's work. User-defined algorithms and parameters, DQAlgorithms and DQParameters, are used to assess the quality of the data. The assessment is reflected in DQflags, which intuitively indicate problems in each monitoring component. Such a flag represents the state of the checked DQRegion and may be red, yellow, green and also include specific

information about potential defects.
All flags are collected and summarised by the IS, published to the conditions database (COOL) and via a graphical user interface to the online shifter.

Offline DQM is performed in the reconstruction chain of the recorded data, allowing for an analysis of more complex physics quantities compared to the online DQM and the assessment of the quality of the reconstructed data. The core part of the DQMF is designed such that it also can be used by the offline DQM, which can retrieve the recorded information from COOL. Problems during data acquisition or in the detector hardware are identified and recorded by means of automated and manual monitoring of a small set of key distributions which contain all relevant information from each detector system.
All information gathered from online and offline DQM are combined into a common database. This database also contains LHC beam conditions and detector status information and is subsequently used to create lists of luminosity blocks suitable for analysis, the good-runs-list (GRL).

4.3.5 Analysis Data Flow in Run II and the xTauFramework

The analysis of recorded data at CERN is generally based on the ROOT C++ software framework, which provides methods for data processing, statistical analysis, visualisation and storage [102]. ATLAS deploys a specific event data model (EDM) to facilitate a common structure for class collections, interfaces and concrete physics objects for the variety of physics analysis groups. The initial EDM used in RUN I [103] revealed shortcomings and limitations over the course of time [104], as for example

- complicated and CPU-expensive C++ features like virtual inheritance
- excessive duplication and reformating of data to allow for processing with ROOT
- difficulties to maintain analysis tools due to the variety of different formats.

During the shutdown phase after RUN I, a large-scale software campaign was started to introduce a new harmonised format - the xAOD[1] - to overcome these limitations [104]. The main feature of this updated EDM is the separation of object properties and the analysis object classes themselves, using a storage container called auxiliary store and a separate object interface. This reorganisation enables a user to partially read objects, to remove selected object data and to add user-defined object properties in a uniform manner via decorations. Furthermore, the xAOD can be directly processed with ROOT without the need for large external libraries, improving the usability of software for performing physics analyses significantly.
This software campaign entailed the complete rewriting of reconstruction and data processing software to cope with the xAOD format. Dedicated studies [104] showed, that the new EDM improved the processing time by 25%, adding to the overall factor 3 speed-up observed for the full updated RUN II data flow. The higher occupancy due to the increased event rate and pile-up can be effectively handled with this updated approach.

The data format used for physics analyses in this thesis is based on simple ROOT files

---

[1] Analysis Object Data, with x denoting the RUN II update.

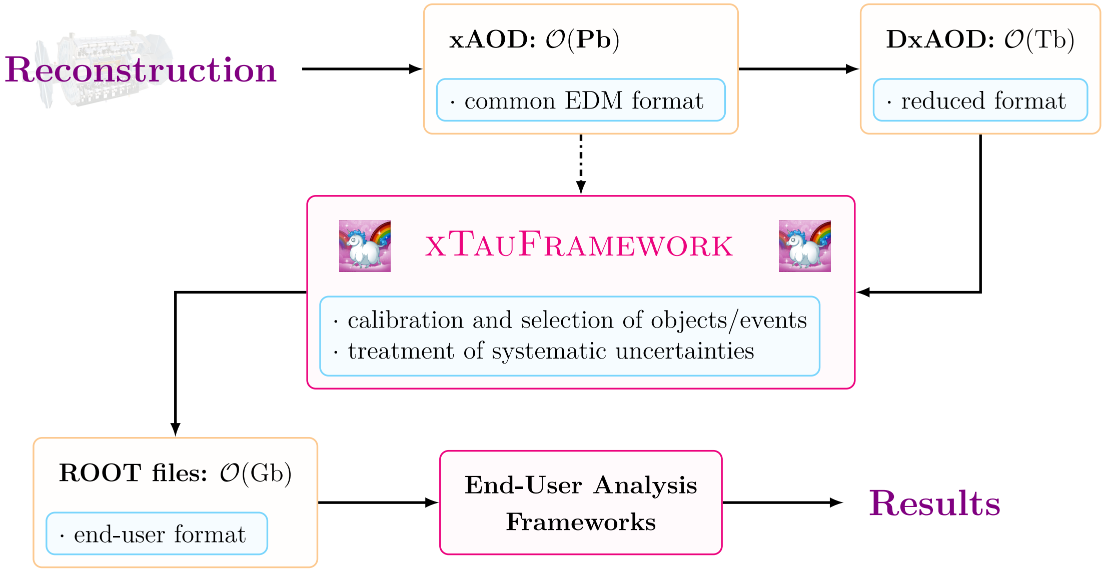


Figure 4.12: Flowchart of the analysis model with the Atlas Run II EDM. Object calibrations and selections are applied with the xTauFramework. The indicated storage specification in orders of Bytes reflect the typical disk usage per processing step for an analysis in Atlas. Unicorn logo taken from [105].

sizing up to several 100 giga-bytes, with a tree-like data structure. In order to prepare the peta-bytes of recorded and simulated data in xAOD format for physics analyses, a complex chain of steps is necessary, as summarised in Fig. 4.12. A first size decrease is achieved by processing the xAOD with a reduction framework, resulting in derived xAODs. This reduction takes advantage of the new EDM by manipulating the xAOD to only contain analysis relevant events, objects and object properties, procedures called skimming, slimming and thinning, respectively. Furthermore, analysis specific object and event properties can be added by decorations. Given the initial size of the xAOD, these derivations are produced with central Grid infrastructures [106].

In the next step of the analysis preparation, events and simulated physics objects, like τ leptons, are calibrated and corrected to auxillary measurements, as explained in Chapter (5). These corrections require external input in form of data files and object interfaces provided by the combined performance (CP) working groups in Atlas, dedicated to reconstruct, identify and correct specific physics objects.
A big project successfully tackled during the course of this thesis was the development and management of a C++-based framework, common across τ-related analyses in Atlas in Run II, capable of applying these steps. This intuitively named xTauFramework is able to process the (derived) xAOD formats and apply calibrations and corrections according to the CP group recommendations. The xTauFramework processes single events, applies analysis specific selections and provides the possibility to process and store the sizable amount of systematic variations. By taking particular care of modularity, steerability, user-friendliness and documentation the design was aimed to incorporate any interested analysis related to τ leptons in Atlas. At peak times, over 25 developers participated actively in

incorporating more than 16 different analyses - a still growing number. After developments for the initial commissioning phase in early RUN II, the management and organisation of this framework became an important task, ensuring streamlined developments and synchronised efforts with the CP groups. Technically, this harmonised approach imposed the usage of various C++ specific features, such as inheritance, polymorphisms and templated functions to ensure full user-defined configurability. The xTAUFRAMEWORK reduces the data size to $\mathcal{O}$(GB) by storing ROOT files with a tree-like data structure and is able to process data in the GRID infrastructure, on batch computing clusters and on local machines. The resulting output can subsequently be processed by ROOT-based analysis-frameworks as indicated in Fig. 4.12, bridging the gap between reconstruction data and end-user analysis.

CHAPTER 5

Object and Event Reconstruction

The physics objects used in the analysis are reconstructed from basic detector signals, which are induced by the object's - or its decay product's - interaction with the detector material. Starting from basic objects described in Sec. 5.1, a plethora of different algorithms is used to differentiate and measure the particles from Fig. 2.1 as detector signatures.
The physics objects relevant to the presented analysis are hadronically decaying τ leptons τ_{had}, jets j and to a lesser extent electrons and muons. Furthermore, the expected presence of neutrinos, which escape the detector without interacting with its material, gives rise to missing transverse energy $E_{\text{T}}^{\text{miss}}$.

Over the course of this thesis, measurements of the τ_{had} identification efficiency were performed which are outlined in greater detail in Sec. 5.4.2.
The final selection criteria to define the physics objects in the analysis are listed in Sec. 6.3.

5.1 Detector Objects

The reconstruction of physics objects in ATLAS is based on two basic detector objects - tracks of charged particles interacting with the inner detector material or the muon spectrometer and energy deposits in calorimeter cells, forming clusters. Technical details of the detector components are summarised in Sec. 4.3.

Tracks The reconstruction of inner detector (ID) tracks is based on three-dimensional signal points in the ID components induced by charged particles traversing the detector layers [107]. An ID track is seeded by clusters of three or more close-by signal points in the IBL and Pixel-layers. These tracks are extrapolated to the strip layers of the inner detector using a fit. This fit removes falsely associated clusters and imposes quality criteria on the track, such as the number of hits in each layer and the goodness of the fit. The so-formed track candidates are matched to combined signals in the TRT, called drift circles, using an additional fit. Non-associated remnants of TRT signals are furthermore combined with track fragments in the Pixel layers to improve the reconstruction efficiency in particular for long-lived particles and tracks from photon conversion.

Several vertex finding algorithms are subsequently applied to identify the primary collision vertex and secondary decay vertices.
The reconstruction of muon spectrometer tracks are based on signals from the monitored drift tubes as introduced in Sec. 4.3.2.4. They can be reconstructed in a stand-alone manner or matched to an inner detector track.

Calorimeter Clusters The ATLAS calorimeter system is designed to provide finely segmented three-dimensional shower information. Energy deposits in individual calorimeter cells are grouped into topological clusters [108]. Starting from an initial cell with an energy deposit exceeding a $4\,\sigma$ noise threshold. Any neighbouring cells are added to the initial cell if they exceed a $2\,\sigma$ threshold, forming a cluster. The clustering stops, if no further adjacent cells fulfilling the threshold criteria are found. In a final step, all directly neighbouring cells with a $0\,\sigma$ noise threshold are added to the so created cluster. The energy deposits of each neighbouring cell in a cluster are compared to determine the local maximum. Clusters are split, if multiple maxima are found until each cluster has one local maximum deposit only. An energy calibration, called local hadron calibration [109], of the clusters is necessary due to non-compensated losses in the detector material. All clusters are initially calibrated to the electromagnetic scale and subsequently identified to originate either from hadronic or electromagnetic showers. This identification is based on shower-shape variables. Depending on the cluster type, different calibrations are applied.

5.2 Electrons and Muons

Electrons and muons are used in the analysis described in this thesis to ensure orthogonality between the $H \to \tau\tau$ analysis channels and to remove potential overlaps between physics objects.

Muons To reconstruct muon candidates [110], a matching between inner detector and muon spectrometer tracks is performed within $|\eta| < 2.5$. The quality of the track-matching is quantified with a χ^2-fit [110] based on the complete track information in both detector systems. The muon energy is corrected for losses in the detector material. For muons with low (high) p_{T}, the momentum measured in the inner detector (muon spectrometer) is used, respectively. All considered muons must fulfil $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. The identification of muon candidates is based on the quality of the associated tracks. The loose working point was chosen for this analysis.

Electrons Electron reconstruction in ATLAS is based on charged inner detector tracks which are matched to electromagnetic calorimeter clusters [111, 112]. By investigating calorimeter-based shower-shape information, track quality criteria such as the number of hits in the TRT and track-cluster association, a likelihood score is constructed, quantifying the identification level of an electron candidate. All electrons considered in the analysis must fulfil the loose identification requirement [111], exceed $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and are only considered if they are within $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$. The latter requirement excludes the transition region between barrel and end-cap calorimeter. The direction of the electron candidate is determined either from the track direction or the associated cluster, depending on the tracks' quality.

5.3 Jets

Hadronic clusters originate from deposits of parton showers induced by quarks/gluons produced in hard interaction or from radiation. By using the anti-k_t algorithm [113] with a distance parameter $R = 0.4$, these clusters are combined to jets, which are used as physics objects in the analysis. The anti-k_t algorithm uses a cluster distance definition, which allows to separate soft radiation constituents from the hard jet. This algorithm is thereby robust against soft radiation, also called infrared-safety, which facilitates jet calibration and theoretical calculations. The anti-k_t algorithm furthermore provides safety against collinear splitting of cluster constituents [114].

An improvement of the jet reconstruction efficiency can be achieved by associating tracks to the jet candidate and identifying the jet vertex. Further details are given in Sec. 6.3.

The jets are calibrated using the electromagnetic+JES scaling scheme, which is based on simulation of the true calorimeter response to hadronic jets and in-situ measurements [115]. Furthermore, it is possible to distinguish b-quark initiated jets due to the longevity of the b by using dedicated b-tagging algorithms [116]. However, for the analysis described in this thesis, b-jet identification is not necessary and therefore not used.

5.4 τ_{had} Leptons

In the following sections an overview of τ lepton decays and reconstruction techniques used in ATLAS is presented.

5.4.1 τ Lepton Decays

With $(1\,776.86 \pm 0.12)\,\text{MeV}$, the τ particle is the heaviest of the known leptons introduced in Chapter (2). It has a mean lifetime of $(2.903 \pm 0.005) \times 10^{-13}\,\text{s}$, implying a proper decay length of $87.03\,\mu\text{m}$.

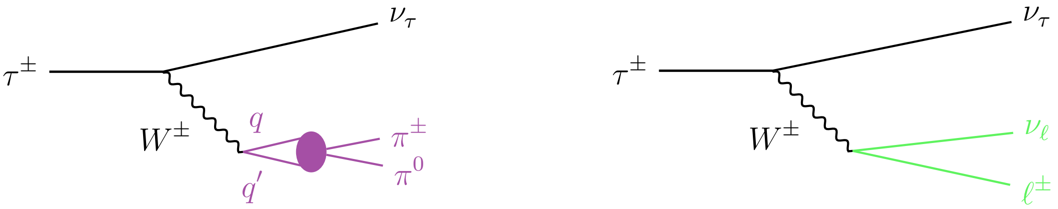


Figure 5.1: Sketches of τ lepton decays, hadronically on the left, leptonically on the right.

In 1975, its discovery was claimed after observing an anomaly in e^+e^--annihilation at SLAC [117]. The τ lepton has the unique characteristic amongst the leptonic particles in the SM to decay both, leptonically and hadronically due to its comparably large mass. The possible decays are sketched in Fig. 5.1.

Its lifetime implies, that the τ lepton typically decays inside the beampipe, allowing only for an indirect observation through its decay products. In Tab. 5.1, the dominant decay modes and branching ratios (BR) are listed.

Type	Decay Mode $\tau^\pm$	BR [%]
leptonic	$\rightarrow \ell^\pm \nu_\ell \nu_\tau$	35.21 ± 0.06
	$\rightarrow e^\pm \nu_e \nu_\tau$	17.82 ± 0.04
	$\rightarrow \mu \nu_\mu \nu_\tau$	17.39 ± 0.04
hadronic	$\rightarrow j \nu_\tau$	~ 64.79
1-prong	$\rightarrow \mathrm{had}^\pm \nu_\tau + X$	50.03 ± 0.08
	$\rightarrow \pi^\pm \nu_\tau$	10.82 ± 0.05
	$\rightarrow \pi^\pm \pi^0 \nu_\tau$	25.49 ± 0.09
3-prong	$\rightarrow 3\mathrm{had}^\pm \nu_\tau + X$	15.21 ± 0.06
	$\rightarrow 2\pi^\pm \pi^\mp \nu_\tau$	9.31 ± 0.05
	$\rightarrow 2\pi^\pm \pi^\mp \pi^0 \nu_\tau$	4.62 ± 0.05

Table 5.1: A selection of the dominant decay modes of the τ lepton [27]. Jets originating from τ-decays are labelled as j and their hadronic constituents as had.

Most noticably, the majority of τ leptons decay hadronically with a BR of 65%. The signature of this decay mode can be understood from Fig. 5.1. A single ν_τ accompanies the hadronic decay product, dominantly formed by one (three) charged pions with a BR of 72% (22%). Furthermore, in 68% of all hadronic τ decays a neutral pion is expected. These hadrons are detectable, hence referred to as visible decay products and labelled τ_{had} throughout the rest of this thesis. The 35% of leptonic decays result in an electron or muon in association with two neutrinos. These leptons are difficult to differentiate experimentally from prompt ones, i.e. leptons produced in the primary vertex.

As detailed in Sec. 6.1, three different, parallel analyses are designed in context of the SM $H \rightarrow \tau\tau$ measurement subject to this thesis. Analyses involving leptons have the advantage of exploiting the electron and muon triggers, which provide a higher efficiency than the τ_{had} ones. However, the decrease in mass resolution due to the two neutrinos in the event and the background contributions from the multitude of processes involving leptons pose additional challenges compared to a fully hadronic $H \rightarrow \tau\tau$ final state analysis.
The hadronic decay mode offers the possibility to directly identify τ leptons via a complex reconstruction and identification chain, detailed in the following sections. Hadronic constituents in a jet formed from τ_{had} typically consist of a small number of charged and neutral pions, which are collimated due to the boost received in the τ lepton decay. Jets

from gluons and quarks created in soft interactions on the other hand are formed from a higher number of hadronic constituents with larger spatial separation.
Typical signatures in ATLAS for such a decay involve one or three charged tracks detected in the ID and a distinctive, collimated cone of energy deposits in the ECAL and HCAL. The associated neutral pions deposit their energy in the ECAL through decays into two photons. On top of this, the tracking resolution in ATLAS allows for a measurement of the τ_{had} vertex displacement with respect to the primary vertex. These features are exploited to distinguish τ_{had} from jets and electrons, the two main background sources.

In the following sections, details about the reconstruction, identification and calibration techniques for hadronically decaying τ leptons τ_{had} is presented. On top of a discussion of the deployed algorithms, an overview of measurements in the 2015 dataset with $3.2\,\text{fb}^{-1}$ is given. Emphasis is laid on the tag-and-probe efficiency measurement of the τ_{had} identification against electrons, which was performed in context of this thesis.

This section is based on [11, 12], where detailed information about the simulated samples used in the various studies can be found.

5.4.2 Reconstruction of τ_{had} Candidates

The τ_{had} candidate reconstruction involves three steps: the seeding from a jet candidate, the reconstruction of the τ-decay vertex and the association of tracks.

τ_{had} Candidate Seed The input to the τ_{had} candidate reconstruction algorithm is formed by jets reconstructed by the procedure explained in Sec. 5.3. These jets must satisfy $p_{\text{T}} > 10\,\text{GeV}$ and be within $|\eta| < 2.5$, corresponding to the coverage of the ATLAS tracking system. Jets in the transition region between barrel and endcap calorimeter $1.37 < |\eta| < 1.52$ are excluded. The initial p_{T} of the τ_{had} candidate is set to the total energy of the topological cluster associated to the jet in the local hadronic calibration. This serves as starting value to the identification and calibration procedure described below. The coordinates of the τ_{had} candidate are defined by the seed jet's η- and ϕ-coordinates. The τ_{had} candidate mass is artificially set to zero, hence transverse momentum and transverse energy of the reconstructed object are identical. These values define the complete τ_{had} candidate four-momentum before further calibrations and corrections are applied.

Vertex Association The decay vertex of the τ lepton is displaced with respect to the primary production vertex due to the τ decay length. Correctly identifying the τ decay vertex reduces contributions from pile-up and increases the reconstruction efficiency. A dedicated algorithm is used for this task, selecting the vertex from the reconstructed primary vertex candidates of the event [118].
The τ_{had} vertex association algorithm considers all tracks within a region $\Delta R < 0.2$ around the jet seed's direction. The transverse momenta of these tracks are summed and matched to each primary vertex. The vertex to which the highest fraction of the summed p_{T} is matched, is chosen as τ_{had} vertex. Impact parameters z_0 and d_0 are subsequently recalculated, which improves the efficiency of track association. The vertex is used together

with the seed jet's topological clusters in $\Delta R < 0.2$ to define the τ_{had} candidate direction before calibration.
In Fig. 5.2 (left), the efficiency for correctly assigning the vertex is shown for τ_{had} candidates with one associated track and compared to the initially associated vertex. An efficiency of $> 95\%$ is achieved for τ_{had} candidates with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, exceeding the initial vertex association efficiency in particular for low p_{T}.

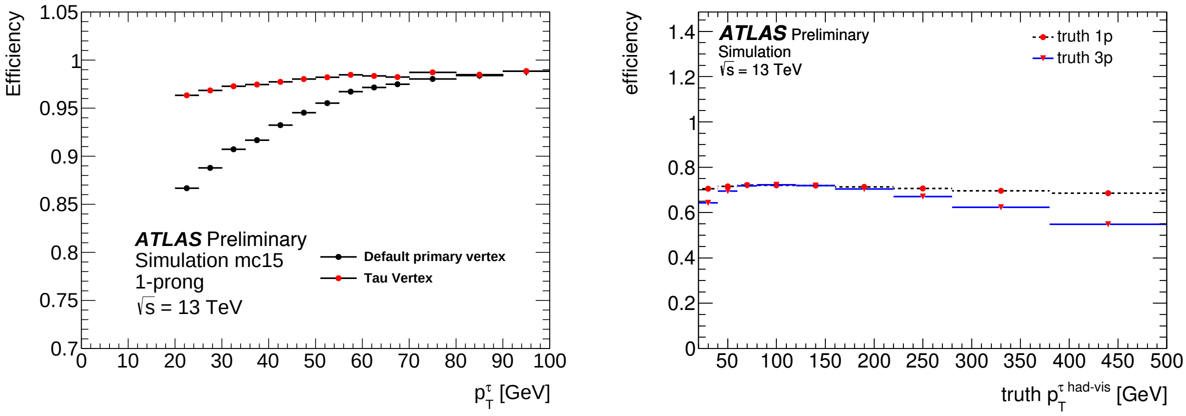


Figure 5.2: Efficiency for τ_{had} vertex assignment (left) and track reconstruction (right) determined in simulation [11].

Track Association With the above definition of the τ_{had} candidate direction, two cone regions around the τ_{had} candidate are defined - the core cone with $\Delta R < 0.2$ and the isolation annulus $0.2 < \Delta R < 0.4$. Any track with sufficient quality, $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ and close to the τ_{had} vertex, quantified by the impact parameters, is associated to the τ_{had} candidate if it lies within the isolation annulus region.
A characteristic feature of τ_{had} decays is the collimated jet shape arising from the boost of the τ_{had} decay products. This implies that charged and neutral pions expected in τ_{had} decays, are close to each other. Therefore, the tracks within the core cone define the number of tracks of the τ_{had} candidate. The charge of the τ_{had} candidates is determined by the sum of charges of the tracks in the core cone.
Following the explanations in Sec. 5.4.1, the majority of τ_{had} candidates are expected to have a number of tracks of 1 (3) corresponding to one (three) associated charged pions. Due to tracking inefficiencies it is however possible, that not all τ_{had} tracks are properly reconstructed or false tracks are associated. In particular, a τ_{had} candidate can have 2 associated core tracks, leading to a charge of 0 or 2. This fact is exploited in the construction of the multijet background estimate as presented in Sec. 7.3.2.
Fig. 5.2 (right) shows the efficiency of reconstructing the same number of tracks as the number of simulated charged decay products for a τ_{had} candidate in dependence on the true p_{T}.

Efficiency and Uncertainties The efficiency of the τ_{had} candidate reconstruction is mainly affected by vertex selection and track association performances. It depends on

the number of tracks and p_{T} regime of the τ_{had} candidates. Inefficiencies originate from overlapping track trajectories for very collimated τ_{had} candidates at high p_{T} or from the track requirements for low p_{T}.
The efficiency has been studied in simulated events by comparing the simulated and reconstructed number of tracks.
Systematic uncertainties on the reconstruction efficiency arise from pile-up and inner detector material effects and are estimated in simulation [11]. They range between 2% to 4.5% depending on the p_{T} of the τ_{had} candidate.

5.4.3 Jet Background Rejection

The τ_{had} candidate reconstruction steps reject only a very small fraction of quark- or gluon-initiated background jets (q/g-jets). To identify τ_{had}, dedicated Boosted Decision Trees (BDTs) were trained separately for τ_{had} candidates with 1 and 3 core tracks, respectively. A thorough explanation of the BDT algorithm is given in Sec. 8.2.1.
The BDTs exploit calorimetric shower shape and tracking information based on the associated tracks and topological clusters of the τ_{had} candidate. The following discriminating characteristics of τ_{had} are used to distinguish them from background jets:

- the collimated structure of τ_{had} stemming from the boost of the τ decay products leads to a narrower signature compared to q/g-jets, reflected in closer energy deposits in the calorimeter
- related to this, the cone containing all associated tracks is smaller for a τ_{had}
- the momentum fraction carried by the track with largest p_{T} is higher for τ_{had} than for q/g-jets, as the total energy for the latter is expected to be more evenly distributed
- the impact parameters of the associated tracks are larger due to the τ decay length
- the electromagnetic calorimeter is sensitive to the π^0 meson content of a jet, which can be used as identification information, as the τ_{had} is typically associated to zero or one π^0 decays and one or three charged pions.

The BDT output is translated into three identification points loose, medium and tight which offer increasing background rejection and targeted signal efficiencies of $0.6, 0.55$ and 0.45, respectively. The efficiencies are constant with respect to the τ_{had} p_{T} and stable against different pile-up conditions.

The performance of the identification procedure was evaluated in a $Z/\gamma^* \to \tau_\mu \tau_{\mathrm{had}}$ tag-and-probe measurement [12], where one τ decays into a muon (the tag-lepton) and associated neutrinos and the other τ hadronically (the probe). Small differences between simulated and measured efficiencies are accounted for by correction factors, which are later applied in the analysis. These depend on the number of associated tracks and are depicted in Fig. 5.3 for the three identification working points. The uncertainties on these correction factors are around $\sim 5\%$.

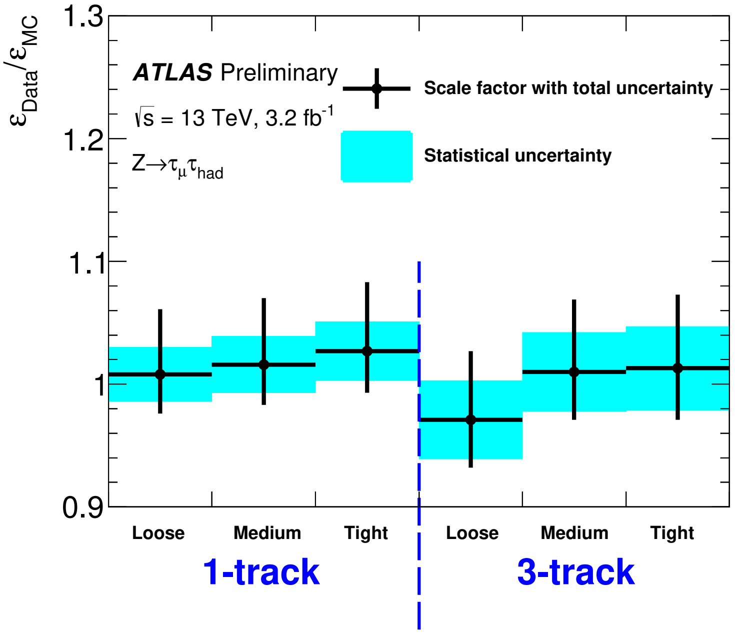


Figure 5.3: A comparison of the measured and simulated τ_{had} identification efficiency expressed as correction factors [12].

5.4.4 Electron Rejection and Efficiency Measurement

The second major background component for the τ_{had} identification is arising from falsely identified electrons, which mimic the signature especially for τ_{had} with 1 core track. An additional identification requirement is constructed based on a likelihood discriminator built for the electron identification in ATLAS [111, 119]. This likelihood is based on shower shape information in the electromagnetic calorimeter and hit information in the inner detector and offers a powerful discrimination between charged π and electrons. As detailed in the following, possible overlaps between τ_{had} candidates and electrons are investigated. The τ_{had} candidate is rejected in dependence on the likelihood score of the matched electron. The performance of this electron rejection procedure, refered to as e-Veto, was measured in the course of this thesis.

5.4.4.1 Electron Likelihood-based Rejection

The construction of the e-Veto is based on simulated $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \tau\tau$ events. A geometrical matching in $\Delta R < 0.4$ is performed between electrons with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ and τ_{had} candidates, to account for losses of electrons in the τ_{had} candidate reconstruction steps. The likelihood score of the matched electrons is then assigned to the τ_{had} candidate. Furthermore, electrons considered for tuning the e-Veto fulfil loose quality requirement applied during their reconstruction, which are based on calorimetric shower shape information like the hadronic leakage of energy deposits. This removes a fraction of the simulated, reconstructed electrons. Therefore, not all reconstructed τ_{had} candidates can be matched, in which case an artificial likelihood score of -4 is assigned to the τ_{had} candidate, reflecting

the τ_{had} like characteristics.
The electron rejection is then performed by applying an upper requirement on the assigned likelihood score and a τ_{had} candidate is rejected if the matched electron has a large score. In Fig. 5.4 (left), the difference between likelihood scores for true electrons and misidentified τ_{had} is presented. The electron rejection is tuned as a function of the τ_{had} p_{T} and track-η such that it yields a constant efficiency of 95% on a reference sample of simulated τ_{had} passing the loose jet identification requirement. Fig. 5.4 (right) shows the background rejection in dependence of the signal efficiency for different η ranges, corresponding to specific detector regions.
The electron likelihood is stable against varying pile-up conditions [119], implying stability also for the e-Veto.

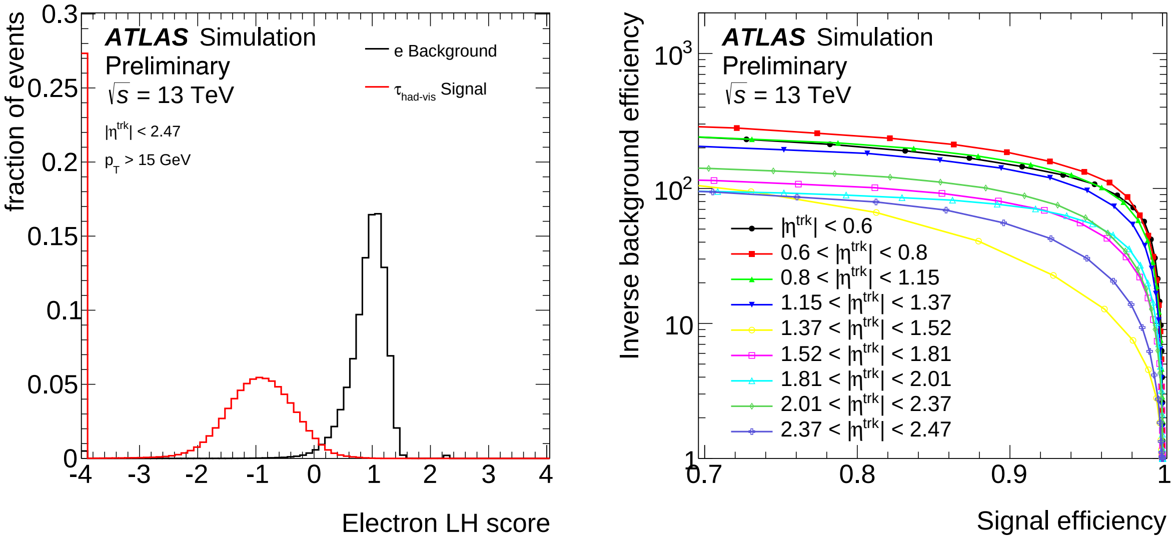


Figure 5.4: Electron likelihood score for prompt electrons and τ_{had} candidates (left) and efficiency of the e-Veto depending on different η ranges.

Various systematic uncertainties are considered for estimating the signal efficiency of the e-Veto, in particular:

- the uncertainty on the modelling of the inner detector material, affecting for example the track reconstruction efficiency
- by varying the noise threshold for the reconstruction of topological clusters, the calorimeter response uncertainty is estimated
- the underlying event tunes used in the simulated samples and the modelling of pile-up
- the hadronic shower model is varied in the respective simulated samples to evaluate the uncertainty on the modelling of interactions with the detector.

Depending on the p_{T} and η of the τ_{had} candidate the total systematic uncertainties range from 2% to 5%.

5.4.4.2 Efficiency Measurement

The performance of the above rejection method was measured by estimating the misidentification probability of electrons as τ_{had} candidates in $Z/\gamma^* \rightarrow ee$ dominated events. A tag-and-probe analysis approach was chosen, tagging a well-defined electron and probing whether the second electron was identified as τ_{had} candidate before and after applying the e-Veto.

Analysed Dataset This measurement is based the full 2015 dataset with $3.2\,\text{fb}^{-1}$ pp-collisions at $\sqrt{s} = 13\,\text{TeV}$. Only events recorded with a fully operational detector state were considered. Different single-electron trigger criteria were used to maximise the acceptance of the analysed dataset. These depend on the p_{T} of the medium quality electron trigger-object [120]. For each trigger, a geometrical matching within $\Delta R < 0.2$ was required between the trigger-object and the reconstructed electron. Inefficient trigger regions were excluded by requiring a minimum p_T threshold for the reconstructed electron depending on the trigger-object $p_T^{e-\text{HLT}}$ as follows:

- for $p_T^{e-\text{HLT}} > 24\,\text{GeV}$, consider reconstructed electrons with $25\,\text{GeV} < p_T < 65\,\text{GeV}$
- for $p_T^{e-\text{HLT}} > 60\,\text{GeV}$, require $65\,\text{GeV} < p_T < 135\,\text{GeV}$
- and finally if $p_T^{e-\text{HLT}} > 120\,\text{GeV}$, then $p_T > 135\,\text{GeV}$.

Event Selection Only events with at least one isolated electron with $p_{\text{T}} > 25\,\text{GeV}$ and tight identification quality were considered [111]. The event must also contain a reconstructed τ_{had} candidate passing the medium jet-identification with $p_{\text{T}} > 20\,\text{GeV}$. This τ_{had} candidate defines the probe lepton if it overlaps with a reconstructed electron in $\Delta R < 0.2$.
Following this, the $Z/\gamma^* \rightarrow ee$ enriched signal region was constructed by applying the following criteria on the selected events:

- the tag electron must exceed $p_{\text{T}} > 30\,\text{GeV}$
- the probe τ_{had} candidate is required to have exactly one core track and be oppositely charged to the electron
- both, tag and probe lepton are required to be within $|\eta| < 2.47$ with the transition region between barrel and end-cap calorimeter $1.37 < |\eta| < 1.52$ excluded
- the invariant mass of the tag and probe leptons is restricted to the Z boson mass peak region $80\,\text{GeV} < m_{e\tau} < 100\,\text{GeV}$
- the transverse mass of electron and the missing transverse energy of the event is less than $m_T < 40\,\text{GeV}$ in order to reject contributions from W boson processes

With this selection a $Z/\gamma^* \rightarrow ee$ purity of 98% was achieved.

Background Estimation The estimation of expected background contributions from $Z/\gamma^* \rightarrow ee$, $W+\text{jets}$, single top quark and $t\bar{t}$ processes is based on simulation using generators and tunes detailed in [12]. The normalisation of the $W+\text{jets}$ background was estimated in a control region defined by requiring $m_{e\tau} < 80\,\text{GeV}$, $m_T > 70\,\text{GeV}$ and a lower threshold on the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}} > 40\,\text{GeV}$. In this $W+$jets dominated region, the modelling was compared to the selected data events and normalisation factors were derived.

Contributions from multijet events are modelled in a fully data-driven way. Shape templates for the p_{T} and η of the τ_{had} candidate were constructed in data by requiring both, tag and probe lepton to have the same charge and subtracting the simulated contribution in this selection. The normalisation of these templates was derived in a multijet control region defined by inverting the isolation criterion of the tag-electron. This enriches the selected sample with jets mimicking electrons and thereby enhances the multijet purity of the selection while reducing other background processes.

In Fig. 5.5, the p_{T} and pseudorapidity η of the τ_{had} probe lepton is shown in the $Z/\gamma^* \to ee$ signal region selection using the described background estimation. Overall, a good modelling was observed as reflected in Fig. 5.5. The predicted η distribution of the τ_{had} probe lepton shows an overestimation for $|\eta| > 2.3$ which originates from a mismodelling of an electromagnetic calorimeter shower variable in the τ_{had} reconstruction. This mismodelling is implicitly corrected with the correction procedure described in the following.

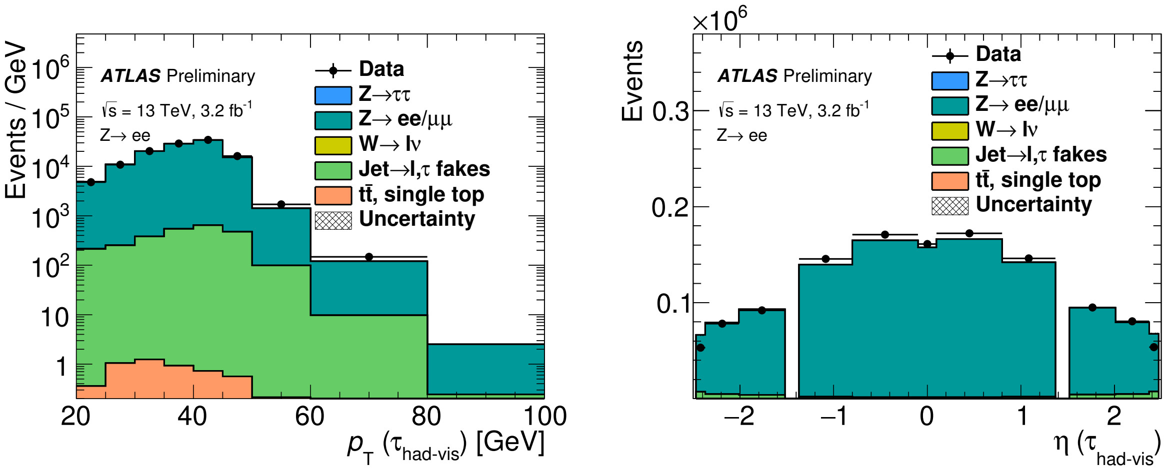


Figure 5.5: Modelling of p_{T} and η distributions used in the e-Veto efficiency measurement before application of the veto requirement.

5.4.4.3 Results

The electron misidentification probability was measured by estimating the number of $Z/\gamma^* \to ee$ events in which the probe-leptons pass both, the medium τ_{had} jet-identification requirement and the e-Veto. Comparing this value to all selected events gives a measure for the rate of electrons not rejected by the τ_{had} identification.

All background contributions not originating from $Z/\gamma^* \to ee$ processes were subtracted from the selected data and then compared to the simulated $Z/\gamma^* \to ee$ sample. The misidentification probability was estimated in dependence of η in both set of events and range between $0.5 - 2.5\%$. Differences between background prediction and measured data efficiency are translated to scale factors, as depicted in Fig. 5.6.

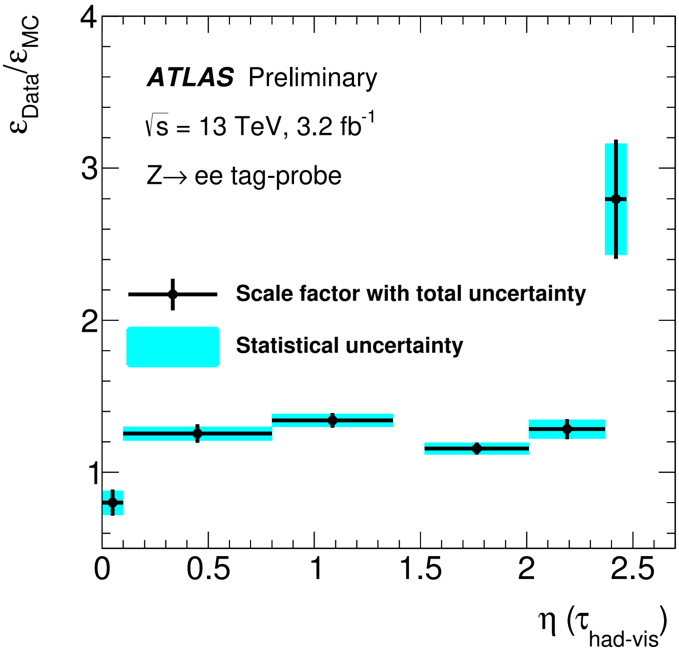


Figure 5.6: The e-Veto efficiencies differ between data and simulation. The correction factors correct the simulated to the observed e-Veto efficiency.

The correction factors range between $0.8 - 1.2$ for detector regions $|\eta| < 2.3$. For $|\eta| > 2.3$ a large correction factor is measured, corresponding to the aforementioned mismodelling in simulation. The factors are stable across the full p_{T} range of the τ_{had} candidate. The correction factor measurement considers the following systematic uncertainties:

- electron trigger uncertainty
- reconstruction, identification and calibration of the electrons
- a statistical and systematical component of the multijet and $W+$jets normalisations

Their impact is summarised in Tab. 5.2. Further details on the estimation and sources of these uncertainties can be found in [12].

Source	Uncertainty
Statistical Uncertainty	4.9%
Multi-jet normalisation	< 1 %
$W+$jets normalisation	< 1 %
Electron Uncertainties	< 1 %

Table 5.2: Summary of the uncertainties on the electron rejection efficiency measurement.

5.4.5 τ_{had} Lepton Trigger

The trigger used in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel is constructed from two hadronically decaying τ trigger-objects τ_{HLT}, referred to as di-hadronic τ lepton trigger. Reconstruction, identification and calibration methods for these τ_{HLT} follow the procedures summarised in the previous and next section, albeit being optimised to cope with the limited processing time available at the trigger stage of the event reconstruction.
τ_{HLT} are formed in a three-staged approach, each stage only being processed if the preceeding stage was passed to cope with the technical limitations of the ATLAS trigger system, which was introduced in Sec. 4.3.

L1 Reconstruction The first stage of the reconstruction is based on Level-1 (L1) calorimeter information only. So called trigger-towers are formed in the ECAL and HCAL within $|\eta| < 2.5$ from calorimeter cells with a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. A 2×2 square of such trigger-towers is combined and defines the core region of a τ_{HLT}-candidate, the region-of-interest (ROI). The initial transverse energy of the candidates is defined as the sum of the two highest energetic neighbouring towers, calibrated at the electromagnetic scale. This results in a worse energy resolution compared to the full τ_{had} reconstruction chain, due to the simple cell clustering using towers and the non-τ_{had} specific calibration. Additional reduction for the event rate can be achieved by combining two such candidates. Topological requirements $\Delta R_{\tau\tau} < 2.9$ and $\Delta R_{\tau\tau} > 0.1$ are applied for the di-hadronic trigger, targeting the relevant event topology and suppressing backgrounds from jets.

HLT Reconstruction The high-level-trigger system provides the possibility to exploit the full detector granularity and run more sophisticated algorithms, run in processing time ordered succession.
In the second stage of τ_{HLT} reconstruction, calorimeter- and track-preselections are applied. ROIs identified in L1 are used as input to the topological clustering algorithm described in Sec. 5.1, with the local hadronic calibration. Using an energy calibration of the associated clusters similar to the one explained in Sec. 5.4.6, the resulting τ_{HLT} energy resolution is significantly improved. At this stage, a minimum p_{T} threshold of 35 GeV and 25 GeV is applied for the two τ_{HLT} used in this analysis.
Thereafter, a trigger-specific pattern recognition algorithm identifies tracks with $p_{\mathrm{T}} > 1$ GeV associated to the τ_{HLT}-candidate.

HLT Identification Finally, a more precise track measurement follows the idea of the one presented in Sec. 5.4.2; its results are subsequently used as input to a BDT-based identification step. Separate variables for τ_{HLT} with one or more than one associated tracks are constructed and used to discriminate signal from jet-background. The BDT score is stable under varying pile-up conditions and provides three working points with different signal efficiencies. For the analysis presented in this thesis, the τ_{HLT} must fulfil the medium identification requirement [121].

Efficiency corrections for the simulation of the single τ lepton triggers have been derived

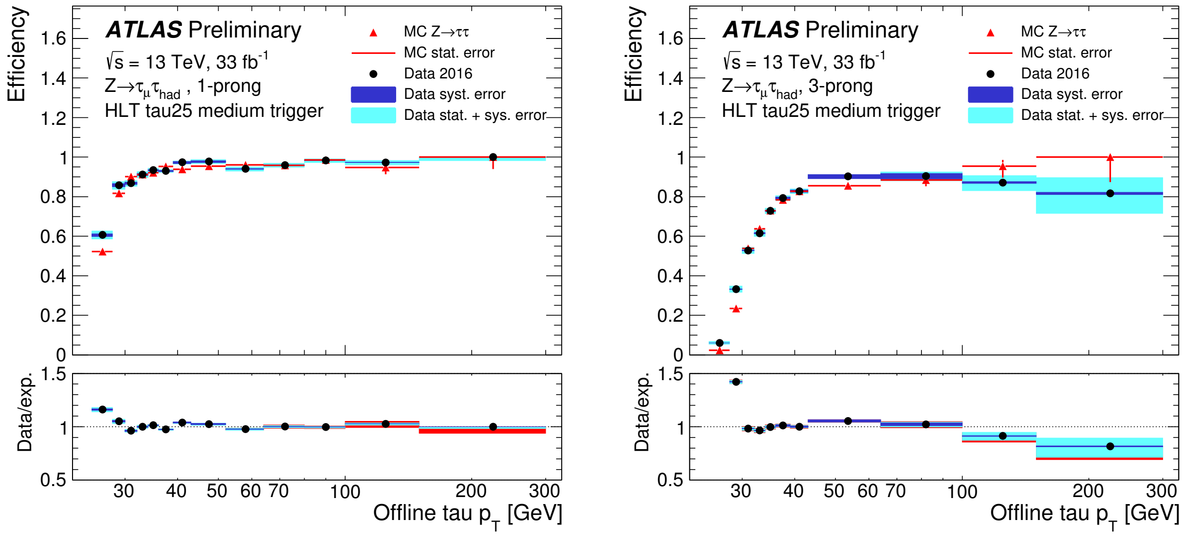


Figure 5.7: Measured and simulated efficiencies of the Run II single τ trigger with medium identification, parametrised in p_{T} of the matched τ_{had}-candidate. The measurement was performed in selected $Z \to \mu\tau_{\mathrm{had}}$ events. The left (right) plot shows the efficiencies for τ_{had}-candidates with one (three) associated tracks. Within combined uncertainties good agreement for $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ is observed [121].

in 2015 and 2016 data, depending on p_{T} and the number of associated track of the τ_{had}-candidates. The measurement is based on a similar $Z/\gamma^* \to \mu\tau_{\mathrm{had}}$ tag-and-probe approach as the identification efficiency measurement presented in Sec. 5.4.3.
The measured efficiency agrees with simulation as demonstrated in Fig. 5.7 and the corrections for $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ are close to 1. Further studies confirmed stability of these corrections against varying pile-up conditions and different η-ranges [12].

5.4.6 Energy Calibration of τ_{had} Leptons

The initial τ_{had} energy calibration used in the reconstruction and identification relies on the local hadronic calibration of the associated topological clusters and is based on calorimeter information only. This scale is corrected to account for the following effects:

- additional energy contributions originating from pile-up interactions
- τ_{had} decay products depositing all their energy in the inner detector material, before reaching the calorimeter
- decay products with energy deposits below the calorimeter threshold
- non-associated decay products outside the core cone of $\Delta R < 0.2$.

Aim of the two-step scale calibration is to correct the reconstructed to the simulated τ_{had} energy, in the following referred to as true energy. The τ_{had} momentum is given by its energy, as the mass is assumed to be zero.

While the first step of the energy calibration of the τ_{had} candidates applied in the

efficiency measurements discussed in Sec. 5.4.3 and Sec. 5.4.4 uses the baseline calibration described in [11, 12], the $H \to \tau\tau$ analysis described in the thesis relies on a multivariate approach. The latter uses Boosted Regression Trees (BRT) [122] and cures the inefficiencies observed in the baseline calibration procedure for low p_{T} regimes [12].
The BRT takes advantage of the Tau Particle Flow method, which provides the possibility to reconstruct individual hadrons in the τ_{had} decay. The information gained from this method improves the energy resolution at low p_{T} significantly due to the precise measurement of charged pions in the inner detector tracking system. They are used together with the baseline calibration quantities and additional calorimeter and tracking information as training input to the BRT. Most noticably, the momenta of individual topological cluster constituents associated to the τ_{had} are powerful input variables to the BRT. Fig. 5.8 demonstrates the resolution improvements achieved with the BRT calibration approach.

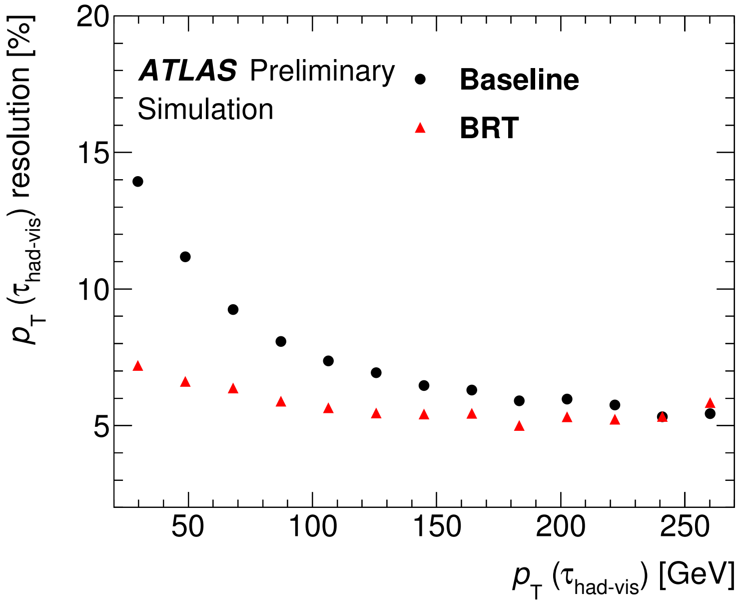


Figure 5.8: The resolution, defined as half of the standard deviation intervals of the ratio of calibrated p_{T} to true p_{T}, of baseline and the BRT-based τ_{had} energy calibration [12].

In a second step the resulting τ_{had} energy scale differences between data and simulation are corrected in an in-situ measurement. A similar tag-and-probe approach in $Z/\gamma^* \to \tau_\mu \tau_{\mathrm{had}}$ events as used in the identification efficiency measurement [12] is followed, estimating the shift of the energy scale in the visible reconstructed mass $m_{\mathrm{vis}}(\tau_\mu \tau_{\mathrm{had}})$ between data and simulation. To facilitate the measurement of the shift, the τ_{had} energy is parameterised as $E_\tau \to (1+\alpha)E_\tau$, while the muon energy scale is based on an independent measurement [110]. For the estimation of α, the data and simulated background model are compared and the value for α is determined, which yield the minimal discrepancy, quantified by a χ^2 hypothesis test. This shift is estimated in dependence of the number of associated core tracks to

$$\begin{aligned}\alpha = &\quad 0.95\% \pm 0.9\% \text{ (stat.) } \pm 1.7 \text{ (syst.)} \quad \text{for } \tau_{\text{had}} \text{ with one core track}\\ \alpha = &\quad -3.1\% \pm 1.1\% \text{ (stat.) } \pm 1.6 \text{ (syst.)} \quad \text{for } \tau_{\text{had}} \text{ with three core tracks .}\end{aligned}$$

The systematic uncertainties are arising from the uncertainty on the BRT measurement, the muon reconstruction, identification and resolution and from the background modelling.

5.5 Missing Transverse Energy

The presence of neutrinos in an event creates an imbalance in the energy distribution in the detector. This imbalance can be quantified by the (negative) vectorial sum of the transverse momenta of all reconstructed physics objects[1] o:

$$E_i^{\text{miss}} = -\left[E_i^{\text{miss,soft}} + \sum_o E_i^{\text{miss},o}\right] \quad \text{for } i = x, y$$

and the transverse momentum of all inner detector tracks not associated to any physics object $E_i^{\text{miss,soft}}$ [123]. These soft tracks are required to originate from the primary vertex, which effectively reduce fake E_i^{miss} contributions from pile-up or detector noise.

The importance of the missing transverse momentum in this analysis arises from the targeted event topology with at least two neutrinos. The vector's magnitude and its direction are given as

$$E_{\text{T}}^{\text{miss}} = \sqrt{\left[E_x^{\text{miss}}\right]^2 + \left[E_y^{\text{miss}}\right]^2} \quad \text{and} \quad \phi^{\text{miss}} = \arctan\left(E_y^{\text{miss}}/E_x^{\text{miss}}\right) .$$

These quantities are crucial to the analysis, in particular for the $\tau_{\text{had}}\tau_{\text{had}}$ mass reconstruction algorithm detailed in the following section.

5.6 $\tau_{\text{had}}\tau_{\text{had}}$ Mass Reconstruction

The reconstruction of the invariant mass of the $\tau\tau$-pair is a difficult task due to the two ν_τ in the event. The visible $\tau_{\text{had}}\tau_{\text{had}}$ mass m_{vis} is determined by the momentum and energy of the decay products τ_{had} in a straightforward manner but gives a broad resolution and a potential bias by neglecting the invisible ν_τ momentum contributions. Including these in the mass estimation introduces unknown quantities but offers significant improvements in resolution in particular for low-mass Higgs boson searches dominated by Z → $\tau\tau$ backgrounds [124]. In the following, two methods for a more precise $\tau\tau$-mass reconstruction are discussed.

---

[1] Namely electrons, muons, hadronic taus and jets in this analysis.

5.6.1 Collinear Approximation

The collinear approximation [124] is based on two assumptions:

- the invisible decay products propagate in the same direction as the respective visible τ_{had}; they are collinear
- all $E_{\text{T}}^{\text{miss}}$ contributions originate from the ν_τ produced in the τ-decay

With these assumptions and information about kinematics and topology of the $E_{\text{T}}^{\text{miss}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ in the event, the true $\tau\tau$-mass can be approximated by

$$m_{\text{coll}} = \frac{m_{\text{vis}}}{\sqrt{x_0 \cdot x_1}}$$

The visible momentum fractions x_i quantify the momentum carried by the visible decay products of τ_i [124]. The assumptions restrict the events for which the mass reconstruction gives a satisfying resolution to a boosted topology. While the assumptions for events with a high p_{T} jet and a close-by $\tau_{\text{had}}\tau_{\text{had}}$ pair work well, they tend to produce long tails of high masses for non-boosted scenarios. This is due to diverging solutions to the equations discussed in [124], when the difference in ϕ between the two τ leptons approaches π. In $H \to \tau\tau$ searches, the collinear mass approximation is hence not ideal, as this resonant decay implies a large contribution of back-to-back τ decays, i.e. $\Delta\phi_{\tau\tau} \to \pi$. A long tail of the $Z/\gamma^* \to \tau\tau$ mass spectrum leads to contamination of the signal mass region. This is reflected in the two distributions shown in Fig. 5.9. Contributions from $Z/\gamma^* \to \tau\tau$ events are prominent across the full signal mass range and particularly pronounced in the non-boosted VBF selection (left).

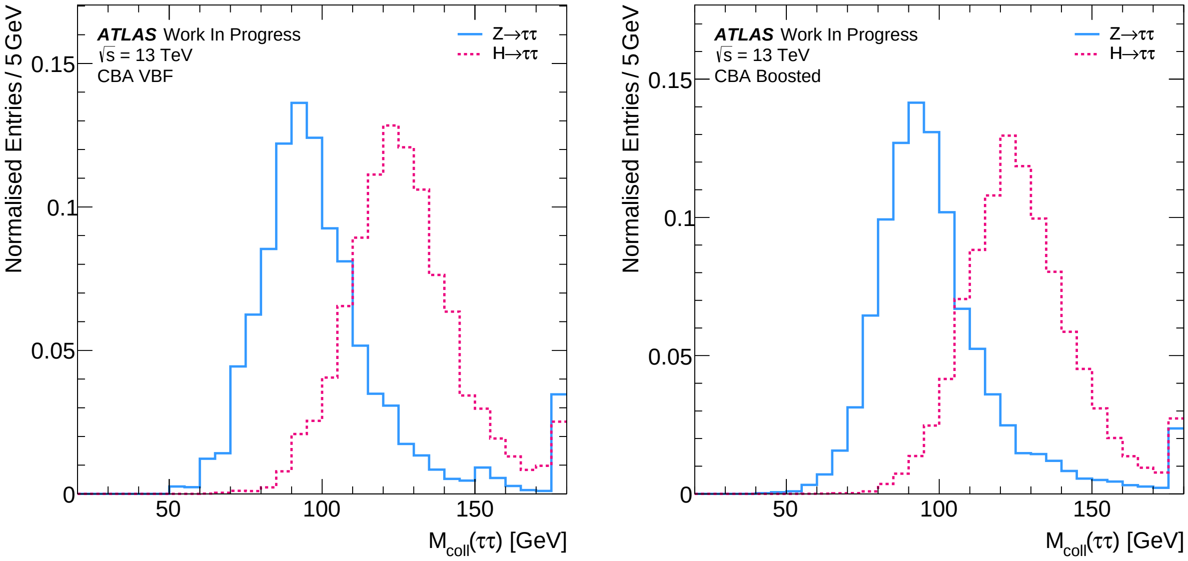


Figure 5.9: The reconstructed $\tau_{\text{had}}\tau_{\text{had}}$ mass in the collinear approximation for the $Z/\gamma^* \to \tau\tau$ and $H \to \tau\tau$ (with $m_H = 125\,\text{GeV}$) processes in the CBA VBF (left) and CBA Boosted (right) selection. The samples and selections are introduced in Chapter (7) and Sec. 6.5, respectively.

5.6.2 MMC Algorithm

The Missing Mass Calculator (MMC) is an algorithm aiming for a full reconstruction of the event kinematics without restricting the event topologies [124].
In the $\tau_{\text{had}}\tau_{\text{had}}$ channel this can be achieved by solving an underconstrained system of four equations with six unknown variables, the components of the invisible momenta p_{miss} carried by the ν_τ. The kinematics of the $\tau_{\text{had}}\tau_{\text{had}}$ system can be expressed as:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \cos\phi^{\tau_0}_{\text{miss}} & \cos\phi^{\tau_1}_{\text{miss}} \\ \sin\phi^{\tau_0}_{\text{miss}} & \sin\phi^{\tau_1}_{\text{miss}} \end{pmatrix} \begin{pmatrix} p^{\tau_0}_{\text{miss}}\sin\theta^{\tau_0}_{\text{miss}} \\ p^{\tau_1}_{\text{miss}}\sin\theta^{\tau_1}_{\text{miss}} \end{pmatrix}$$

$$\text{for } \tau_i \text{ with } i = 0,1: \quad m_\tau^2 = m_{\text{vis}}^2 + 2p_{\text{miss}}\left(\sqrt{p_{\text{vis}}^2 + m_{\text{vis}}^2} - 2\, p_{\text{vis}}\cos\left(\theta_{\text{vis}} - \theta_{\text{miss}}\right)\right)$$

with unknown variables $\phi_{\text{miss}}, \theta_{\text{miss}}$ and p_{miss} describing the angles and momenta of the ν_τ.
The τ mass is given by $m_\tau = 1.777\,\text{GeV}$ [27].
For a given point in the $(\phi^{\tau_0}_{\text{miss}}, \phi^{\tau_1}_{\text{miss}})$-plane, the angular distance ΔR between visible and invisible decay products can be calculated. On top of that, the m_{MMC} algorithm uses prior knowledge about kinematic constraints and τ decay topologies encoded in probability density functions (PDF) to define topological event likelihoods. The dependence of ΔR on the initial momentum of each τ is parameterised[1] and the resulting PDFs are used to evaluate probabilities for a given τ decay topology.

By performing a grid scan over points in the $(\phi^{\tau_0}_{\text{miss}}, \phi^{\tau_1}_{\text{miss}})$-plane, the configuration with the highest probability is determined. This configuration yields the most likely mass of

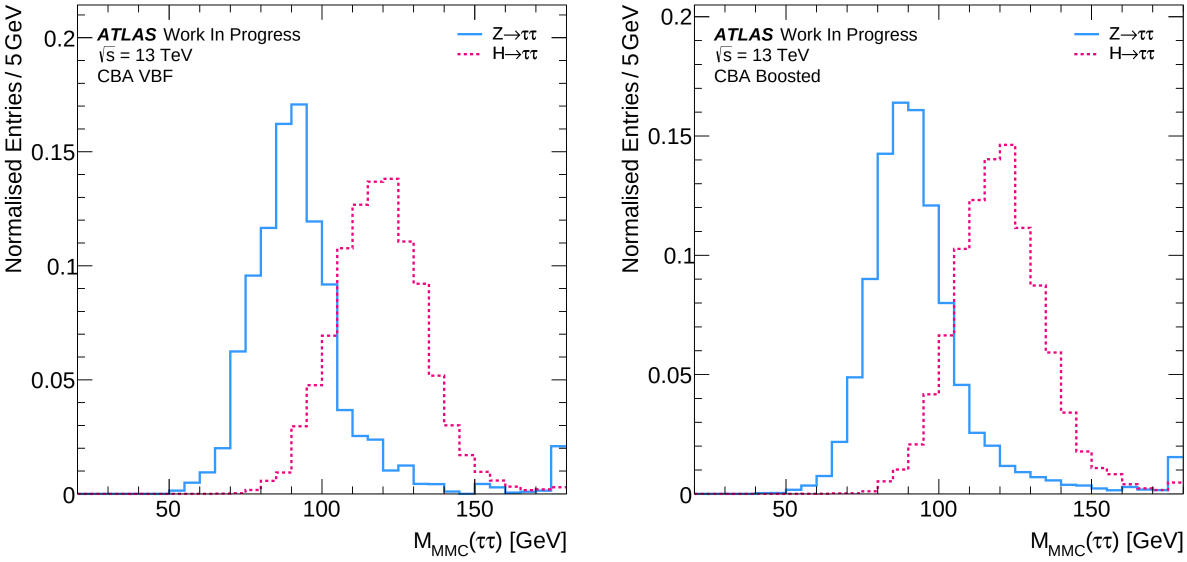


Figure 5.10: The reconstructed m_{MMC} in the $\tau_{\text{had}}\tau_{\text{had}}$ channel for the $Z/\gamma^* \to \tau\tau$ and $H \to \tau\tau$ (with $m_H = 125\,\text{GeV}$) processes in the CBA VBF and CBA Boosted selection. The samples and selections are introduced in Chapter (7) and Sec. 6.5, respectively.

---

[1] Using a linear combination of Gaussian- and Landau-functions.

the $\tau\tau$-system m_{MMC} and provides a significantly improved mass resolution for all decay topologies compared to the collinear approximation [124].
Due to the high correlation between MMC performance and $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, the MMC algorithm is designed to allow for possible mis-measurements of $E_{\mathrm{T}}^{\mathrm{miss}}$. If the kinematics of a given event are outside the scan range or in case of large $E_{\mathrm{T}}^{\mathrm{miss}}$ fluctuations, it is possible that the MMC does not find a solution. The MMC efficiency given as the ratio of events with solution and the initial events is $> 99\%$ for $H \to \tau\tau$ and $Z/\gamma^{*} \to \tau\tau$ processes but lower for background events, providing an additional handle on them. In Fig. 5.10, the m_{MMC}-distributions for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ CBA VBF and CBA Boosted categories are shown for signal and the $Z/\gamma^{*} \to \tau\tau$ processes. The peaks reconstructed from these resonant decays are clearly separated for both distinct event topologies. Including detector resolution effects, the MMC-algorithm achieves a resolution of 14% [124]. The contamination of signal events by $Z/\gamma^{*} \to \tau\tau$ events with large reconstructed mass is less pronounced compared to the collinear approximation technique.

CHAPTER 6

Event Selection and Categorisation

In the following, the object and event selection for the $\tau_{\text{had}}\tau_{\text{had}}$ analysis is motivated and summarised.

After a brief overview of the analysis strategy in Sec. 6.1, the triggers used to record the analysed dataset and selection requirements yielding well-identified physics objects are presented. Thereafter, the preselection, first stage of the event selection and starting point of the background estimation is discussed in Sec. 6.5. Subsequently, the categorisation details the second stage of the selection, the definition of signal regions.

The general strategy for selecting events follows ideas of the RUN I-analysis [9]. Various studies and cross checks were performed to adapt and improve the selection efficiency for RUN II conditions. Given the higher centre-of-mass energy of 13 TeV in the RUN II-dataset compared to the RUN I-analysis, events are expected to receive additional boost from their respective mother-particles. Additionally, the increased luminosity and pile-up activity lead to an increase of multijet events, hence, emphasis was put on the separation of signal from these background contributions.

6.1 Analysis Strategy

As a consequence of the possible final states in the $H \to \tau\tau$ channel, three orthogonal analyses were designed to target the different decay mode combinations of the τ-pair:

- the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel, investigating the decays into two leptons,
- the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, analysing events where one τ decays leptonically and the other hadronically and
- the $\tau_{\text{had}}\tau_{\text{had}}$ channel, where both τ decay hadronically.

Due to different background compositions, independent background estimations were performed in each channel. The analyses are designed to facilitate a combination of all three channels in a common fitting procedure after the channel-specific construction of final discriminants in dedicated signal regions. In order to ensure stability of this combination, categorisation and object definitions are harmonised where possible between the three channels. In this thesis, the complete, standalone $\tau_{\text{had}}\tau_{\text{had}}$ analysis is presented.

The general selection strategy of the $\tau_{\text{had}}\tau_{\text{had}}$ analysis focuses on Higgs boson production in vector-boson fusion (VBF) and gluon-gluon fusion (ggF). Consequently, two distinct event topology classes are investigated, the VBF and Boosted categories. Signal contributions from other production modes, the most sizeable being the Higgs-strahlung production mode (VH) associated to a vector boson V, are also taken into account but not targeted explicitly. Two separate $\tau_{\text{had}}\tau_{\text{had}}$ analysis approaches were investigated:

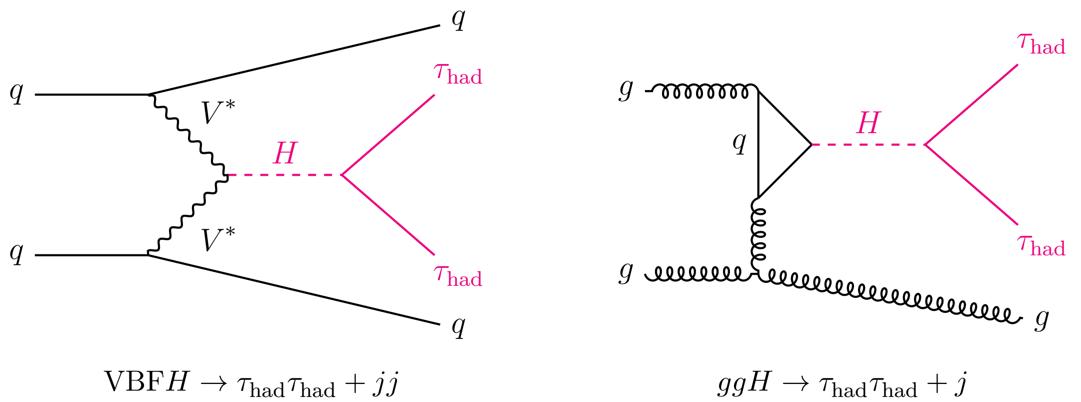


Figure 6.1: Sketch of Feynman diagrams for an example tree-level signal processes. The resulting signature is given by the $\tau_{\text{had}}\tau_{\text{had}}$ pair and jets j formed by the associated partons and motivates the event selection.

- the cut-based analysis (CBA), where requirements are imposed on signal-sensitive variables in order to select phase space regions with high signal-to-background ratio
- the multivariate analysis (MVA), in which a multivariate classifier is constructed to separate signal from background events.

Both approaches share the same background model, whereas the signal extraction procedure is independent, allowing for a cross check between the two.
The MVA event selection criteria were chosen to be looser than the CBA criteria in order to take full advantage of the separation power of the multivariate classifier.

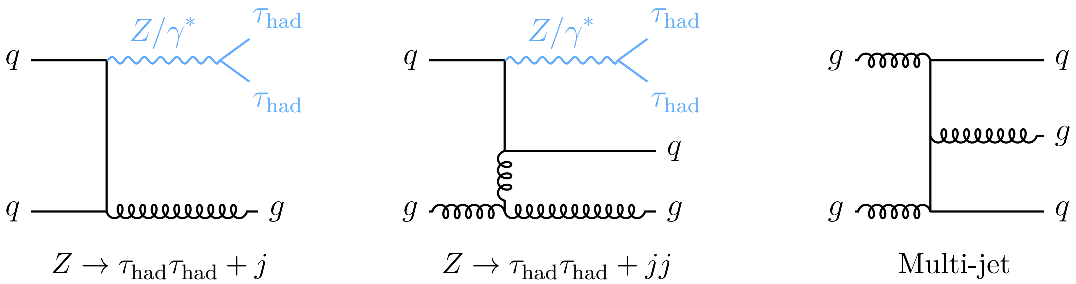


Figure 6.2: Sketch of Feynman diagrams for the dominant background processes. The background process composition depends on the targeted signal topology. While $Z/\gamma^* \to \tau\tau$ decays yield two real τ_{had}, contributions from multijet events emerge from misidentified jets faking τ_{had} signatures.

The main backgrounds in the $\tau_{\text{had}}\tau_{\text{had}}$ channel originate from τ_{had} pairs produced in decays of the Z boson and from jets faking τ_{had} formed in multijet events. Contributions from other processes are also considered but are small compared to the aforementioned processes. A thorough overview of estimation techniques is provided in Sec. 7.3. In Fig. 6.1 and Fig. 6.2, examples of tree-level Feynman diagrams are sketched, underlining the processes of the main signal- and background-event contributions.

6.2 Trigger Requirement

As motivated in Sec. 4.3, the recorded data was filtered during data taking. The $\tau_{\text{had}}\tau_{\text{had}}$ channel uses di-hadronic τ triggers constructed from two hadronically decaying τ trigger objects τ_{HLT}, reconstructed at the high-level trigger stage introduced in Sec. 4.3.3.
Assuming that both single τ lepton triggers are uncorrelated, corrections for the di-hadronic τ triggers are given by the product of the single τ lepton trigger corrections. This has been validated for $\text{Z} \rightarrow \tau\tau$ and signal processes in [121, 125].
The τ_{had} candidates used in the analysis are geometrically matched to the trigger objects by requiring their angular distance to be $\Delta R(\tau_{\text{HLT}}, \tau_{\text{had}}) < 0.1$.

Due to the increased luminosity and consequently higher trigger rates in the 2016 dataset compared to 2015, two different triggers were used:

- 2015: di-hadronic τ lepton trigger with (sub-)leading τ_{HLT} $p_{\text{T}} > 35\,(25)\,\text{GeV}$ and medium identification at HLT
- 2016: di-hadronic τ lepton trigger with the same requirements as in 2015 but an additional jet trigger object.

The requirement of an extra jet trigger-object with $p_{\text{T}} > 25\,\text{GeV}$ and $|\eta| < 3.2$ in 2016 reduces the trigger rate, while leaving freedom to construct the signal regions as detailed later in this chapter. The leading jet in the event is required to be within $\Delta R < 0.4$ of the jet trigger-object. In order to avoid biases from inefficient trigger-regions, a requirement of $p_{\text{T}} > 70\,\text{GeV}$ and $|\eta| < 3.2$ is imposed on the leading jet. Outside these ranges, simulation and measured trigger efficiencies differ significantly while satisfying agreement is observed within.

6.3 Object Selection

The considered objects for this analysis are a consequence of the topology of the $H \rightarrow \tau\tau$-events. In the $\tau_{\text{had}}\tau_{\text{had}}$ final state, two τ_{had} are required, as well as $E_{\text{T}}^{\text{miss}}$, describing the energy of the undetected ν_{τ} from the τ decays. Additional jets play an important role for defining signal categories and background suppression. Furthermore, electrons and muons are used to ensure orthogonality to the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{lep}}$ analyses. All objects are selected and corrected in accordance with the ATLAS RUN II recommendations.

τ_{had} candidates are required to have either one or three associated tracks, targeting decays with one or three $\pi^{\pm}$, respectively. An additional criteria on the reconstructed charge q of the τ_{had} candidate is set to $|q| = 1$. For the estimation of the multijet background contributions τ_{had} candidates with two tracks are also considered.

Candidates from non-fiducial detector regions with $|\eta| > 2.5$ are discarded from the analysis. Additionally, the transition region $1.37 < |\eta| < 1.52$ between barrel and end cap calorimeter is excluded in order to avoid contributions from misidentified true electrons. All τ_{had} candidates must fulfil $p_{\text{T}} > 20\,\text{GeV}$. For the two τ_{had} with highest p_{T}, this requirement is tightened to 40 (30) GeV as explained in Sec. 5.4.5. The resulting p_{T} distribution for the leading τ_{had} for the signal and two main background sources are shown in Fig. 6.3 (left) for the CBA Preselection region introduced in Sec. 6.4.

The τ_{had} candidates are distinguished from quark- and gluon-initiated jets by using the BDT-identification criteria described in Sec. 5.4.3. The two τ_{had} with highest p_{T} in the event must fulfil the tight identification threshold, corresponding to an identification efficiency of $45-60\%$[1] [12]. For the multijet background estimation and the training of BDTs these criteria are loosened to the medium ($60-75\%$ efficiency) and loose ($75-85\%$ efficiency) threshold. Fig. 6.3 (right) shows the different distributions for real τ_{had} originating from signal and $Z/\gamma^{*}\rightarrow\tau\tau$ processes and fake τ_{had} from multijet processes. The former tend to high scores around 1 while fake τ_{had} score lower.

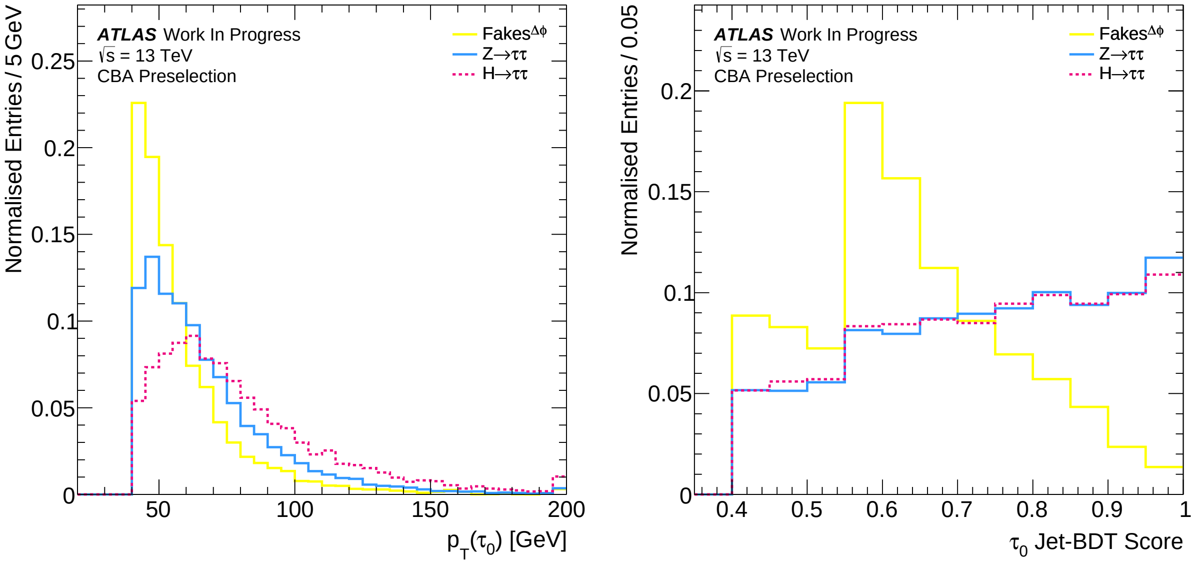


Figure 6.3: The leading τ_{had} p_{T} (left) and jet-BDT score (right) distribution for the signal (dashed, magenta) and the $Z/\gamma^{*}\rightarrow\tau\tau$ process (blue) and the multijet background (yellow) at CBA preselection. The samples and selections are introduced in Chapter (7) and Sec. 6.4, respectively. The step in the jet-BDT score distribution around 0.55 is an artifact of the τ_{had} trigger identification requirements.

Contributions from misidentified electrons are reduced by a p_{T} and η dependent factor using the electron-Likelihood as described in Sec. 5.4.4, with an efficiency of 95% [11, 12].

---

[1] Depending on the number of associated tracks and p_{T}

Jets are restricted to detector regions with $|\eta| < 4.5$ and are only considered in the analysis if they exceed a p_{T} threshold of 20 GeV.
Additional jets in the detector arising from pile-up contributions from the same or close-by bunch-crossings are rejected using the jet vertex tagging-Likelihood (JVT) [126].
The JVT was introduced in RUN II to cope with increased luminosity and to distinguish jets from hard-scattering processes from pile-up activity. It is based on the RUN I jet vertex fraction, defined as the ratio of summed transverse momenta of the jet's tracks associated to the primary vertex and associated to any vertex in the event [127]. All jets with $p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.4$ must fulfil $|\mathrm{JVT}| > 0.59$.
The so-called forward-JVT algorithm is used to suppress pile-up jets with $p_{\mathrm{T}} < 50$ GeV in the detector's forward region $|\eta| > 2.5$ by requiring a lower cut of 0.4 on this forward-JVT Likelihood score [128].

Electrons and Muons are defined following the criteria used in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ channel. By vetoing events with electrons or muons, orthogonality with the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ channels is achieved. The quality criteria for tracks forming electrons and muons are detailed in Sec. 5.2. Electrons must pass the loose identification criteria and exceed a transverse momentum of $p_{\mathrm{T}} > 15$ GeV. Detector regions above $|\eta| > 2.47$ and in-between $1.37 < |\eta| < 1.52$ are excluded, constraining the electron reconstruction to the fiducial inner detector volume and excluding the transition region between barrel and end-cap calorimeters. For muon candidates a minimum threshold for p_{T} is set at 10 GeV, while restricting the detector region to $|\eta| < 2.5$. The loose identification criteria is required, which is based on track information from the inner detector and muon spectrometer.

Overlap removal The reconstruction of the above objects is based on the same elementary output of the ATLAS detector, in particular tracks and clusters. Hence, different physics objects can be associated to the same basic detector signals. In order to avoid double-counting of objects, a so called overlap removal is used to discard any duplication based on the same true physics object in the detector. A geometrical match in ΔR between the reconstructed objects is performed and any ambiguity removed from the event. A common strategy is defined across the three $H \to \tau\tau$ channels by performing the overlap removal procedure in order of misidentification probabilities for objects, i.e.

1. if a jet overlaps with a τ_{had} candidate in $\Delta R < 0.2$ it is removed in favour of the τ_{had} candidate
2. if a jet is closer than $\Delta R < 0.4$ of an electron or muon, it is discarded
3. if an overlap between τ_{had} candidate and an electron or muon is found in $\Delta R < 0.2$, it is removed
4. if an electron is within $\Delta R < 0.2$ of a muon, it is rejected

In the $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel, this procedure simply results in a removal of any jet within $\Delta R < 0.2$ of a τ_{had} candidate, since events with additional electrons or muons are discarded.

6.4 Event Preselection

The preselection categories for the $\tau_{\text{had}}\tau_{\text{had}}$ analyses are constructed starting from the trigger conditions and object definitions. As will be discussed in the following, these criteria remove badly reconstructed events and ensure the presence of two well-defined τ_{had} in a kinematic region relevant for this analysis. These preselection categories are used to validate the background modelling and form the basis for the signal regions as described in Sec. 6.5.

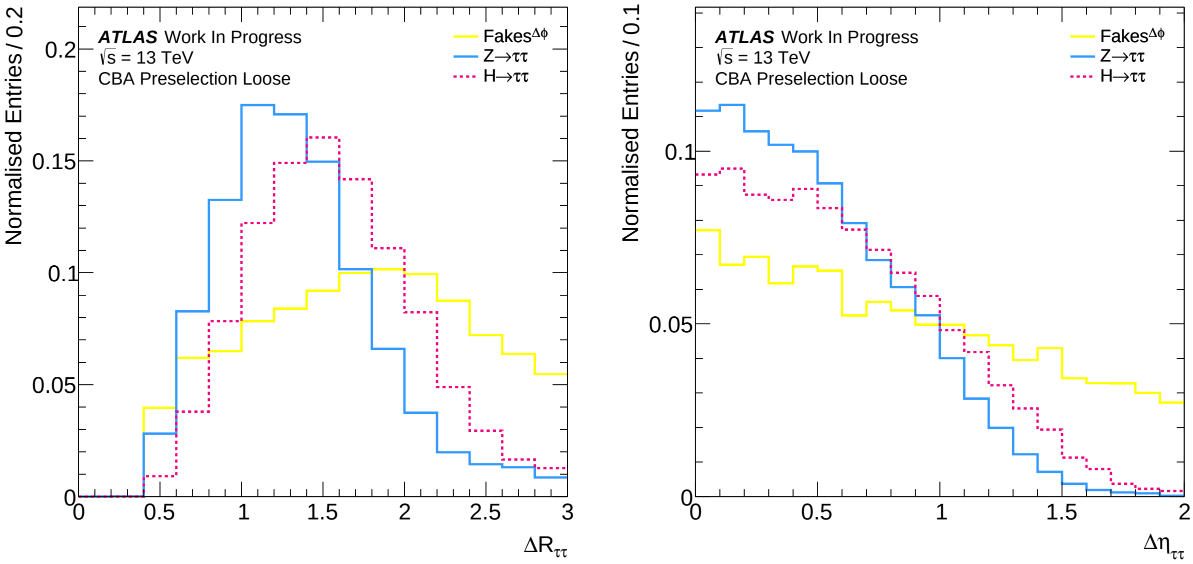


Figure 6.4: The $\Delta R_{\tau\tau}$ (left) and $\Delta\eta_{\tau\tau}$ (right) distribution for the signal, $Z/\gamma^* \rightarrow \tau\tau$ and the multijet background in the CBA preselection loose region. The event model is introduced in Chapter (7).

During the first step of the preselection, events with problematic data taking conditions are rejected as follows:

- At least one primary vertex is required with a minimum of four associated tracks. This effectively rejects events from cosmic radiation [129].
- Any event with a jet reconstructed from calorimeter noise and energy spikes, machine effects like varying LHC beam conditions or showers induced by cosmic radiation is discarded [130].
- Events where the reconstruction of the m_{MMC} mass failed are rejected.

Orthogonality to the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{lep}}$ channel is achieved by vetoing any event with an electron or muon.
The trigger requirement discussed in Sec. 6.2 imposes the presence of at least one jet with $p_{\text{T}} > 70\,\text{GeV}$ and $|\eta| < 3.2$.

After these initial steps, only events with two suitable τ_{had} are selected, in particular:

- Exactly two τ_{had} candidates with tight identification threshold are required, in order to reject fake τ_{had}-candidates arising from multijet contributions.

- Both selected τ_{had} candidates originate from the same primary vertex and are matched to the dihadronic τ lepton trigger objects.
- Events are discarded if more than two τ_{had} candidates with medium identification threshold are found.
- Ordered by their respective p_{T}, the (sub-)leading τ_{had} candidate has to exceed $p_{\mathrm{T}} > 40\,(30)$ GeV. These thresholds where chosen according to the τ_{HLT} trigger object p_{T} thresholds described in Sec. 6.2 and the corresponding trigger efficiencies.

The following criteria are applied for preselection and categorisation, but are altered for construction of the regions used to estimate the multijet background shapes:

- Both τ_{had} candidate have opposite charge $q(\tau_0) \times q(\tau_1) = -1$, each with an absolute value of $|q(\tau_{\mathrm{had}})| = 1$
- Related to this, each τ_{had} candidate must be associated with either one or three tracks

The two ν_τ from the τ_{had} decays carry a sizeable fraction of the τ-pairs energy, which cannot be measured directly. Therefore, a lower boundary of 20 GeV is required for the $E_{\mathrm{T}}^{\mathrm{miss}}$ in all events, targeting events with low $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions from multijet processes.

τ_{had}-pairs produced from Higgs or Z bosons are expected to decay back-to-back in their rest-frame but receive a boost when observed in the detector (i.e. the laboratory frame) due to the momentum of the respective boson. Given the non-resonant structure of multijet events, this is not the case for jets faking τ_{had}. This behaviour is reflected in the angular distance $\Delta R_{\tau\tau}$ and the pseudo-rapidity gap $\Delta\eta_{\tau\tau}$. These quantities are used to define

- the **Preselection loose** region by requiring $0 < \Delta R_{\tau\tau} < 3$ and $|\Delta\eta_{\tau\tau}| < 2$, used to extract the template reweighting for fake τ_{had} backgrounds
- the **Preselection Fit** region by tightening to $0 < \Delta R_{\tau\tau} < 2.5$, used to fit normalisations and
- the **MVA Preselection** region where also $|\Delta\eta_{\tau\tau}| < 1.5$ is imposed.

The requirements on $\Delta R_{\tau\tau}$ remove possible overlaps between the τ_{had} and, together with the $\Delta\eta_{\tau\tau}$ requirements, reduce the contribution from multijet events effectively. This is visualised Fig. 6.4 for the CBA Preselection loose category.

Additional topological selection requirements are used to define the **CBA Preselection** regions. In the collinear approximation [124], it is assumed that the τ-pair decay products propagate in the same direction as the respective τ due to the boost of their mother particles. Details are given in Sec. 5.6. To ensure that the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction is aligned to the τ_{had}, a cut on the visible momentum fractions x_0 (x_1) of the (sub-)leading τ_{had} is applied:

$$0.1 < x_{0/1} < 1.4\ .$$

This requirement further reduces contributions from non-resonant decays. The corresponding distribution for x_0 is shown in Fig. 6.5 (left).

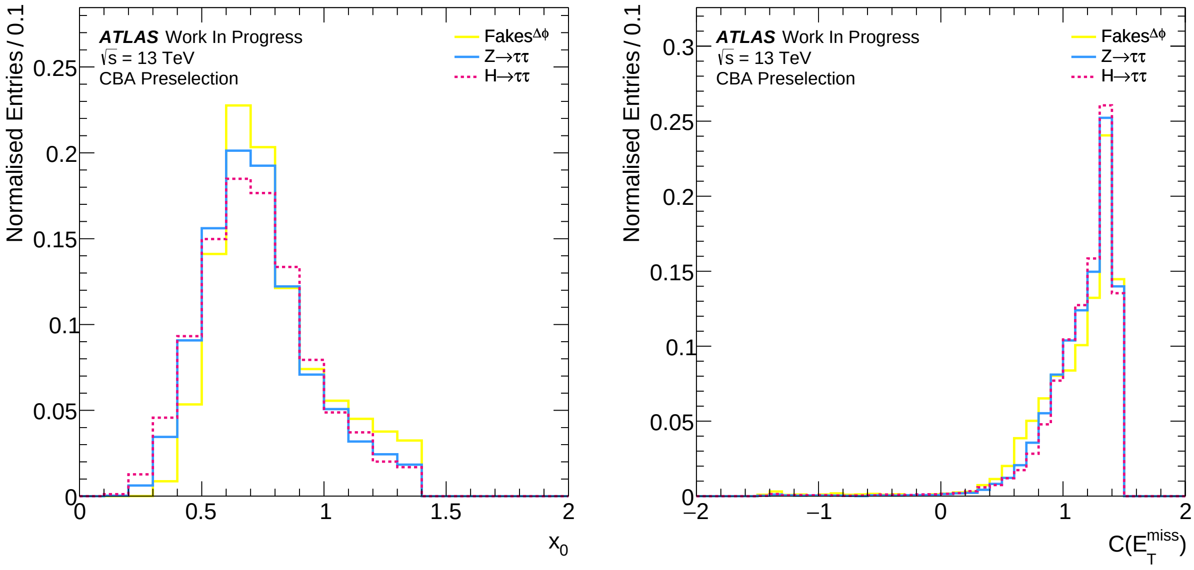


Figure 6.5: Visible momentum fraction x_0 for the leading τ_{had} (left) and $\mathcal{C}_{\tau\tau}(E_{\text{T}}^{\text{miss}})$ (right) distributions for the signal, $Z/\gamma^* \to \tau\tau$ and the multijet background in the CBA preselection region. The event model is introduced in Chapter (7).

6.5 Categorisation

Following the preselection of events, a categorisation in the aforementioned VBF and Boosted event classes is defined. Distinct features of the signal processes' kinematics and topology are used to achieve a maximal separation from background contributions.

6.5.1 Topology of Signal Events and Discriminating Variables

In the following, the definition of distinct $H \to \tau\tau$ event characteristics used to define VBF and Boosted categories is discussed. The VBF category exploits unique features of the VBF-production mode to distinguish signal from Z $\to \tau\tau$ and multijet processes, in particular the kinematics and topology of the two well-separated jets from the VBF-scattering.

The Boosted category, targets events with Higgs bosons produced in ggF, which are expected to have a high p_{T} as a consequence of being recoiled against a jet. This results in a boosted decay topology for the τ_{had} pair and a high-p_{T} jet.

Both categories share common discriminating features, in particular:

m_{MMC} The mass of the $\tau_{\text{had}}\tau_{\text{had}}$ system, reconstructed as detailed in Sec. 5.6, offers the highest separation power between signal and background. This variable helps to discriminate resonant from non-resonant events and provides a distinct region around 125 GeV, where signal events are populated. The cut-based analysis takes advantage of the separation power by using m_{MMC} in the profile likelihood fit. The implications for the multivariate analysis are discussed in Sec. 8.2.2.2.

$\boldsymbol{\Delta R_{\tau\tau}}$ describes the angular distance between the two τ_{had} and is calculated from the η and ϕ coordinates of the detector's cylindrical coordinate system. Low values are a result of two close-by τ_{had}, which is in particular the case in the boosted category and a characteristic of resonant decays. This variable is strongly correlated to the p_{T} of the Higgs boson candidate, as a high momentum of a mother-particle implies a boost of the daughter particles, bringing them closer together. This is not the case for non-resonant background contributions, which provides a handle specifically on multijet induced backgrounds.

$\boldsymbol{\mathcal{C}(E_{\mathrm{T}}^{\mathrm{miss}})}$ The centrality of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is a measure of the $E_{\mathrm{T}}^{\mathrm{miss}}$ position relative to the τ_{had} pair. A projection of $E_{\mathrm{T}}^{\mathrm{miss}}$ in a plane where the τ_{had} are orthogonal to each other is constructed using ξ and ζ, which results in the $E_{\mathrm{T}}^{\mathrm{miss}}$ centrality definition as follows:

$$\begin{aligned}
\xi =& \frac{\sin\phi(E_{\mathrm{T}}^{\mathrm{miss}}) - \phi(\tau_0)}{\sin\phi(\tau_0) - \phi(\tau_1)} \\
\zeta =& \frac{\sin\phi(E_{\mathrm{T}}^{\mathrm{miss}}) - \phi(\tau_0)}{\sin\phi(\tau_0) - \phi(\tau_1)} \\
\mathcal{C}(E_{\mathrm{T}}^{\mathrm{miss}}) =& \frac{1}{\sqrt{2}} \frac{\xi + \zeta}{\sqrt{\xi^2 + \zeta^2}} \qquad (6.1)
\end{aligned}$$

If the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction is exactly central between the τ_{had} pair, this centrality is 1. In case the $E_{\mathrm{T}}^{\mathrm{miss}}$ is perfectly aligned with one of the τ_{had}, the formula reduces to $1/\sqrt{2}$. Finally, in case the $E_{\mathrm{T}}^{\mathrm{miss}}$ is pointing directly opposite of the central bisection of the τ_{had} pair plane, the centrality is -1.
Due to the correlation between the boson's decay products, $\mathcal{C}(E_{\mathrm{T}}^{\mathrm{miss}})$ is expected to yield higher values for signal and $\mathrm{Z} \rightarrow \tau\tau$ events than for events from multijet background. This is reflected in Fig. 6.5 (right).

The following variables are specific to the VBF category and require at least two jets in the event.

$\boldsymbol{m_{jj}}$ The invariant mass of the two jets in the VBF category quantifies the inherent mass of the dijet system. On top of jet-p_{T} requirements, this ensures that both jets form a high-energetic system, a core feature of the VBF-scattering. Inspecting Fig. 6.6 (left), it becomes apparent that signal events produce a longer tail for large m_{jj} values. This fact will be exploited for efficient background suppression and categorisation.

$\boldsymbol{\Delta\eta_{jj}}$ is defined as the difference in pseudo-rapidity between the two VBF jets. As both jets produced in VBF are expected to be in the opposite forward regions of the detector, large values indicate a VBF-signal like event, while low values are typical for background contributions. This variable is strongly correlated with m_{jj} and also exhibits pronounced tails towards large values for the signal processes, as can be seen in Fig. 6.6 (right).

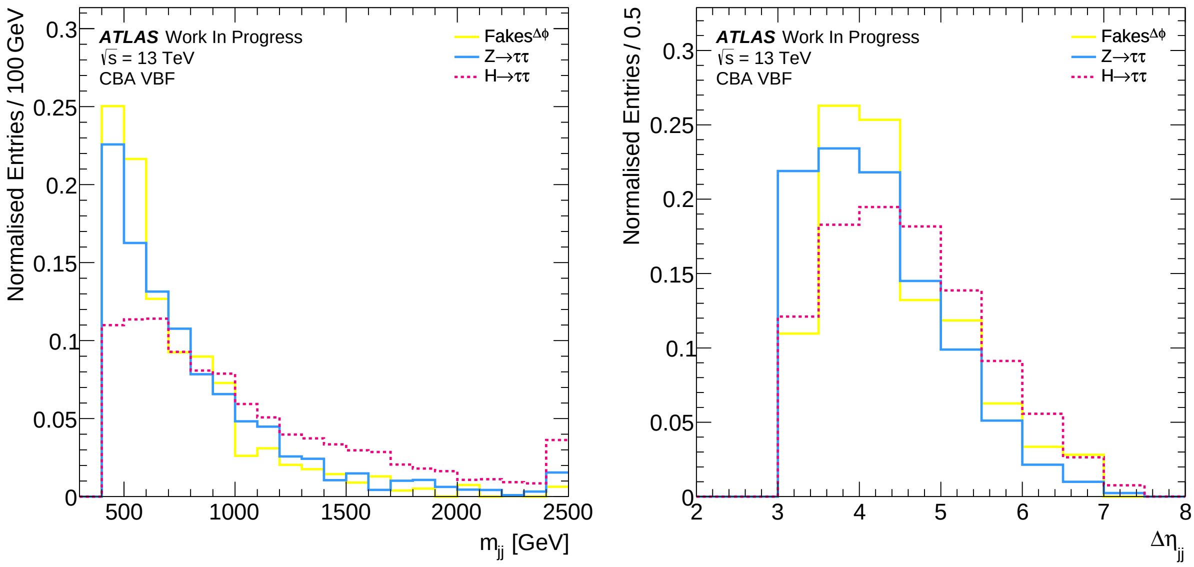


Figure 6.6: The m_{jj} (left) and $\Delta\eta_{jj}$ (right) distribution for the signal, $Z/\gamma^* \to \tau\tau$ and the multijet background in the CBA VBF inclusive region. The event model is introduced in Chapter (7).

$\eta_{j_0} \times \eta_{j_1}$ Furthermore, the product of the jet pseudo-rapidities $\eta(j_0) \times \eta(j_1)$ is differentiating between events where both jets are in opposite (VBF-like) or in the same detector-hemisphere.

$\mathcal{C}_{jj}(\tau_{\text{had}})$ The relative positions of the τ_{had} to jets in VBF events, called τ-centrality, quantifies another distinguishing characteristic of VBF signal events. It is constructed as follows:

$$\mathcal{C}_{jj}(\tau_{\text{had}}) = \exp\left[\frac{-4}{\left(\eta\left(j_0\right) - \eta\left(j_1\right)\right)^2}\left(\eta(\tau_i) - \frac{\eta(j_0)+\eta(j_1)}{2}\right)^2\right] \quad \text{with } i = 0, 1.$$

and resembles a Gaussian distribution, centred at the point exactly between the two VBF jets in η. The distributions' width in this definition is defined by the pseudo-rapidity gap of the jets.

For τ_{had} lying centrally between the two jets, this variable evaluates to 1. In case the τ_{had} are perfectly aligned to one of the jets, the resulting value for the centrality is $1/e$. Finally, values $[0, 1/e]$ are a result of a τ_{had} direction not in-between the two jets, indicating a background-like event. The distribution for signal and the two main backgrounds is presented in Fig. 6.7 (left).

$p_T(\tau\tau, jj, E_{\text{T}}^{\text{miss}})$ The total p_{T} of the τ_{had}, jets and $E_{\text{T}}^{\text{miss}}$ in the event is another measure for the Higgs boson system's energy, tending towards high values for signal events.

p_T^H The p_T of the Higgs boson candidate is calculated as the vectorial sum of the τ_had pair's momentum in the transverse plane and $E_\mathrm{T}^\mathrm{miss}$. It quantifies the transverse momentum of the resonance, resulting in larger values for signal than for background. The p_T^H spectra for the CBA VBF and Boosted inclusive region are presented in Fig. 6.7 (right) and Fig. 6.8 (right) respectively.

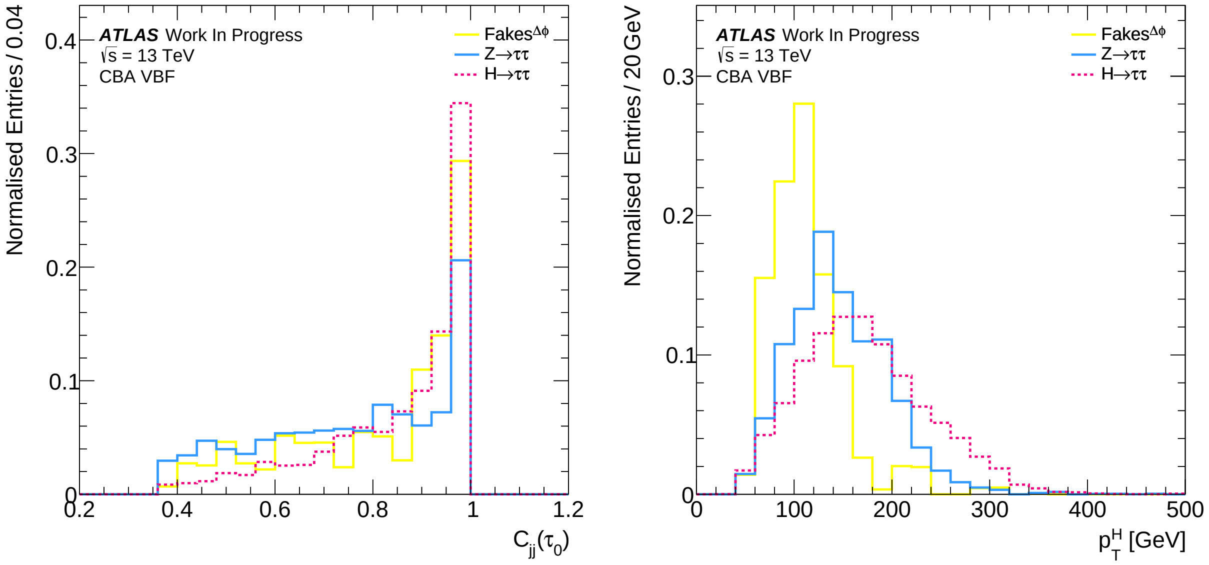


Figure 6.7: The $\mathcal{C}_{jj}(\tau_0)$(left) and p_T^H (right) distributions for the signal, $Z/\gamma^* \to \tau\tau$ and the multijet background in the CBA VBF inclusive region. The event model is introduced in Chapter (7).

6.5.2 VBF Categories

The variables introduced in the previous section are used to construct the VBF categories, as described in the following.

For cut-based and multivariate approach (CBA and MVA), the event is required to have:

- at least one additional jet with $p_\mathrm{T} > 30\,\mathrm{GeV}$
- which is separated from the leading jet by $|\Delta\eta_{jj}| > 3.0$
- and forms an invariant mass with the leading jet of $m_{jj} > 300\,\mathrm{GeV}$

For the MVA, this defines the VBF region called **MVA VBF**.

The selection criteria for **CBA VBF** are tightened using

- a higher cut on the di-jet mass $m_{jj} > 400\,\mathrm{GeV}$,
- a criterion ensuring that both jets are located in opposite hemispheres $\eta_{j_0} \times \eta_{j_1} < 0$
- and that the τ_had are lying in-between the two jets, characterised by their pseudo-rapidities.

From this inclusive CBA VBF region, three subcategories are defined by targeting different phase spaces in the CBA. This sub-categorisation strategy is based on [9, 131].

CBA VBF low-$\Delta R_{\tau\tau}$ By requiring $\Delta R_{\tau\tau} < 1.5$ and $p_{\mathrm{T}}^{H} > 140\,\mathrm{GeV}$, a phase space almost completely depleted of multijet backgrounds is constructed. Despite the strong correlation of these variables, both requirements are necessary for efficient background rejection. As shown in Fig. 6.9, the remaining background is dominated by $\mathrm{Z} \rightarrow \tau\tau$ events. Of all CBA regions, this category provides the highest signal-to-background ratio. The kinematic properties of the dijet system in the VBF event are exploited on the remaining CBA VBF events with high $\Delta R_{\tau\tau}$ or low p_{T}^{H}.

CBA VBF high-$\Delta R_{\tau\tau}$ tight Events with high m_{jj} and large pseudo-rapidity gap $\Delta\eta_{jj}$ are characteristic for signal processes. To exploit this feature, a two-dimensional requirement is applied, by requiring $m_{jj} > -250\,\mathrm{GeV} \times |\Delta\eta_{jj}| + 1\,550\,\mathrm{GeV}$ The so defined category gives the highest signal-to-background ratio and is the second most sensitive VBF region in CBA.

CBA VBF high-$\Delta R_{\tau\tau}$ loose All remaining CBA VBF events are collected in the high-$\Delta R_{\tau\tau}$ loose category. This category has a similar background composition as the category above, but includes less signal events.

In summary, the VBF categorisation yields one inclusive signal region for MVA and three distinct signal regions for CBA.

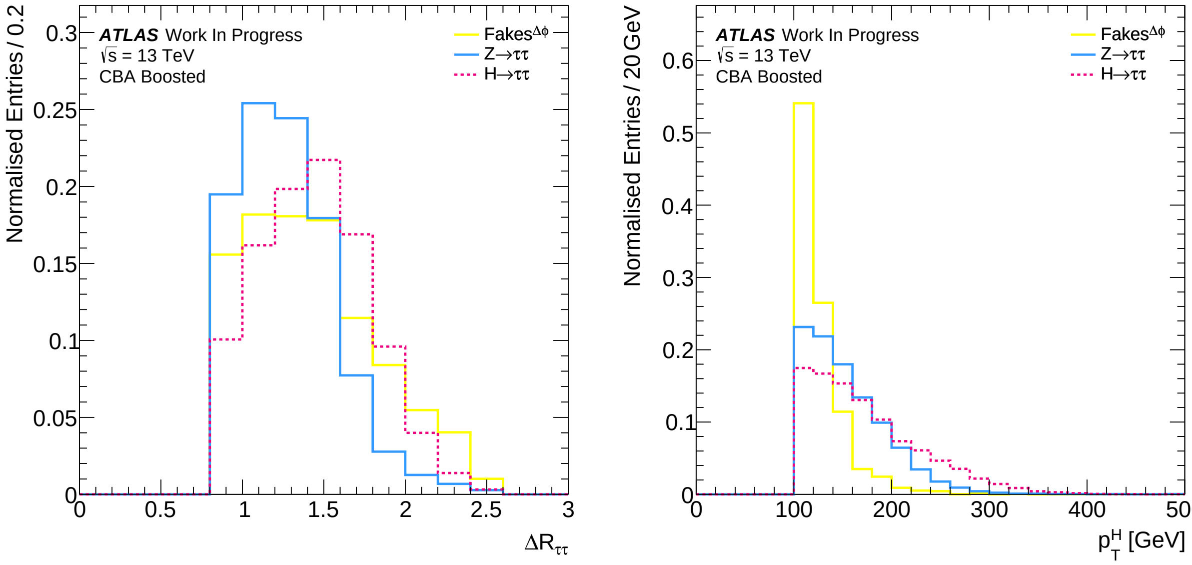


Figure 6.8: The $\Delta R_{\tau\tau}$ (left) and p_{T}^{H} (right) distributions for the signal, $Z/\gamma^{*} \rightarrow \tau\tau$ and the multijet background in the CBA Boosted inclusive region. The event model is introduced in Chapter (7).

6.5.3 Boosted Categories

All events which pass the respective preselection but fail the VBF requirements are collected in the **CBA/MVA Boosted** categories, if they fulfil $p_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$.

By differentiating between events with high and low p_{T}^{H}, the CBA Boosted category is split further, targeting different background compositions. The choice of thresholds can be understood by inspecting Fig. 6.8.

CBA Boosted tight is the overall second most sensitive category in CBA due to a high signal-to-background ratio and almost no contributions from multijet events. It is defined from all events passing the CBA Boosted selection and additionally fulfilling $\Delta R_{\tau\tau} < 1.5$ and $p_{\mathrm{T}}^{H} > 140\,\mathrm{GeV}$. Both variables are strongly correlated as is the case in CBA VBF, but again both requirements are needed for efficient multijet background suppression. As visualised in Fig. 6.9, background contributions in this category arise almost solely from $\mathrm{Z} \to \tau\tau$ events which is evident given that the criteria target resonant decays.

CBA Boosted loose Finally, an event is selected for the Boosted loose category, if it passes the Boosted inclusive requirement but fails the tight selection. Hence, events with either high $\Delta R_{\tau\tau}$ or $p_{\mathrm{T}}^{H} < 140\,\mathrm{GeV}$ are considered.

In summary, the Boosted categorisation results in one inclusive signal region for MVA and two distinct signal regions for CBA.

6.5.4 Signal and Background Compositions

The composition of processes in the selected categories are summarised in Fig. 6.9 and Fig. 6.10. The strategy results in different signal-to-background ratios and background compositions per subcategory. The estimation techniques for each contribution are detailed in Sec. 7.3.

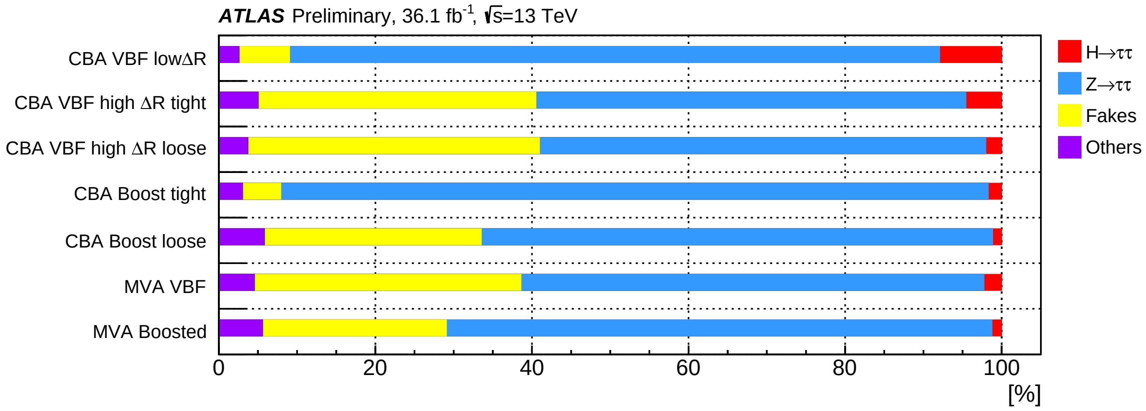


Figure 6.9: Relative process contributions for each signal region.

In Fig. 6.9, the relative fractions of background and signal events is illustrated. Using the selection described in the previous sections, dominant background contributions arise from Z $\rightarrow \tau\tau$ between $(50-90\%)$ and from multijet events $(5-40\%)$ depending on the category. Other background processes amount to $\approx 5\%$ across all categories. Furthermore, the composition of signal contributions is illustrated in Fig. 6.10 for the MVA categories. This information is later used to decide which signal process is considered in the training of the multivariate classifier.
In the MVA VBF, the dominant contribution is coming from VBFH signal events with $\sim 60\%$, followed by events produced in gg-Fusion with $\sim 35\%$. For MVA Boosted, the ggH production mode is dominant amounting to $\sim 75\%$ of the signal processes. Event migration of VBF-produced events to a more boosted topology is observed to $\sim 15\%$.

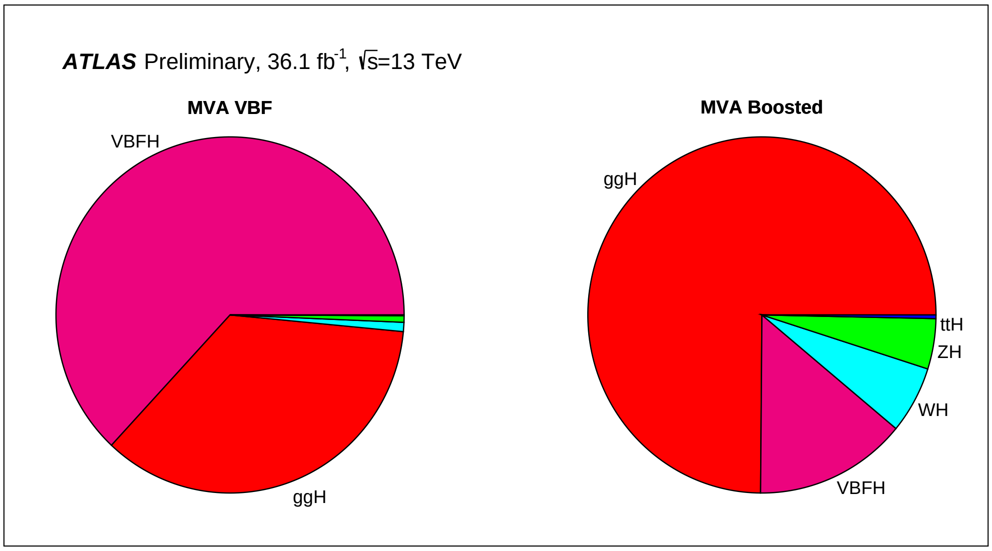


Figure 6.10: Relative signal contributions for the MVA VBF and Boosted categories. The labels for the sub-dominant production modes VH and ttH are suppressed in the left plot for readability.

CHAPTER 7

Event Modelling and Systematic Uncertainties

In the following chapter the modelling strategy for the processes considered in this analysis is presented. Furthermore, details about process-specific and general systematic uncertainties are described.

The estimation of contributions from the various signal and background processes is based on Monte-Carlo Simulation as well as data-driven techniques to estimate shapes, normalisations and uncertainties.
Following the official ATLAS recommendations, all objects and event properties which are based on simulation are corrected to better describe the recorded data. This is done with a variety of object-specific software packages during data processing. Each correction is based on a dedicated measurement and introduces systematic uncertainties, which are detailed in Sec. 7.6.

For all simulated events the ATLAS detector response is modelled using a full simulation of the detector based on the GEANT4 software [132, 133]. Pile-up effects from the same and nearby bunch-crossings of the simulated event are included in the simulation. For this, the simulated events are overlaid with minimum bias events generated with PYTHIA8 using the MSTW2008LO PDF in the A2TUNE [134, 135]. All simulated events pass the same reconstruction software as recorded data. Furthermore, differences between the simulated and observed pile-up description are corrected in a data-driven way. Independently for all luminosity blocks, each event receives a weight calculated from the difference in observed and simulated $\langle\mu\rangle$ distributions[1]. Therefore, the simulated luminosity profile of events corresponds to the profile of the recorded data set after correction within uncertainties.

7.1 Data Sample

The data sample analysed in this thesis was produced in 2015 and 2016 pp collisions at the LHC with a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and recorded with the ATLAS detector.

---

[1] $\langle\mu\rangle$ is defined as the average number of interactions per bunch crossing

The details on the experimental setup are discussed in Chapter (4).

By using the central data quality assessment in ATLAS, it is ensured that only well-measured events are used in the analysis. In particular it is required, that all detector subsystems, including the IBL, were fully operational, no defects were found and the physical objects are well-reconstructed.
ATLAS uses lists of luminosity blocks, called Good Runs List (GRL), to identify these events. In this analysis, a GRL for 2015 and 2016 was used to filter the recorded events.

The trigger requirements discussed in Sec. 6.2 restrain the available dataset further, as not all trigger items are always active during data taking. In particular, a so-called pre-scaling can be deployed to switch off parts of the HLT-processing. The small amount of events recorded with such a pre-scaling are excluded from the analysed dataset.

Given these constraints, the total usable integrated luminosity is $3.21\,\text{fb}^{-1}$ for 2015 and $32.86\,\text{fb}^{-1}$ for 2016. The combined dataset, henceforth referred to as Data amounts to $36.07\,\text{fb}^{-1}$. Dataset-specific analysis steps, like the trigger-correction explained in the following section, are done separately by splitting all simulated events according to the relative size of the 2015 and 2016 datasets. The background estimation and statistical analyses are performed on the combined datasets.

7.2 Signal Modelling

The expected signal contributions are estimated using Monte Carlo Simulation of the hard scattering process, parton shower, hadronisation, underlying event contributions and detector response. All production modes and τ decays described in Chapter (3) are taken into account. Given the measurements of the Higgs boson mass in other decay modes and the previous results of the search for the $H \to \tau\tau$ process, only one mass point $m_H = 125\,\text{GeV}$ is considered in this analysis.

The signal processes are simulated at a fixed perturbative QCD order and corrected for effects from higher order contributions by using cross-section values calculated at the highest available precision as discussed in Chapter (3). These are applied as global event weights. Corrections to the p_T-dependent differential cross sections are taken into account by applying event-by-event weights. The cross sections and simulated samples used in this analysis are listed in Appendix A. The generators used in the analysis are summarised in Tab. 7.1 and detailed in the following.

The underlying parton distribution functions (PDF) parametrisation for the matrix element calculation of all signal production modes is the same, namely PDF4LHC15 [136].
For the modelling of non-perturbative effects the PDF set CTEQ6L1 [78] with AZNLO tune [137] is used.
Furthermore, photon and other QED emissions from electroweak vertices and charged leptons are modelled using PHOTOS++ 3.52 for all samples [138].
The branching ratio for $H \to \tau\tau$ used for all signal processes is 0.06256 [54]. The hadronic decay branching ratio of the τ decays was introduced in Sec. 5.4.1.

Process	Generator		PDF set		Tune	Order
	ME	PS	ME	PS		
$H \to \tau\tau$						
ggF	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	NNLO+NNLL
VBF	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	(N)NLO
VH	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	NNLO
ttH	aMC@NLO	Pythia8	NNPDF23		A14	NLO
Simulated Backgrounds						
$V +$ jets	Sherpa 2.2.1		NNPDF30		Sherpa	NNLO
$t\bar{t}$	Powheg	Pythia6	CT10	CTEQ6L1	Perugia2012	NNLO+NNLL
Single top	Powheg	Pythia6	CT10	CTEQ6L1	Perugia2012	NNLO
Di-Boson	Sherpa 2.2.1		NNPDF30		Sherpa	NNLO

Table 7.1: Summary of the generators used for simulated samples. Details and references are listed in the corresponding sections.

ggFH Signal contributions arising from ggF-processes are calculated at QCD next-to-next-to-leading order (NNLO) accuracy with POWHEG NNLOPS [139]. The cross sections are calculated at the next higher order N^{3}LO with additional use of NLO electroweak corrections [55–58].
The ggF-production involves a quark loop as shown in Fig. 6.1 with main contribution from the top-quark with mass m_t. Higher order processes are important for this production mode, as they give sizeable corrections to the (differential) cross sections. The large-m_t approximation [140] is used for the differential cross section calculation. Loop contributions from other quarks, namely the b-quark, are neglected for this phase space.
To increase the available number of events in the phase spaces relevant for the $\tau_{\text{had}}\tau_{\text{had}}$ analysis while saving computing resources, a filter is applied during event generation. Both τ leptons originating from the Higgs boson are required to decay hadronically and to exceed a p_{T} of 30 GeV(20 GeV), respectively.

VBFH For simulation of the VBF-production mode, POWHEG-BOX v2 [141–144] with the MiNLO approach [145] is used for a next-to-leading-order (NLO) accuracy.
The parton shower model used on the generated events is provided by PYTHIA 8.186 [146]. To correct for higher order effects, the normalisation is taken from an approximate-NNLO QCD cross-section calculation with NLO electroweak corrections applied [59–61]. These EW corrections depend on the Higgs boson p_{T} and are up to 20% for $p_{\text{T}} > 300$ GeV. For the VBFH sample the same $\tau_{\text{had}}\tau_{\text{had}}$ filter strategy was used as in the ggF simulation.

VH The production with an associated Z- or W-boson is also simulated with POWHEG-BOX v2 in the MiNLO [145] approach and interfaced to the Pythia 8.186 parton shower model. The normalisation is at NNLO QCD precision with corrections from radiative electroweak effects at NLO applied [62–64].

ttH Finally, the signal contributions from associated production of the Higgs boson with a $t\bar{t}$-pair is calculated with QCD NLO accuracy using MG5_aMC@NLO v2.2.2 [147] and PYTHIA8 for parton showering.
The matrix element calculation uses the NNPDF30LO [148] PDF-set, while the NNPDF23LO [77] PDF set with A14 [149] tune is used for non-pertubative effects.

7.3 Background Modelling

The $\tau_{\text{had}}\tau_{\text{had}}$ channel background contributions are dominated by Z/γ^* and multijet processes as illustrated in Fig. 6.9. The estimation of background contributions with at least one true τ_{had} is based on simulation, while contributions from events with two fake τ_{had} candidates are estimated in a data-driven way.
In Sec. 7.4, light will be shed upon the determination of normalisations for each background component.

7.3.1 $Z/\gamma^* \rightarrow \tau\tau$-Background

As depicted in Fig. 6.9, the fully hadronic decay of $Z/\gamma^* \rightarrow \tau\tau$ is the dominant irreducible background contribution across all signal regions. The estimation of this component is based on Monte Carlo simulation, using a normalisation determined in a dedicated data-driven fitting procedure.
Both, electroweak- (EW) and QCD production modes are considered, with the EW component only amounting to a very small contribution given its small cross section.

The samples used in the analysis are summarised together with their cross sections in Appendix A. A dedicated study has been performed in context of the $\tau_{\text{had}}\tau_{\text{had}}$ analysis, comparing the modelling at preselection, VBF and Boosted regions between the POWHEG, MADGRAPH and SHERPA generators. The SHERPA 2.2.1 generator [73] provided the best overall modelling and largest number of events in the relevant phase spaces. Thus, it was chosen as default generator for Z/γ^* process modelling across all three channels [125].
Using the Comix [150] and OpenLoops [151] matrix-element generators, $Z/\gamma^* \rightarrow \ell\ell$ contributions are calculated up to two partons with a NLO and up to four partons at LO accuracy. These calculations are using the NNPDF30NNLO [148] PDF set merged to the dedicated SHERPA parton shower model [152] with the ME+PS@NLO prescription [153]. The cross sections are calculated at NNLO precision with the FEWZ3.1 software [154, 155] using CT10NNLO PDFs [156]. The small kinematic corrections on the differential cross sections are also considered through event-by-event weights.

The SHERPA internal modelling of the τ lepton decays was extensively tested in context of the τ identification measurements described in Sec. 5.4.3. During event generation a filtering is applied to enrich events in important $\tau_{\text{had}}\tau_{\text{had}}$ phase spaces, similar to the strategy deployed for signal samples. The τ pair from the Higgs boson decay is required to decay fully hadronically, each τ exceeding 30 GeV and 20 GeV respectively.

In order to validate the modelling of the $Z/\gamma^* \rightarrow \tau\tau$ processes, in particular the jet production rate associated with the Z boson, a study was performed in context of the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel [125]. Given that the tail of the Z boson mass peak extends up to a range of the expected Higgs boson mass of 125 GeV, a validation region close to the signal regions was investigated. As it proved difficult to construct a clean, signal depleted $Z/\gamma^* \rightarrow \tau\tau$ region in this regime, the validation was performed selecting $Z/\gamma^* \rightarrow \ell\ell$ events, with $\ell = e, \mu$. The validation regions are constructed as follows:

- two oppositely charged leptons of same flavour, with leading lepton $p_{\text{T}} > 40\,\text{GeV}$
- with an invariant mass $m_{\ell\ell} > 80\,\text{GeV}$,
- fulfilling $|\Delta\eta_{\ell\ell}| < 1.5$ and $\Delta R_{\ell\ell} < 2.5$
- at least one jet with $p_{\text{T}} > 70\,\text{GeV}$ is required.

From this preselection, VBF- and Boosted-like categories are constructed as follows

- VBF: sub-leading jet $p_{\text{T}} > 30\,\text{GeV}$, $m_{jj} > 400\,\text{GeV}$, $\Delta\eta_{jj} > 3$, $\eta_{\text{j}_1} \times \eta_{\text{j}_2} < 0$,
- Boosted: fail VBF-like selection and $p_{\text{T}}^{\ell_1} > 100\,\text{GeV}$.

With this selection a $Z \rightarrow \ell\ell$ purity of 99.0% for the VBF-like and 98.5% for the Boosted-like category is achieved.
Detailed modelling checks have been performed on these events, including:

- a reweighting of jet-multiplicity, m_{jj} and leading jet-p_{T} to observed data
- dependence on pile-up by splitting in high- and low-$\langle\mu\rangle$ regions
- the impact of the underlying event model by varying the fraction of multiple parton interactions

All tests confirmed that within the given theoretical uncertainties, described in Sec. 7.6, the default modelling is satisfying. A more detailed description of the procedures and results can be found in [125].

7.3.2 Background from Multijet Events

The background composition in Fig. 6.9 shows, that the second most prominent background component arises from events, where both τ_{had} candidates are reconstructed from misidentified jets. Events with misidentified jets can originate from gluon- or quark-initiated multijet processes but also from electroweak decays of the W/Z bosons associated with jets or the decay of top-quarks. In this analysis, different background modelling approaches are chosen dependent on the number of misidentified τ_{had} candidates. This is realised by using generator-based information about the simulated object type[1]. Events with at least one τ_{had} candidate geometrically matched within $\Delta R < 0.2$ to a simulated τ or electron are described by simulation, as specified in the next section. Other background contributions,

---

[1] Encoded in the pdg-ID: 13 (11) for τ leptons (electrons) [27]

with both τ_{had} misidentified from a jet will be labelled as **Fake background** throughout the rest of this thesis and are estimated in the data-driven way detailed in the following.

7.3.2.1 Shape Templates for the Fake Background

Several of the $\tau_{\text{had}}\tau_{\text{had}}$ event selection requirements can be altered to construct control regions (CR), which are orthogonal to the signal regions (SR) and where Fake background are enriched. From these CR, the kinematic shapes of the Fake background contribution is estimated and extrapolated to the SR using histogram templates and a fit of the normalisation to observed data. Hence, it is important that the CR correctly reflects the kinematics and topology of the SR.

The Preselection requirements investigated in this analysis to construct Fake background enriched CR are

- the product of the charges of the τ_{had} candidates is strictly negative and the absolute charge per τ_{had} is $|q| = 1$
- each τ_{had} candidate must have either one or three associated tracks (nTracks)
- both τ_{had} candidates fulfil the tight identification requirement.

The resulting CR definitions which were investigated are summarised in Tab. 7.2.

Region Name	Charge Product	nTracks	ID(τ_1)	Region Type
`OS`	$q(\tau_1) \times q(\tau_2) = -1$	1, 3	tight	SR
`SS`	$q(\tau_1) \times q(\tau_2) = 1$	1, 3	tight	CR
`nOS`	$q(\tau_1) \times q(\tau_2) \neq -1$	1, 2, 3	tight	CR
`OS aID med`	$q(\tau_1) \times q(\tau_2) = -1$	1, 3	medium && not tight	CR
`nOS aID med`	$q(\tau_1) \times q(\tau_2) \neq -1$	1, 2, 3	medium && not tight	CR
`OS aID loose`	$q(\tau_1) \times q(\tau_2) = -1$	1, 3	loose && tight	CR

Table 7.2: Summary of the Fake background control regions definitions investigated in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

`OS` Defines the default SR setting of the Preselection requirements in question for the analysis.

`SS` The same-sign CR `SS` includes only $\tau_{\text{had}}\tau_{\text{had}}$ pairs with the same charge q. This constellation occurs when two uncorrelated jets with same reconstructed charge are identified as τ_{had}-candidates, typical for non-resonant processes expected in Fake background. The `SS` serves as cross-check to the main CR, introduced in the following.

`nOS` For the not-opposite sign CR **`nOS`** the charge product requirement is inverted. Like the `SS`, this enriches the Fake background contribution but adds τ lepton candidates with only two associated tracks and a resulting charge of $q = 0$. This can occur if a track is outside the cone $\Delta R < 0.2$ used to associate tracks to τ_{had} candidates, characteristic for wider jets originating from gluons or quarks. The **`nOS`** is the main CR in the analysis, used to extract Fake background shape templates for all considered variables. The identification requirements are not altered with respect to the SR.

`OS aID med` is constructed by requiring the sub-leading τ_{had} candidate τ_1 to fulfil the medium but fail the tight identification requirement, thereby ensuring orthogonality to `OS`. This control region is used to correct the shape of the Fake background template, as described in Sec. 7.3.2.2

`nOS aID med` combines the `nOS` and `OS aID med` criteria forming another orthogonal CR and is used to validate the Fake background template reweighting.

`OS aID loose` follows the idea of the `OS aID med` but with a looser identification requirement on τ_1: the sub-leading τ_{had} candidate must fulfil the loose identification requirement but fail the tight one. As will be shown in Sec. 7.3.2.3, this region can be used to assign a systematic uncertainty on the reweighting method.

Cut	Data	$Z \to \tau\tau$	Others	$H \to \tau_{\text{had}}\tau_{\text{had}}$
`OS`	14871.0±121.9	10342.6±98.3	678.1±29.6	208.7±1.7
`SS`	3133.0±56.0	105.5±9.0	100.6±14.4	4.1±0.5
`nOS`	7582.0±87.1	743.5±24.3	206.4±19.1	22.7±1.3
`OS aID med`	5816.0±76.3	1720.1±35.4	238.2±16.2	48.6±1.9
`nOS aID med`	7503.0±86.6	290.4±13.7	147.7±16.5	8.4±0.8
`OS aID loose`	12562.0±112.1	2810.7±44.9	523.5±24.7	85.2±2.5

Table 7.3: Event yields in the CBA Preselection Fit category for all investigated CR. The **`nOS`** CR was chosen as default region to extract the Fake background template due to the comparably high statistics and low signal contamination.

In Tab. 7.3, the corresponding selection yields for each CR are presented. In principle it would be possible to further loosen the identification requirements for the anti-ID regions to enhance the statistical power of this region and to suppress contributions from true τ_{had}. However, studies in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel have revealed a strong kinematic dependency of the τ_{had} identification requirement when extrapolating the Fake background templates to the signal regions. This imposes the use of correction factors to avoid significant biases. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, deriving such factors requires an efficient quark/gluon separation as well as a large sample of simulated multijet events. A similar method has been studied as documented in [125] but introduced very large systematic uncertainties and was hence abandoned. The aforementioned `OS aID med` region avoids the issue of kinematic dependencies as it is chosen to be close to the signal regions but is statistically limited and has a large contamination of true τ_{had} of 30% as can be seen in Tab. 7.3. Furthermore, a relatively high signal contamination of 0.6% is observed in `OS aID med`, as shown in Tab. 7.3. For the latter reasons, the `OS aID med` region is not used to construct the nominal fake template. On all regions, the `nOS` provides the highest number of Fake background enriched events while only having comparably small contribution from true τ_{had} (10%) and a comparably small signal contamination (0.3%). This can be seen in Fig. 7.1, which shows the m_{MMC} and $\Delta\phi_{\tau\tau}$ distributions in the CBA `nOS` Preselection for Data and all simulated samples. The distribution for the simulated samples show that the true τ_{had} contribution in these regions are exclusively under the Z mass peak.

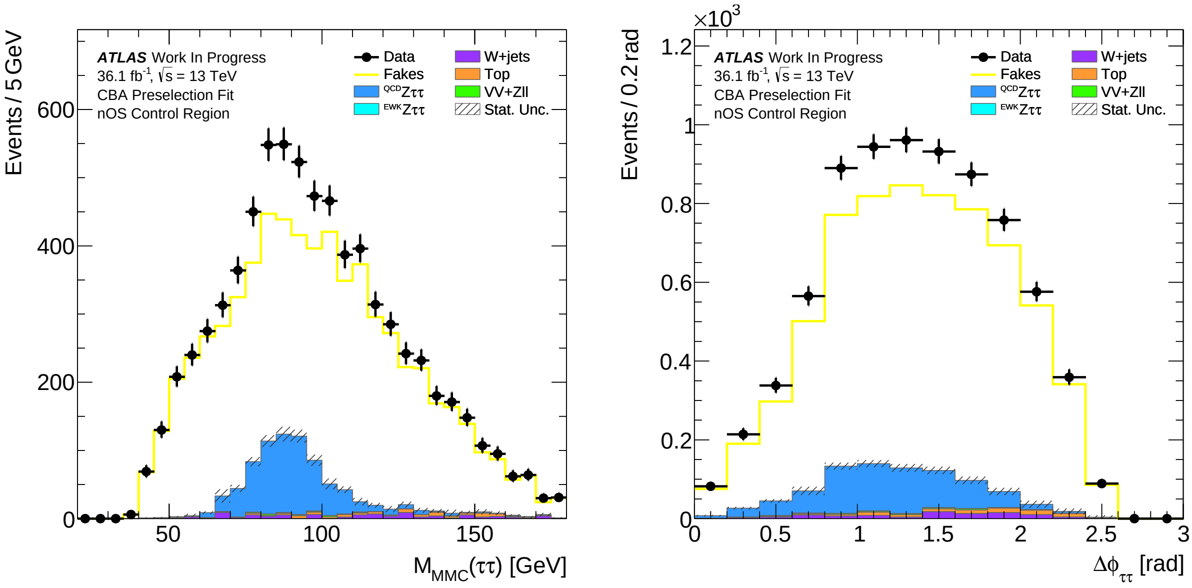


Figure 7.1: m_{MMC} (left) and $\Delta\phi_{\tau\tau}$ (right) distributions in the `nOS` CR in the CBA Preselection Fit category. True τ_{had} contributions estimated from simulation are subtracted from the Data distribution to obtain the `nOS` Fake background template (yellow).

The Fake background template is constructed separately for each signal region and variable by altering the selection requirements according to Tab. 7.2 in Data but still applying the respective kinematic and topological requirements. The resulting distribution is then corrected for true τ_{had} contamination by subtracting the simulated shapes from the `nOS` Data histogram.

This finally forms the nominal Fake background template, which is subsequently improved by a reweighting method introduced in the following.

7.3.2.2 Reweighting of the Fake Background Template

Small differences in kinematic distributions between `nOS` and `OS` were observed in Fake dominated, signal-depleted kinematic regions, like $m_{\text{MMC}} < 60\,\text{GeV}$. This can be seen in

Fig. 7.2 (left). These are corrected by a reweighting of the Fake background template. This difference originates from different quark/gluon compositions between `OS` and `nOS`. In a scenario, where the `nOS` exhibits a higher contribution of gluons splitting into a pair of quarks than `OS`, the kinematic properties of Fake background events could be biased. Studies of this effect proved difficult due to the lack of simulated multijet events in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ phase space.

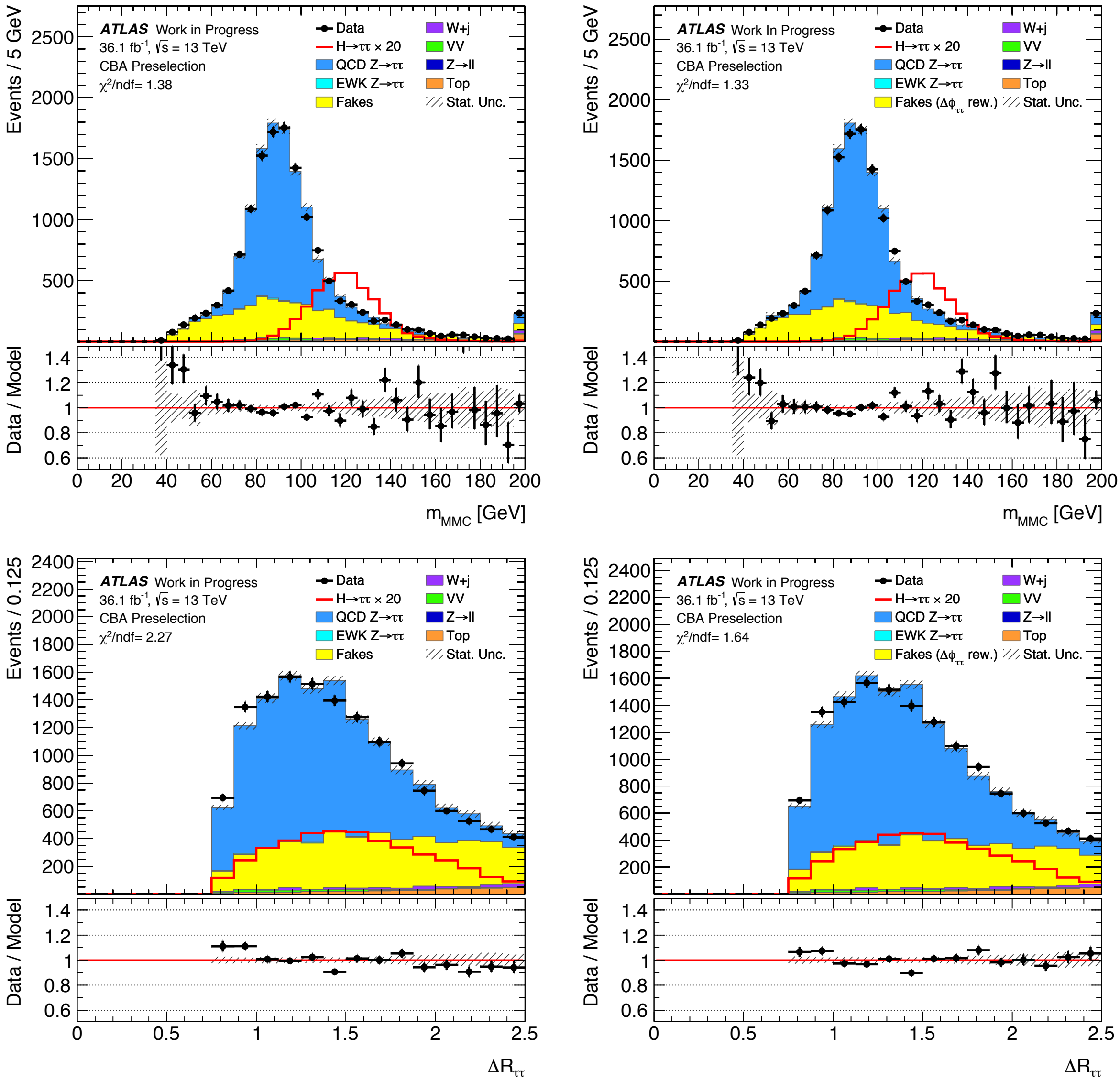


Figure 7.2: Distributions of the m_{MMC} and $\Delta R_{\tau\tau}$ in the CBA Preselection category. The background predictions are stacked and their combined statistical uncertainty is presented as shaded band. The left-hand plots illustrate the agreement between Data and event model for the nominal `nOS` Fake background template, while the right-hand side shows the modelling with $\Delta\phi_{\tau\tau}$ weights applied. The improvement of the reweighting is quantified by a χ^2_{red} test between Data and background prediction. Systematic uncertainties are not included. The correction is effective for $m_{\mathrm{MMC}} < 80\,\mathrm{GeV}$, as higher mass ranges are dominated by the $Z/\gamma^* \to \tau\tau$ background. The $\Delta R_{\tau\tau}$ distribution shows improvement for small and large values.

It was observed, that the shapes of the Fake background contribution in the `OS aID med` region reflects the distributions for kinematic properties of `OS` Fake background the best for the investigated CR. Hence, the `nOS` template is reweighted to this region, which has two advantages:

- the nominal Fake background template is corrected to a shape from a control region closer to the signal region
- and this allows for a streamlined treatment of systematic uncertainties on the Fake background template.

Out of all investigated variables, the mismodelling was found to be most emphasised in the angular distance $\Delta\phi_{\tau\tau}$ between the two τ_{had}. As mentioned before, this indicates a difference in the quark/gluon composition between `OS` and `nOS`. Jet pairs initiated from a gluon splitting into a pair of quarks are expected to be closer to each other than uncorrelated jets. This can bias the resulting $\Delta\phi_{\tau\tau}$ distribution. It was observed that the reweighting of the $\Delta\phi_{\tau\tau}$ `nOS` template effectively corrects the mismodelling in correlated variables. The kinematic selection of the CBA Preselection Loose region was chosen to perform this reweighting.

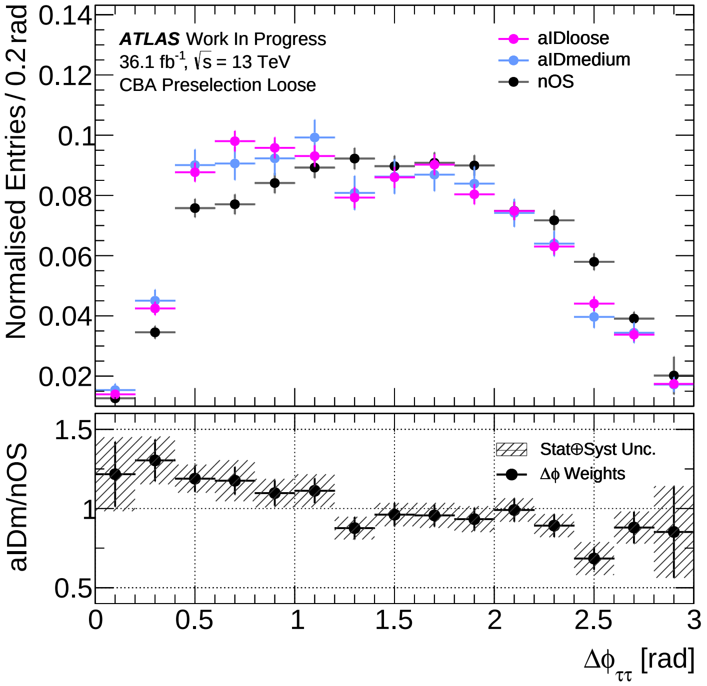


Figure 7.3: The top panel shows the $\Delta\phi_{\tau\tau}$ distributions after true τ_{had} subtraction in the `nOS`, `OS aID med` and `OS aID loose` CR in the CBA Preselection Loose category. The bin-by-bin difference between `nOS` and `OS aID med` shapes is shown as ratio in the bottom panel. The nominal values serve as reweighting factors to correct the `nOS` template to the `OS aID med` shape. The shaded uncertainty band is constructed comparing the `nOS` and `OS aID loose` distributions, symmetrising the bin-by-bin differences and adding them in quadrature to the statistical uncertainties of the reweighting factors.

In Fig. 7.3, the normalised Fake background templates for $\Delta\phi_{\tau\tau}$ are shown for the two CR, as well as for the `OS aID loose` region. The true τ_{had} contamination is subtracted from the Data distributions for both CR. For the `OS aID med`, identification and trigger correction factors are adapted according to the loosened identification criteria. The bottom panel shows the bin-by-bin ratio of the distributions in `OS aID med` and `nOS` and reveals discrepancies up to 20% for low $\Delta\phi_{\tau\tau} < 1.0$. Furthermore, regions with $\Delta\phi_{\tau\tau} > 2.0$ are populated at a higher rate for the `nOS` template, corresponding to an increased contribution also from back-to-back τ_{had}-candidates. The `OS aID loose` distribution is used to construct a systematic uncertainty, as will be discussed in Sec. 7.3.2.3.

The bin-by-bin ratio is applied as event weight to all `nOS` events when constructing the Fake background templates in each region. The $\Delta\phi_{\tau\tau}$ reweighting therefore is derived at the CBA Preselection Loose region and was found to correct shapes when propagated to the signal regions. Noticeably, the improvement was observed to a similar extent also in the MVA definition of each category.

The impact of the reweighting on the modelling is illustrated in Fig. 7.2 for the m_{MMC} mass and the angular distance $\Delta R_{\tau\tau}$. Being applied to the sub-dominant background contribution, the Fake background reweighting has a small impact on the bulk of the distributions. However, a clear improvement can be seen for MMC values smaller than 70 GeV as well as for the overall modelling of $\Delta R_{\tau\tau}$. This improvement was quantified using a χ^2_{red}-test between observed Data and predicted background contributions for all kinematic variables of the $\tau_{\text{had}}\tau_{\text{had}}$ system used in the analysis. In Tab. 7.4, the observed improvement is summarised for characteristic variables at CBA Preselection.

Variable	χ^2_{red} Data vs Background-Model	
	Nominal	$\Delta\phi_{\tau\tau}$-Reweighted
m_{MMC}	1.38	1.33
$\Delta\eta_{\tau\tau}$	1.61	1.47
$\Delta\phi_{\tau\tau}$	2.67	1.71
$\Delta R_{\tau\tau}$	2.27	1.64
p_{T}^{H}	1.26	1.05
$\mathcal{C}_{\tau\tau}(E_{\text{T}}^{\text{miss}})$	1.03	0.95
x_0	1.91	1.89
x_1	0.93	0.89
$p_T^{\tau_0}/p_T^{\tau_1}$	1.00	1.00
m_{vis}	1.21	0.87

Table 7.4: The modelling of characteristic kinematic variables used in the analysis was compared to Data between reweighted and nominal `nOS` Fake background template for the CBA Preselection category. The improvement is quantified with a χ^2_{red} test and observed across all variables. Systematic uncertainties are not included in this test.

7.3.2.3 Fake Background Template Uncertainties

Three separate modelling uncertainties for the Fake background templates are constructed, targeting specific assumptions of the presented methods. These include

- the statistical and systematic uncertainty of the $\Delta\phi_{\tau\tau}$-reweighting procedure,
- the uncertainty of the extrapolation from `nOS` to `OS`,
- the uncertainty arising from the subtraction of the simulated true τ_{had} contributions.

The input for the latter two uncertainty components is shown in Fig. 7.4.

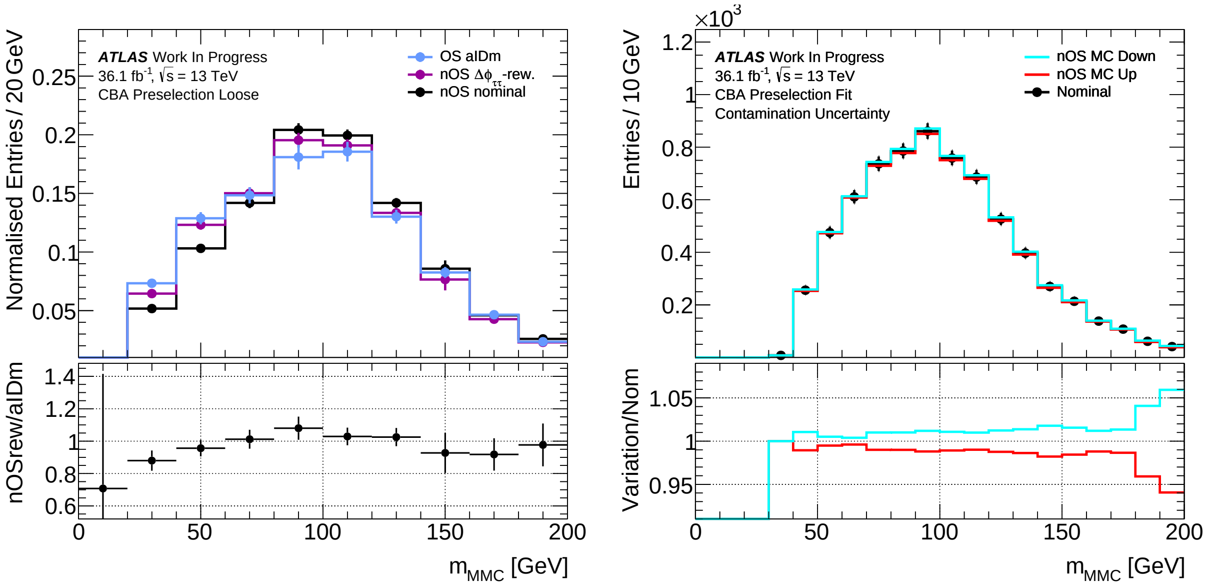


Figure 7.4: The left-hand figure shows the distributions used to extract the extrapolation uncertainty. The m_{MMC}-shape of the $\Delta\phi_{\tau\tau}$-reweighted `nOS` template is compared to the `OS aID med` distribution. Bin-by-bin deviations are translated to systematic uncertainties. This provides a closure to the reweighting method.
The right-hand figure shows the impact of the contamination uncertainty. The reweighted `nOS` template is shown after subtracting the simulated true τ_{had} contributions, varied within their statistical uncertainties.

Reweighting Uncertainty The correction factors of the reweighting procedure are derived from two statistically limited CR. To assign a statistical uncertainty, the bin-by-bin uncertainties of each Fake background template is propagated to the factors, as depicted in the ratio of Fig. 7.3.
Furthermore, a systematic uncertainty is constructed using a third CR, the `OS aID loose`. Bin-by-bin differences to the `OS aID med` template are interpreted as shape uncertainties and propagated accordingly. The shaded band in bottom panel of Fig. 7.3 shows the two uncertainty components added in quadrature. They amount up to $5-10\%$ depending on the signal region in a mass range $100\,\text{GeV} < m_{\text{MMC}} < 150\,\text{GeV}$. For the statistical analysis, both uncertainties are treated independently.

Extrapolation Uncertainty The constructed `nOS` Fake background template is taken as default shape to describe Fake backgrounds in `OS`. For this extrapolation, the transition of `nOS` to `OS`, a systematic uncertainty is constructed by comparing the shape of different regions for the m_{MMC} mass.
Using the same arguments as above, the m_{MMC} template extracted in the `OS aID med` region is resembling the shape of Fake backgrounds in `OS` better. Remaining differences to `nOS`-template after reweighting are taken as uncertainties. These differences arise from additional kinematic effects of the extrapolation, not covered by the $\Delta\phi_{\tau\tau}$ specific reweighting procedure.
Fig. 7.4 (left) presents the size of this uncertainty in the region used to extract the reweighting factors. For the signal regions, the so constructed variations are around 5% for the m_{MMC} mass range between 100 GeV and 150 GeV and up to 10% for $m_{\mathrm{MMC}} < 80$ GeV.

Contamination Uncertainty The Fake background templates are constructed by subtracting the true τ_{had} contribution estimated in simulation. All systematic reconstruction and identification uncertainties of the simulated objects discussed in Sec. 7.6 are propagated to the templates.
Additionally, a statistical uncertainty on this subtraction term is considered due to the limited amount of simulated events in the `nOS`-region. The simulated contribution at `nOS` Preselection is varied within its total statistical uncertainty. The resulting shape difference in the Fake background template for this statistical component is propagated to the statistical analysis and illustrated in Fig. 7.4 (right). This uncertainty amounts to $\sim 2.5\%$ for m_{MMC} smaller than 180 GeV.

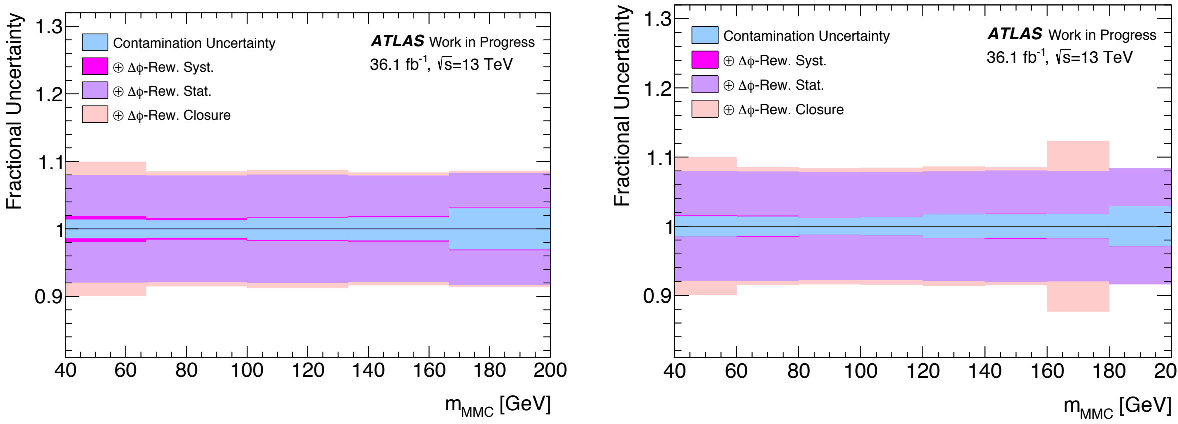


Figure 7.5: Fake background template uncertainty depicted as fraction of the nominal Fake background template for the CBA VBF/Boosted signal regions. The components are stacked and added in quadrature for illustration purposes only.

For illustration purposes the relative impact of all components on the final Fake background template is summarised in Fig. 7.5 for two CBA signal regions. The components are stacked in quadrature in these figures to highlight mass regions with large uncertainties. As can be seen from the fluctuations of these variations, the components describing statistical uncertainties are comparably flat, while the impact of shape uncertainties highly depend on the mass range. Overall, the Fake background template uncertainties are $\sim 10\%$ if added in quadrature. The MVA signal regions were investigated in the same manner and show a similar behaviour.

7.3.3 Other Background Contributions

For simplification, other background contributions from W + jets, top-quark and di-boson processes with at least one true τ_{had} in the event are combined as Others in the statistical analysis. The estimation of these processes is based on Monte Carlo simulation and normalisations are taken from theory predictions. All background generators are summarised in Tab. 7.1.

Contributions from W/Z bosons associated with jets, di-boson processes involving an electroweak boson pair VV and the electroweak VBF production of W/Z bosons are all estimated using the Sherpa 2.2.1 generator [73]. PDF sets and tunes for matrix element calculation and parton shower are identical to the settings used for the simulation of the $\mathrm{Z} \rightarrow \tau\tau$ process.

Events originating from $t\bar{t}$ processes are modelled using Powheg-Boxv2 [141–143, 157] with the CT10 PDF sets.
The Powheg-Box v1 [158, 159] generator is used to simulate electroweak single top quark production. A four-flavour scheme is used for the NLO calculations together with the PDF set CT10f4. For the t-channel decay modelling, MadSpin [160] is used.
All events involving top quarks use the Pythia 6.428 [161] with the CTEQ6L1 PDF set in the Perugia 2012 tune [162] to model parton shower, fragmentation and underlying event.

7.4 Normalisation Strategy

The initial normalisations of simulated samples are given by their respective theoretical cross sections and the integrated luminosity of the analysed dataset. The various object-specific scale factors for correcting simulation to the observed dataset impact the normalisation of each background. The processes constituting the Others background are normalised to the theory predictions. The $Z/\gamma^* \rightarrow \tau\tau$ background and the Fake background template normalisations are fitted to Data in two steps:

- A data-driven estimation at preselection is performed to facilitate modelling validations. The resulting factors are referred to as pre-fit normalisations.
- In the statistical analysis, both normalisations are free floating parameters, the post-fit normalisations.

The pre-fit normalisations are estimated using a maximum Likelihood fit, as detailed in the following. The results of this fit provide the starting values for the statistical analysis.

The complete $\tau_{\text{had}}\tau_{\text{had}}$ background model in the `OS` signal regions can be summarised by

$$\mathbf{n}(\text{Bkgd})^{\texttt{OS}} = r_{\text{QCD}} \cdot \underbrace{\left(\mathbf{n}(\text{Data})^{\texttt{nOS}} - r_{\text{Z}} \cdot \mathbf{n}(Z/\gamma^{*}\!\to\tau\tau)^{\texttt{nOS}} - \mathbf{n}(\text{Others})^{\texttt{nOS}}\right)^{\Delta\phi_{\tau\tau}}}_{\text{Reweighted Fakes Template}}$$

$$+\, r_{\text{Z}} \cdot \mathbf{n}(Z/\gamma^{*}\!\to\tau\tau)^{\texttt{OS}} + \mathbf{n}(\text{Others})^{\texttt{OS}} \tag{7.1}$$

where r_{QCD} and r_{Z} are the pre-fit normalisation factors[1]. $\mathbf{n} = (n_0, \ldots, n_N)$ describes the nominal yield per histogram bin of the $\Delta\eta_{\tau\tau}$ distribution.

The $\Delta\eta_{\tau\tau}$ variable was chosen to provide the input distribution for the normalisation fit. Due to the non-resonant topology of the Fake background contributions, the shape of the $\Delta\eta_{\tau\tau}$-distributions differs from the $Z/\gamma^{*}\!\to\tau\tau$ background. The shape differences in $Z/\gamma^{*}\!\to\tau\tau$ and Fake backgrounds allow for a simultaneous constraint of r_{QCD} and r_{Z}. This can be seen in Fig. 7.6, where the $\Delta\eta_{\tau\tau}$-distribution is illustrated after the normalisation fit in the Preselection Fit region. In particular $1.5 < \Delta\eta_{\tau\tau} < 2.0$ is dominated by Fake background contributions, which helps to reduce the uncertainty on r_{QCD}.

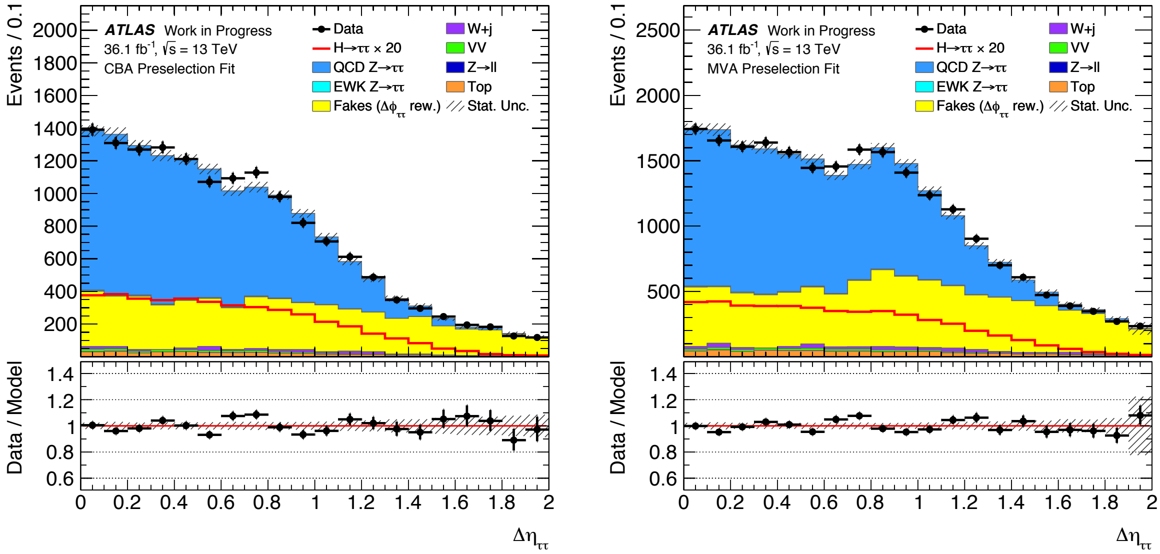


Figure 7.6: Distributions of the $\Delta\eta_{\tau\tau}$ variable after the fit in the CBA (left) and MVA (right) Preselection Fit region. Only statistical uncertainties are shown.

To perform this fit, the CBA/MVA Preselection Fit regions were chosen due to the minimal expected signal contamination of $\sim 1\%$. Signal events in Data hence have a negligible impact on the fit result.

The maximum Likelihood fit[2] is performed using the RooFit tool-kit [163]. The dependence of the subtracted $Z/\gamma^{*}\!\to\tau\tau$ contribution in the Fake background template on r_{Z} is considered in this fit. In Tab. 7.5, the resulting pre-fit normalisation factors are summarised. The differences between CBA and MVA Preselection Fit regions cause the

---

[1] r_{QCD} for the Fake background template and r_{Z} for the $Z/\gamma^{*}\!\to\tau\tau$ background, respectively

[2] For more details see Sec. 8.3.1

Fit Category	r_{QCD}	r_{Z}
CBA Preselection Fit	0.74 ± 0.03	0.89 ± 0.02
MVA Preselection Fit	0.67 ± 0.02	0.94 ± 0.02

Table 7.5: Pre-fit normalisations for the $Z/\gamma^{*}\rightarrow\tau\tau$ (r_{Z}) and Fake background template (r_{QCD}) separately for MVA and CBA Preselection Fit regions. These serve as starting point for the free floating normalisations in the statistical analysis. The differences between MVA and CBA normalisations are due to the looser requirements in MVA.

nominal values for r_{QCD} and r_{Z} to differ slightly. They are however compatible within statistical uncertainties. The resulting normalisation factors have been cross checked by using different variables for the normalisation fit, like the m_{MMC}, which yielded consistent results. Various studies have been performed to ensure stability of the fit results and are detailed in Appendix A. These checks included:

- repeating the fit independently for the 2015 and 2016 dataset
- varying the p_{T} threshold of the leading jet between $50 - 100$ GeV to exclude issues arising from the jet criterion in the trigger
- raising the p_{T} thresholds of the selected $\tau_{\text{had}}\tau_{\text{had}}$ pair to check for a possible impact of inefficient regions of the τ lepton triggers
- repeating the normalisation fit for varying criteria on the number of tracks associated to either τ_{had} candidate
- performing the background estimation using a different generator for the $Z/\gamma^{*}\rightarrow\tau\tau$ component
- using p_{T} dependent identification scale-factors for the $\tau_{\text{had}}\tau_{\text{had}}$ pair
- changing the identification criteria from tight to medium for both τ_{had}
- applying a lower cut on m_{MMC} between $50 - 80$ GeV to exclude problems in the fit arising from the low-mass region in the Fake background template
- removing the τ_{had} identification scale-factors, to investigate their impact on the normalisation.

Only the latter test had a significant impact on the observed $Z/\gamma^{*}\rightarrow\tau\tau$ normalisation. Without the τ_{had} identification factors shown and discussed in Sec. 5.4.3, the normalisation is compatible with 1.

7.5 Background Model Validation

For each signal region, the expected and observed event yields are listed together with their Poisson uncertainty in Tab. 7.6. Detailed tables showing the individual constituents of each event model component are attached in Appendix A. For these numbers, the previously presented normalisation factors have been applied independently for CBA and

MVA. A visualisation of the process composition can be found in Fig. 6.9 and Fig. 6.10.

	Region	$Z/\gamma^*\to\tau\tau$	Fakes	Others	$H\to\tau_{\mathrm{had}}\tau_{\mathrm{had}}$	Data	Total Background
CBA	Preselection Loose	10715.6±92.3	6665.5±89.0	1084.4±51.3	229.2±1.8	18622.0±136.5	18465.5±138.1
	Preselection Fit	9215.0±87.5	4978.5±70.1	678.1±29.6	208.7±1.7	14871.0±121.9	14871.6±115.9
	Preselection	9137.7±87.1	4247.4±65.5	630.8±28.5	202.8±1.7	14002.0±118.3	14015.9±112.7
	VBF	328.4±12.1	169.5±12.4	22.4±3.3	30.0±0.5	520.0±22.8	520.3±17.7
	VBF Low ΔR	158.7±7.0	19.3±5.1	5.5±1.1	16.6±0.3	179.0±13.4	183.4±8.8
	VBF High ΔR Tight	108.6±8.4	98.0±9.1	11.6±2.7	10.4±0.3	220.0±14.8	218.2±12.7
	VBF High ΔR Loose	61.0±5.2	52.3±6.8	5.3±1.6	3.0±0.2	121.0±11.0	118.6±8.7
	Boosted	6954.1±63.4	2213.4±50.7	472.3±23.4	142.3±1.5	9558.0±97.8	9639.8±84.5
	Boosted Tight	3617.3±39.2	297.3±25.8	137.2±7.4	76.0±1.1	4103.0±64.1	4051.7±47.5
	Boosted Loose	3336.9±49.8	1916.1±43.6	335.1±22.2	66.2±1.0	5455.0±73.9	5588.1±69.8
MVA	Preselection Fit	12523.0±101.5	8314.1±86.4	1136.0±62.2	249.5±1.9	21958.0±148.2	21973.1±147.1
	Preselection	12321.9±100.8	6828.0±79.0	1058.1±38.4	238.5±1.8	20246.0±142.3	20208.1±133.7
	Rest	2636.9±67.1	3344.3±51.9	276.3±25.6	36.4±0.8	6519.0±80.7	6257.6±88.6
	VBF	982.2±22.9	570.4±22.5	77.0±6.1	44.7±0.6	1516.0±38.9	1629.7±32.7
	Boosted	8702.8±71.6	2913.3±55.1	704.7±28.0	157.4±1.5	12211.0±110.5	12320.8±94.6

Table 7.6: Event yields for all categories used in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ analysis. The selection requirements are detailed in Sec. 6.5. The total background yield is calculated as the sum of $Z/\gamma^*\to\tau\tau$, Fake and Others backgrounds.

All investigated distributions exhibit good agreement between expected and observed Data within statistical and systematic uncertainties.
The variables used to define the MVA VBF and Boosted signal regions are depicted in Fig. 7.7 for the MVA Preselection region. The shaded uncertainty bands depicted in the upper and lower pads of Fig. 7.7 illustrate the statistical uncertainty added in quadrature to the systematic uncertainties, which are discussed in detail in Sec. 7.6. For these plots, object-reconstruction, -identification and -calibration uncertainties are considered, as well as theoretical uncertainties on the simulation of the background components. These bands also include the systematic uncertainty arising from the Fake background model.
All jet-related variables agree well within uncertainties, in particular also for the tails of the distributions. These distributions confirm the background modelling strategy and furthermore illustrate the signal region definitions described in Sec. 6.5.

Furthermore, Fig. 7.8 shows the $\Delta R_{\tau\tau}$ and p_{T}^{H} distributions which are essential to define signal regions in the CBA VBF and Boosted selection. The modelling agrees with the observed distribution within uncertainties, demonstrating that modelling and pre-fit normalisation techniques are valid in a more signal-like phase space.

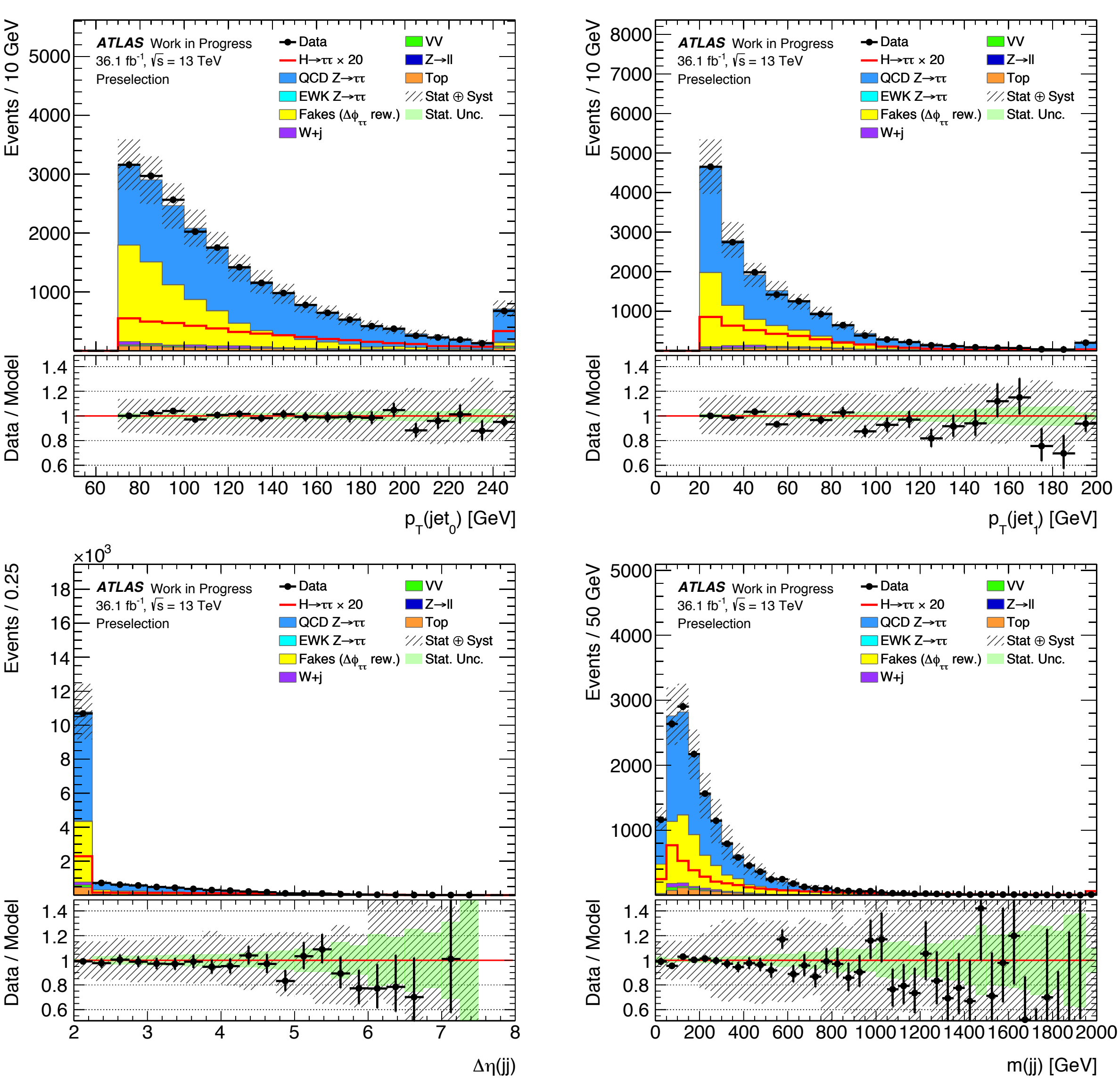


Figure 7.7: Distributions of the leading and sub-leading jet p_{T}, pseudo-rapidity gap $\Delta\eta_{jj}$ and di-jet mass m_{jj} in the MVA Preselection region. All distributions are depicted with statistical and systematic uncertainties and show good agreement between observed and expected Data. These variables are used to define the MVA signal regions.

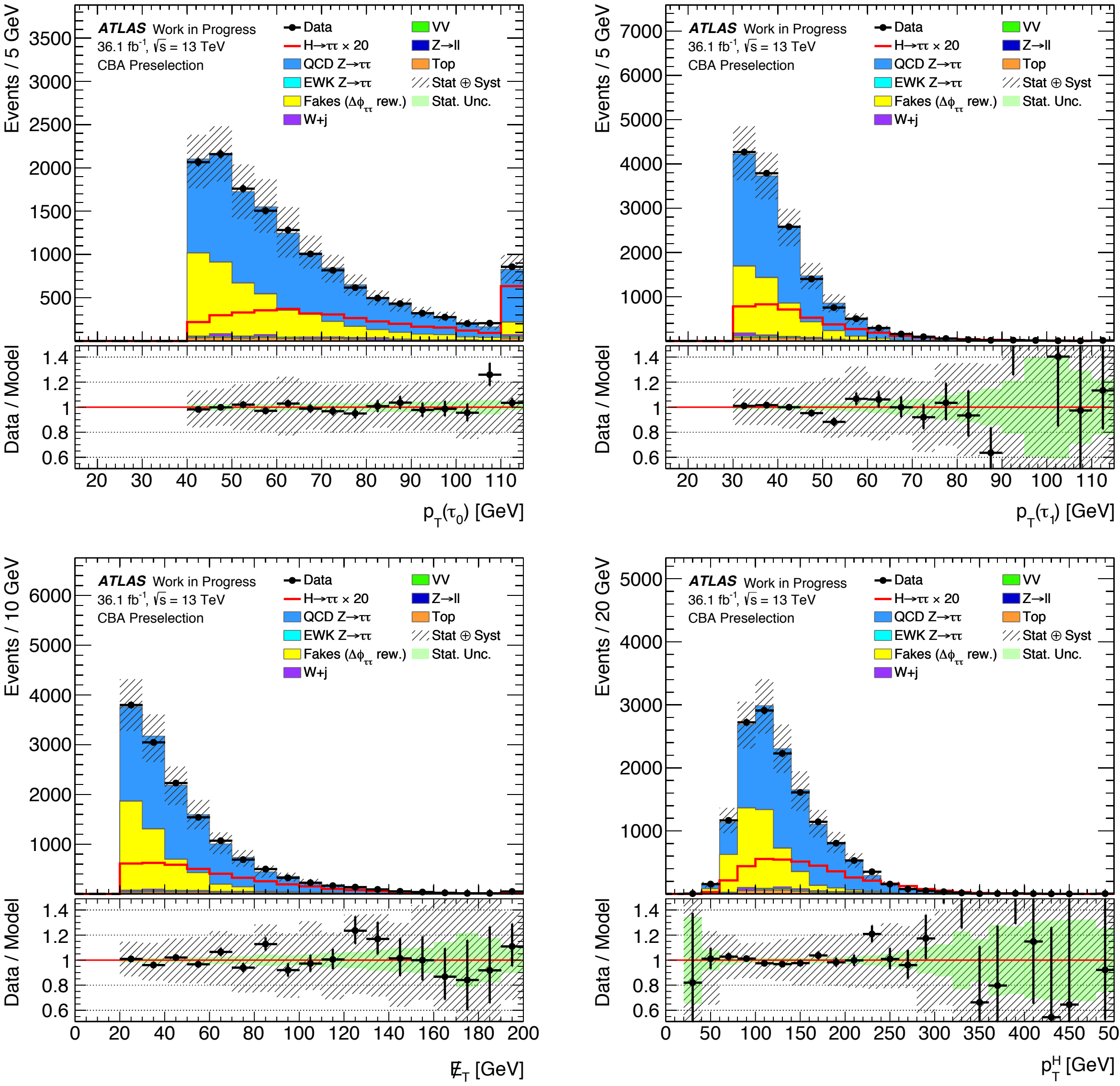


Figure 7.8: Distributions of the leading and sub-leading τ_{had} p_{T}, the $E_{\text{T}}^{\text{miss}}$ and p_{T}^{H} at CBA Preselection.

7.6 Systematic Uncertainties

The experimental techniques to reconstruct the event and theoretical predictions for the event model introduce systematic uncertainties which are summarised in the following. The impact of these uncertainties on the statistical analysis is discussed in Chapter (9). The effects are reflected in changes of the total event yields and changes in the shapes of the final discriminant distributions. For each considered simulated process, the corresponding quantity is varied within ±1 standard deviation of its nominal value. For experimental uncertainties, these values and uncertainties have been determined in dedicated auxillary measurements. The uncertainty arising from the data-driven Fake background template estimation have been detailed in Sec. 7.3.2. The impact of each uncertainty on the yield can be found together with a more thorough discussion of each component in [125].

7.6.1 Experimental Uncertainties

The various sources of experimental uncertainties arise from object reconstruction and identification techniques, trigger efficiencies and energy scale and resolution measurements.

τ_{had} Three uncertainties are considered for the energy calibration of the τ_{had} [11, 12]. The modelling of τ_{had} decays in simulation depends on pile-up, underlying event and detector response. An additional uncertainty is assigned for the impact of the GEANT4 detector model used in simulation. The energy calibration is estimated in a dedicated tag-and-probe in-situ measurement [12], where the simulated visible mass of a $Z/\gamma^* \to \tau\tau$ calibration dataset is fitted to observed data. These calibration uncertainties are used for both, the true τ_{had} and the small fraction of jets mis-identified as τ_{had} candidates in simulation.
The efficiency corrections for the τ_{had} identification discussed in Sec. 5.4.3 are based on tag-and-probe measurements, which account for differences between simulation and observed efficiencies. These depend on the number of tracks and the p_{T} of the true τ_{had} and are varied within their uncertainty for each event. The identification uncertainties are split in four components targeting the jet- and electron rejection in low- and high-p_{T} scenarios, respectively.
Finally, the τ_{had} trigger efficiencies have been measured and compared in data and simulation independently in the 2015 and 2016 dataset [121]. Four systematic and statistical uncertainties result for these corrections - the statistical uncertainty of the measurement arising from the limited number of events in 2015 data and simulation, the systematic uncertainty of the 2015 measurement and a total uncertainty for the 2016 correction.

Jets Various sources of uncertainties for the jet reconstruction and identification are considered. These depend on the p_{T} and η of the considered jet. Simulated events were compared to 2015 and 2016 data [164] and result in 32 jet uncertainties. These cover:

- the uncertainty on the jet energy scale measurement, performed in $Z+$jets events and di-jet data

- the dependence of the jet calibration relative to different η-regions of the detector
- the difference in detector response to quark/gluon-initiated jets and different jet compositions
- effects from the pile-up subtraction procedure, which corrects the jet energy reconstruction for additional contributions
- and the efficiency of the jet-vertex-tagger algorithm [128].

$E_{\mathrm{T}}^{\mathrm{miss}}$ The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction depends on the energy calibration of the jets, τ_{had}, electrons and muons in each event. When varying these calibrations in an event, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is also re-evaluated. Therefore, the $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainty arises from the $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions from energy deposits in the detector not associated to any physics object, called soft-track term. The measurement of this soft term uncertainty depends on the total p_{T} of all physics objects and was estimated in a dedicated measurement on 2015 and 2016 data [164]. In particular, the scale and resolution of these soft-track terms are varied within their uncertainties.

Pile-Up reweighting The correction of the simulated pile-up contributions mentioned in Chapter (6) introduces an uncertainty. In practice, this is done by shifting the simulated $\langle\mu\rangle$ distribution by a factor derived from the $\langle\mu\rangle$ distribution in the measured dataset. The uncertainty of this factor is propagated as event weight.

Luminosity The integrated luminosity of the analysed datasets is used to scale the simulated processes and has an uncertainty of $\sim 3.2\%$. Scans of the $x-y$ beam separation were performed in 2015 and 2016 to calibrate the luminosity scale [165]. The uncertainty on the combined luminosity measurement assumes uncorrelated uncertainties between the two measurements.

7.6.2 Theory Prediction Uncertainties

All simulated processes, signal and background, are subject to systematic uncertainties arising from theoretical assumptions, which can be grouped as follows:

- matrix element calculation uncertainties, depending in particular on the chosen QCD scale
- parton distribution function (PDF) uncertainties and the running strong coupling constant
- uncertainty on the Higgs boson branching ratio prediction $BR(H \to \tau\tau)$
- underlying event and Higgs boson p_{T} modelling.

These uncertainties affect the cross sections and modelling and are summarised in the following.

Signal Modelling The uncertainties on the signal modelling arise from the calculations introduced in Sec. 7.2 and detailed in [54]. The calculations are varied with respect to the assumed QCD scale, matching procedures, PDFs and parton shower model.
The Higgs boson production cross-sections have an uncertainty of ±4% for the ggF and $^{+0.4\%}_{-0.3\%}$ for the VBF production mode originating from the QCD scale uncertainty. PDFs assumed for the cross section calculation give an uncertainty of ±3.2% for ggF and ±2.1% for VBF. The $H \to \tau\tau$ branching ratio is varied within $^{+1.17\%}_{-1.16\%}$.

In practice, the various signal theory uncertainties are implemented as follows. The nominal weight for a given simulated event is replaced by a respective systematically varied one. Effects on the total normalisation are removed by using a modified overall normalisation of the signal samples to only account for changes in acceptance and shape.

$Z/\gamma^* \to \tau\tau$ Modelling Uncertainties on the modelling of the simulated $Z/\gamma^* \to \tau\tau$ process shapes arise from the scales assumed during simulation and the parton matching procedures which have been discussed in Sec. 7.3.1. Sherpa provides event weights to account for renormalisation and factorisation scale variations. The event weights result in varied shapes for the final discriminant distributions. Furthermore, the resummation and CKKW matching scale are varied in dedicated simulated samples. The uncertainty is extracted from ratio histograms binned in the Z boson p_{T} and jet multiplicities. The final variation is constructed by taking the maximum variation of any of the scale uncertainty per bin. Finally, a global 5% uncertainty is used for the inclusive cross section to account for the aforementioned uncertainties in the cross section calculation.

CHAPTER 8

Signal Extraction

In particle physics, searches for new particles follow a general strategy by testing the observed data against a hypothesis, assuming all events were produced from background processes. If a significant deviation from this background-only hypothesis is found, it is rejected in favour of a hypothesis including also signal processes.
This can be quantified by the signal strength μ, parameter of interest in this analysis. It is defined as the ratio of measured and the expected cross section times branching ratio:

$$\mu = \frac{(\sigma \times \mathrm{BR})^{\mathrm{obs}}}{(\sigma \times \mathrm{BR})^{\mathrm{SM}}} \quad . \tag{8.1}$$

Therefore, a μ of 0 corresponds to an absence of signal events, whereas a μ of 1 confirms the SM signal assumptions.
To perform such a hypothesis test, a robust statistical formalism is constructed based on a binned and profiled maximum Likelihood fit. Input to this fit are the final discriminants in the selected signal regions of both analysis approaches.

In the following chapter, the procedure to extract the signal strength and excess significance in the selected dataset is described. Selection criteria and event modelling are according to the preceding chapters.
Firstly, the construction and validation of the final discriminant for the cut-based and multivariate analysis approach are discussed in Sec. 8.1 and Sec. 8.2.2. Thereafter in Sec. 8.3, the statistical model and in particular the profile Likelihood function are presented. Various validation studies to ensure stability of the statistical procedure are summarised in Sec. 8.3.3.

8.1 Final Discriminant in the Cutbased Approach

In the cut-based approach, the invariant mass of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ pair reconstructed with the MMC-algorithm[1] is chosen as final discriminant and input to the Likelihood fit. The resonant characteristics of the H decays result in a peak around 125 GeV for signal

---

[1] For details of the algorithm compare to Sec. 5.6.

events. This implies separation power of the m_{MMC} between signal events and the bulk of background events produced in Z $\to \tau\tau$-decays; events produced from signal processes populate the region between $100\,\mathrm{GeV} < m_{\mathrm{MMC}} < 150\,\mathrm{GeV}$. The m_{MMC} is therefore sensitive to the Higgs boson mass within the given experimental resolution.
This can be seen in the m_{MMC} distributions for the selected dataset in the CBA signal regions depicted in Fig. 8.1 for the CBA VBF and in Fig. 8.2 for the CBA Boosted categories. Each background component considered in the analysis is presented, for the

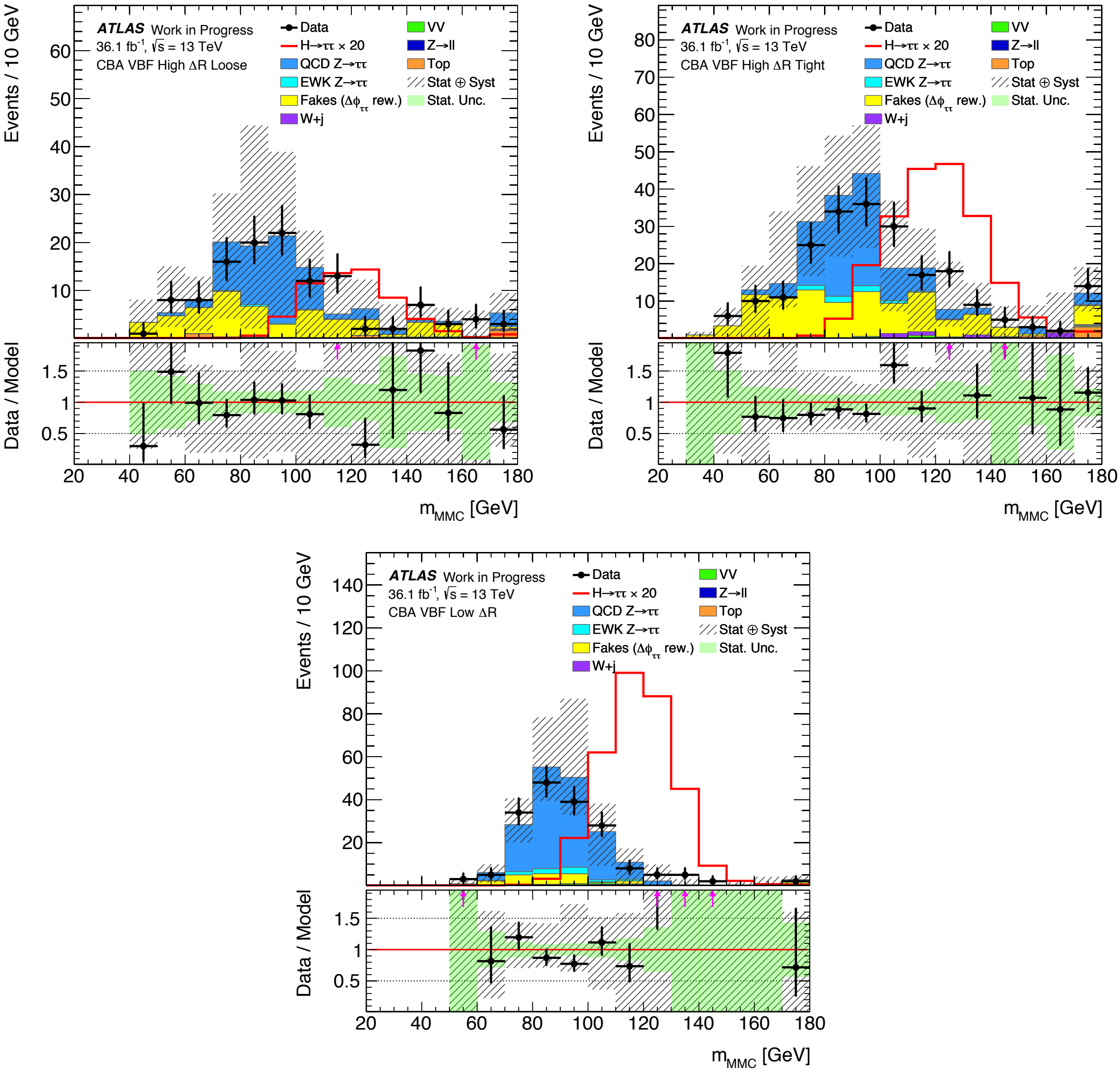


Figure 8.1: Pre-fit m_{MMC} distribution for the CBA VBF signal regions.

statistical analysis however, the processes with minor contributions are summarised as Others. Contributions from Fake background and Z $\to \tau\tau$ decays are scaled with pre-fit normalisation factors, as discussed in Sec. 7.4.

The binning of the distributions depends on the signal region due to different event yields of the selection. It was optimised in a way to ensure maximum separation power between

signal and combined background processes. The optimisation procedure is discussed in Sec. 8.3.5 and takes into account systematic uncertainties.
The distributions motivate the definition of the side-band regions $m_{\mathrm{MMC}} < 100\,\mathrm{GeV}$ and $150\,\mathrm{GeV} < m_{\mathrm{MMC}} < 200\,\mathrm{GeV}$, where only small signal contributions are expected. In these side-bands the free floating normalisations of Z $\rightarrow \tau\tau$ and Fake backgrounds are constrained during the fit in the statistical analysis. Within the given uncertainties, the side-bands reflect the adequate performance of the background model. The bin-by-bin uncertainties range between $50-80\%$ in the CBA VBF and $10-25\%$ in the CBA Boosted categories in the signal-sensitive mass regions. A dedicated treatment of the systematic uncertainties is performed before the final measurement, to reduce the impact of bin-by-bin migration and other statistical effects. The distributions in Fig. 8.1 furthermore show that the main background contribution in regions under the signal-peak arise dominantly from $Z/\gamma^{*}\rightarrow\tau\tau$ and multijet events.

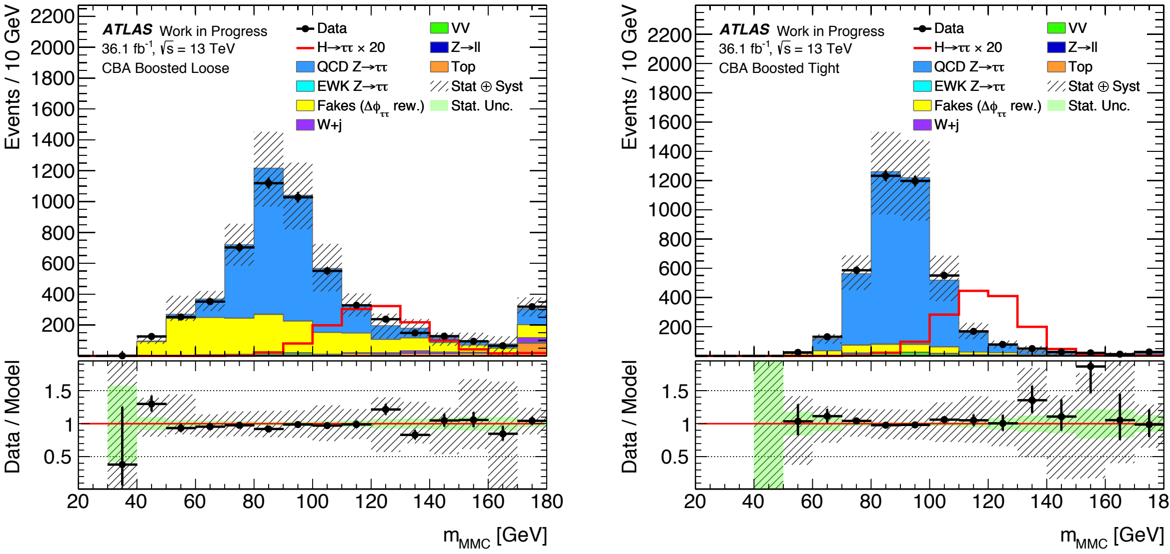


Figure 8.2: Pre-fit m_{MMC} distribution for the CBA Boosted signal regions.

These binned distributions form the input to the profile Likelihood fit, which is explained in detail in Sec. 8.3.1.

8.2 Multivariate Classification with Boosted Decision Trees

The main focus of this thesis is the multivariate approach to construct a final discriminant. This discriminant is defined by the output of the Boosted Decision Tree (BDT) algorithm. On top of separation characteristics of various signal sensitive variables like the MMC, the BDT exploits correlations between these kinematic observables to separate the events into signal and background contributions.

The BDT-algorithm is an example of supervised machine learning algorithms, which use example inputs to acquire insights on general rules how to map independent input data to a desired output. In particle physics this task typically involves labelling events as signal- or background-like, a so called binary classification problem. The algorithm uses characteristics of a set of labelled events to conclude on the classification of an independent set of events.

A variety of machine learning algorithms were studied in context of particle physics analyses. Besides BDTs, neural networks proved to be successful for data analysis tasks [166]. Neural networks use a multitude of connected nodes in hidden layers to transform input into a response, usable to classify events. The non-linear output and complex connection of nodes make the output of neural networks difficult to interpret.

BDTs on the other hand are comparably straightforward to interpret and offer a similar performance [167]. They are based on Decision Trees, which subsequently partition the available N-dimensional variable space[1] in hypercubes to enhance the purity of signal- or background events. The aim of this partitioning is to find optimal thresholds to distinguish signal-enhanced kinematic regions in the example input, thereby maximising the sensitivity of an analysis. The error of the classification in the independent input data is dictated by the choice of these thresholds.

In the following sections, the construction of BDTs in context of the $H \to \tau_{\text{had}}\tau_{\text{had}}$ searches is presented. Due to the limited available number of events in the MVA signal regions, special emphasis is put on the optimisation of the BDTs. The BDT-implementation of the Tmva tool-kit was used for this analysis [168]. The final input to the measurement of the $H \to \tau_{\text{had}}\tau_{\text{had}}$ signal strength are the outputs of the BDTs in all signal regions, the BDT score distributions, which are used in the profile likelihood fit.

8.2.1 Boosted Decision Trees BDT

8.2.1.1 Decision Trees

The basis of any BDT is formed by decision trees, outlined in the following. Starting from a set of events with known labels (i.e. “Signal” or “Background”), signal events are normalised so that signal and background contributions each have the same total weight. In the first step of the decision tree construction, the root node, all events are sorted in all input variables x_i with $i = 0, \ldots, N$ provided to the algorithm. The threshold resulting in the best separation between signal events s and background events b is determined for each variable. A metric to quantify this separation is the Gini index g, defined as:

$$\begin{aligned} g = p_s(1-p_s) &= \frac{2sb}{(s+b)^2} \\ \text{with} \qquad p_s &= \frac{s}{s+b} \end{aligned} \tag{8.2}$$

where p_s is the signal purity for each of the resulting event sets after splitting by the threshold in question. The best separation is achieved if g is minimal, indicating a pure

---

[1] For N input variables.

selection of either signal or background events. The variable yielding the highest separation is then chosen to define the first splitting of events. This split yields two daughter nodes, i.e. events passing or failing the selected threshold. This is visualised in Fig. 8.3, in which

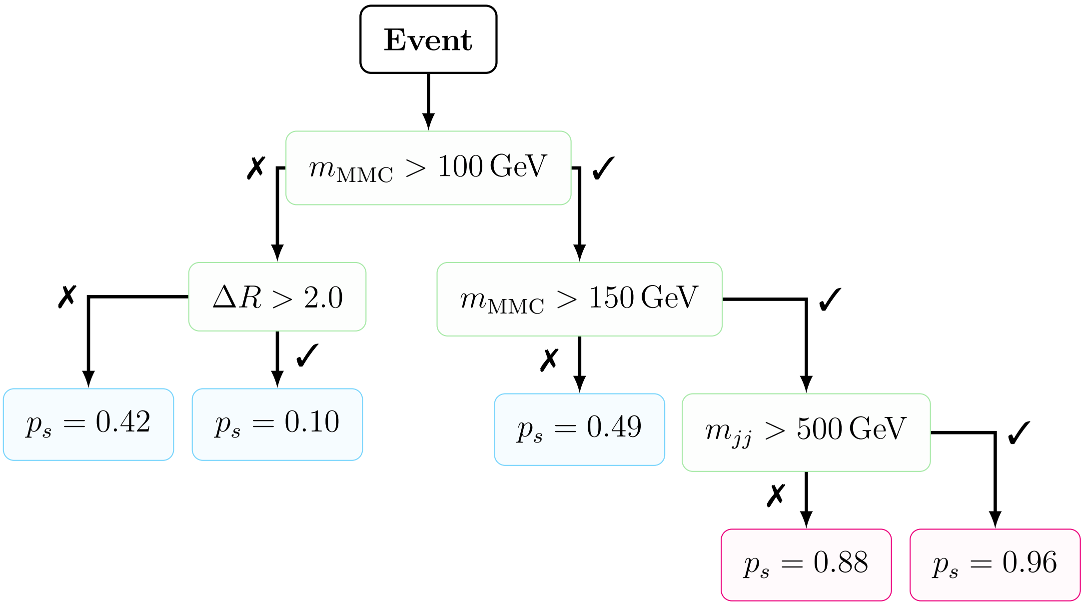


Figure 8.3: Sketch of a Decision Tree for the MVA VBF category. The green boxes represent the splitting nodes. After each split, the signal purity p_s is evaluated. Once the stopping criteria are met, the leaves are labelled as Signal (magenta) or Background (blue) depending on their final signal purity.

the initial set is split into two subsets of events, the nodes, depending on whether an event fulfils $m_{\mathrm{MMC}} > 100\,\mathrm{GeV}$. In each iteration, the separation is maximised using all variables, in particular allowing for multiple usage of the same variable as indicated for m_{MMC} on the right-hand side of Fig. 8.3.

For each of the resulting nodes the procedure is repeated recursively until one of the following stopping criteria is reached:

- a maximum depth of the decision tree, placing an upper limit on the number of subsequent nodes
- a minimum percentage of events in the resulting nodes, typically between $1 - 10\%$.

The final nodes after the splitting stops are called leaves. Each leaf is assigned a label according to the purity p_s of events in the leaf, where leaves with $p_s > 0.5$ are dominated by signal events. This label is ± 1 for signal and background respectively. Thus, events are classified into signal and background components in the resulting variable-space hypercubes defined by the splitting thresholds in each node up to the final leaf. The tree has learned how to distinguish signal and background events, as it was trained on a labelled input.

The principle of splitting the variable space is also used in simple cut-based approaches, albeit only one hypercube is formed to define signal and background-like event categories. Decision Trees could in principle partition the available variable space in any number of hypercubes, taking into account non-linear and possibly hidden correlations.

The limitation of decision trees in context of particle physics data analyses lies in the typically limited amount of input events.

8.2.1.2 Overfitting

Statistical fluctuations in the input events can lead to non-general decision trees with a high classification error on an independent set of events, the testing set. This is called overfitting (or overtraining), a situation where a trained decision tree is sensitive to statistical noise. On top of the underlying characteristics of signal and background events, nodes and thresholds in such a decision tree depend on specific fluctuations of the training set. These small variations can lead to fundamentally different tree structures, an undesirable effect as the classification on independent events will get unreliable.

8.2.1.3 Boosting of Decision Trees

Because of these shortcomings, the constructed decision tree is a so called weak learner. Boosting is a method to construct a stable, general strong learner with a smaller misclassification rate on the testing set by combining an ensemble of decision trees. For Decision Trees such an ensemble is called decision forest and consists of trees with subsequently weighted events. Misclassified events receive a common, higher event weight in the next training iteration derived from the misclassification rate of the previous tree. The factor modifying the event weight is called the boost weight. The new, reweighted training set is renormalised to the initial total event weight.

By averaging the decisions of all trees in the resulting forest, a continuous boosted event classification score between $[-1, 1]$ is constructed. A large value for any given event suggests signal-like event characteristics, as the majority of decision trees labelled the event as signal.

This boosting can be regularised by different algorithms to avoid overfitting as described previously. A popular algorithm for this boosting is AdaptiveBoost, where the boost weight is modified by a parameter, the learning rate [168].

The GradientBoost algorithm takes a similar approach using a learning rate called Shrinkage. It differs however in the construction of the misclassification rate used to derive the boost weight. These rates can be parametrised by loss functions, which quantify the deviation from the true event label to the classified one. AdaptiveBoost relies on an exponential loss function while the current Tmva implementation of GradientBoost uses a binomial log-likelihood loss function [168]. A minimisation of the loss function corresponds to finding the set of averaging parameters of the boosting algorithm for which the decision forest response is closest to the true label of the training event. For GradientBoost, this minimisation is based on the gradient of the loss function [168]. This generalised approach is expected to have a more stable performance in situations with limited number of events and to be less sensitive to outliers in input data. As GradientBoost promises to improve the accuracy in noisy conditions, it was chosen as boosting algorithm in this analysis.

The technical implementation of GradientBoost BDTs in Tmva offers several parameters to regularise the algorithm. These are defining characteristics of the decision tree construction and boosting procedure, namely:

- number of trees in the forest, `nTrees`,
- `Shrinkage` of the GradientBoost algorithm,
- the stopping criteria maximum depth of each decision tree `MaxDepth` and
- minimum number of signal events in a leaf `MinNodeSize`
- and the granularity used to scan each variable during node splitting to determine the best separation threshold. Ideally, this scan would be based on a continuous distribution of the input variable but due to computing expenses, the variable range is binned into `nCuts` parts.

The mostly empirical procedure to determine the parameter setting yielding the best level of separation performance and model generalisation is detailed in Sec. 8.2.2.3.

8.2.2 BDT Training Strategy and Optimisation

In the next sections, several strategies to construct stable and well-performing BDTs for the $\tau_{\text{had}}\tau_{\text{had}}$ MVA are presented. difficulties for the training of BDTs arise from the limited amount of available events. The BDTs trained on a small number of events tend to be overfitted, as statistical fluctuations of the training set become important during the node splitting. By relaxing selection criteria of MVA compared to CBA as introduced in Chapter (6), additional events are available for the BDT-training, thereby improving the signal and background separation power and stability of the BDT.

Two separate BDTs are constructed for the VBF and Boosted topologies introduced earlier. By this, specific kinematic characteristics can be exploited to optimise separation in each phase space.

Only VBFH signal events are considered for the MVA VBF BDT training to exploit the characteristic topology with two forward jets enclosing the τ_{had} pair. The MVA Boosted training uses both, the VBFH and ggH production modes in order to account for signal events failing the VBF requirements and thereby increasing the number of available training events. For the statistical analysis, these BDTs are evaluated on all signal events. With this strategy, category specific features are targeted effectively by the BDT, an example being the kinematics of the di-jet system in the MVA VBF category. The inclusion of the VBFH process in the boosted BDT is motivated by the signal composition presented in Fig. 6.10 and was observed to improve the sensitivity in this category.

For both categories, contributions from $Z/\gamma^* \rightarrow \tau\tau$ and the Fake background were considered for the training as detailed in Sec. 8.2.2.2.

8.2.2.1 Cross Evaluation

The overfitting problem of BDT algorithms is related to another practical issue concerning the training: a BDT should not be evaluated on events which were used to train it. This would lead to a bias towards training events and a different performance on an independent

event set.
In order to retain the information from the full set of events while avoiding this bias, a cross evaluation scheme was deployed. The available events in each region are split according to their event numbers in an even and odd set. For simulated (recorded) events, the event number is subsequently increased per simulated (recorded) event. As these events are uncorrelated, the splitting criterion is unbiased. This creates two statistically independent set of events.

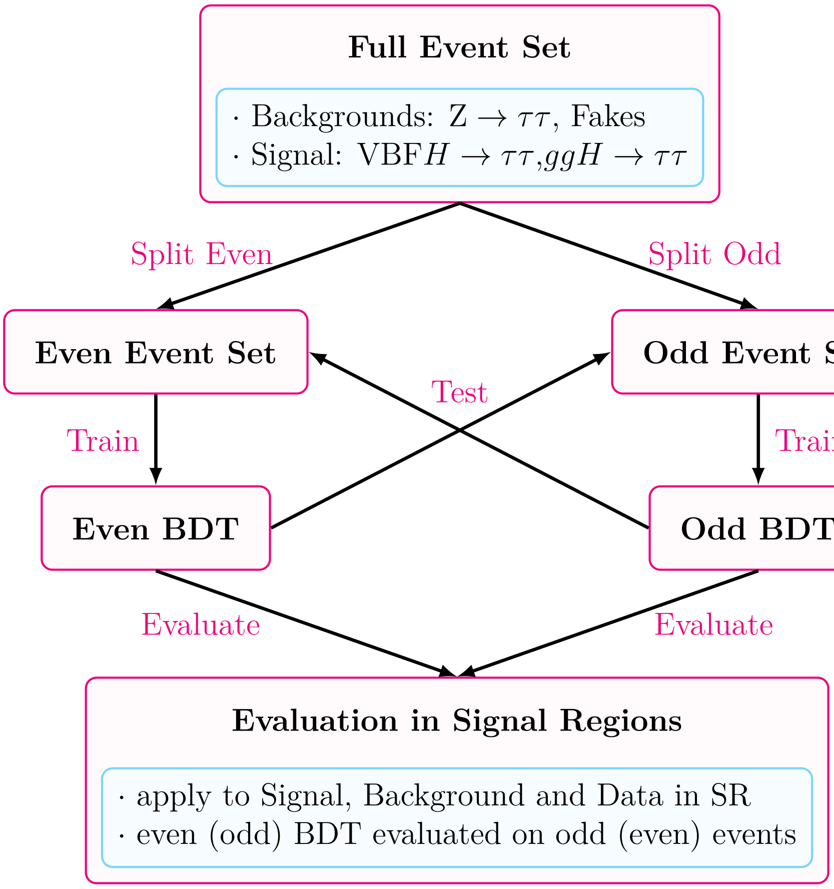


Figure 8.4: Cross Evaluation scheme as used in the $\tau_{\text{had}}\tau_{\text{had}}$ analysis, for both the MVA VBF and MVA Boosted category independently. The strategy retains the information of all available events and simultaneously provides a handle on the validation of the performance of a given BDT.

Two separate BDTs are then constructed for both categories by independently training on each event set. To investigate whether a given BDT is overfitted, the performance of a BDT can be tested on the respective set which was not used for the training, forming the aforementioned testing set. This scheme is sketched in Fig. 8.4.

For evaluation of the trained and tested BDTs on the events used in the statistical analysis, both are used: the even (odd) BDT on all of the odd (even) events in the signal regions. This strategy has the advantage that no information is discarded in training, all available events in each signal region can be used to train the BDTs.

8.2.2.2 BDT Training Input

The variables used as input to the BDT algorithm determine the separation power and depend on the considered category. The initial list of variables is listed in Tab. 8.1. This list was determined and harmonised between all three channels of the $H \to \tau\tau$ analysis and serves as starting point for the final input variable selection described in Sec. 8.2.2.3.

Variable	Description	VBF	Boost
m_{MMC}	MMC-based $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$-mass	✓	✓
p_{T}^{H}	p_{T} of the Higgs candidate.	✓	✓
$\Delta R_{\tau\tau}$	Angular distance between the two τ_{had}	✓	✓
$\Delta\eta_{\tau\tau}$	Pseudorapidity gap between the two τ_{had}	✓	✓
$p_T(\tau_0, \tau_1, j_0, j_1, \mathrm{MET})$	Total transverse momentum	✓	
$\mathcal{C}_{\tau\tau}(E_{\mathrm{T}}^{\mathrm{miss}})$	$E_{\mathrm{T}}^{\mathrm{miss}}$ centrality	✓	✗
$\mathcal{C}_{jj}(\tau_0)$	Leading τ_{had} centrality	✓	
$\mathcal{C}_{jj}(\tau_1)$	Subleading τ_{had} centrality	✓	
m_{jj}	Invariant mass of the two leading jets	✓	
m_{vis}	Visible mass of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$	✗	✓
$\Delta\eta_{jj}$	Pseudorapidity gap between the two leading jets	✗	
$\eta_{j_0} \times \eta_{j_1}$	Product of the jet's pseudorapidity	✗	
$p_T^{\tau_0}/p_T^{\tau_1}$	Ratio of the transverse momenta of the two τ_{had}	✗	
$\Delta\phi(H, j_1)$	Azimuthal distance between the Higgs and the sub-leading jet.	✗	
$m_T(\tau_0, \mathrm{MET})$	Transverse mass of the leading τ-candidates and $E_{\mathrm{T}}^{\mathrm{miss}}$	✗	
$m_T(\tau_1, \mathrm{MET})$	Transverse mass of the subleading τ-candidates and $E_{\mathrm{T}}^{\mathrm{miss}}$	✗	
$\Delta p_T(\tau_0,\tau_1)/\sum_i p_T^i$	Ratio of difference and sum of τ_{had} momenta	✗	
$\Delta\phi_{\tau\tau}$	Azimuthal distance between the two τ_{had}	✗	✗
$m^{\mathrm{vis}}_{\tau_0,\tau_1,j_0}$	Visible mass of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and the leading jet.	✗	✗
$\mathrm{MET}/p_T(\tau_0)$	Ratio of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the p_{T} of the leading τ_{had}	✗	✗
$\mathrm{MET}/p_T(\tau_1)$	Ratio of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the p_{T} of the leading τ_{had}	✗	✗
$p_T(\tau_0, \tau_1)$	p_{T} of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ system	✗	✗
$p_T(\tau_0, \tau_1, j_0, \mathrm{MET})$	p_{T} of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ system, the leading jet and $E_{\mathrm{T}}^{\mathrm{miss}}$		✗
$p_T(\tau_0) + p_T(\tau_1)$	Scalar sum of the τ_{had} transverse momenta		✗

Table 8.1: Full list of input variables considered for the training of the BDTs. The final selection for the categories in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel is determined in a dedicated study. ✗ marks variables initially considered per category and ✓ the final selected set.

The modelling of selected variables used for training the final BDT is presented in Fig. 8.5 for MVA VBF and Fig. 8.6. The complete set of figures can be found in Appendix B. The good modelling of all variables validates the event model for the MVA and promises an equally satisfying output modelling of the BDT score.

As the BDT algorithm scans the optimal threshold of cuts with a limited granularity

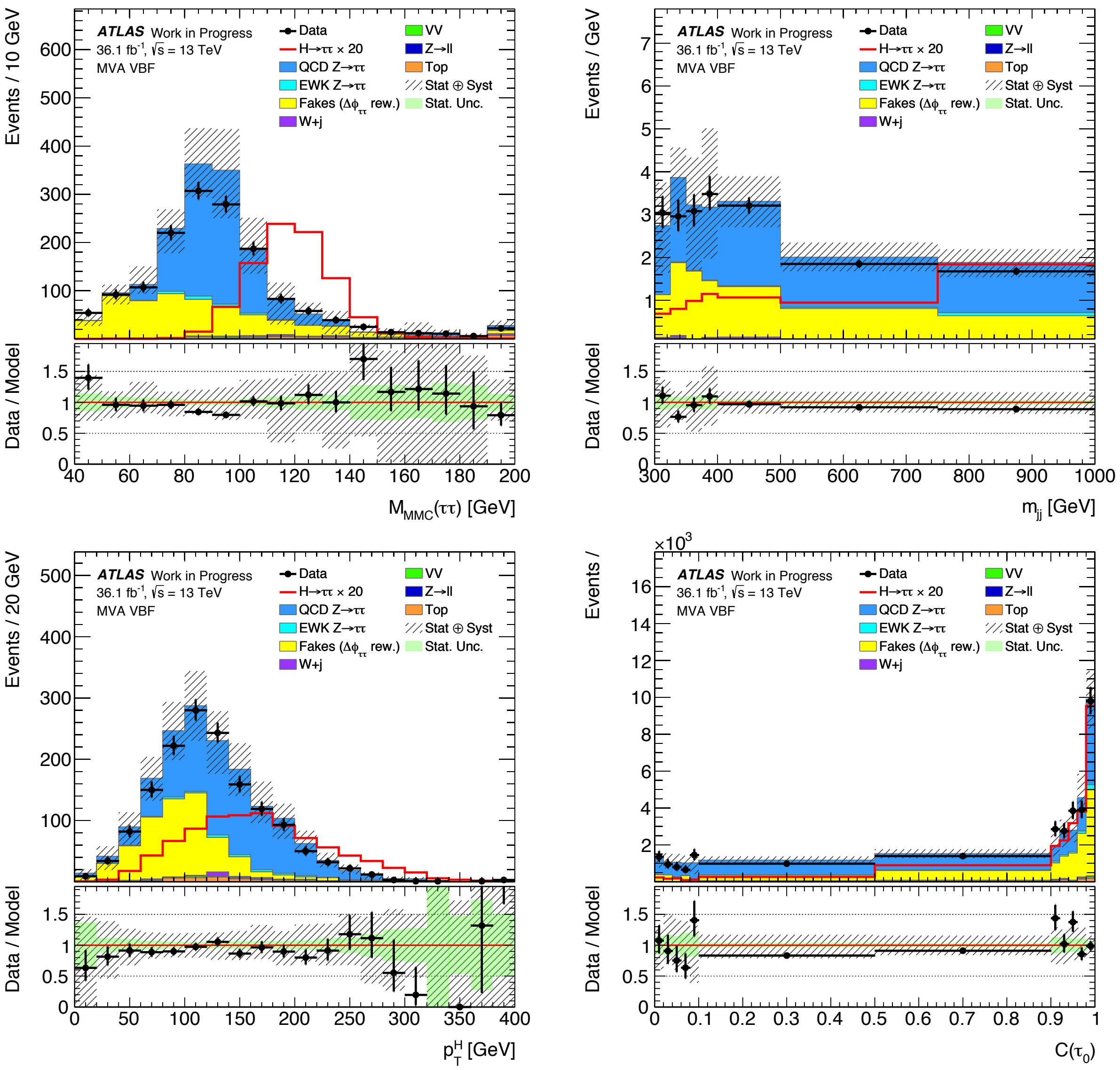


Figure 8.5: Modelling of selected input variables used in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ VBF BDT. The uncertainty band includes statistical and systematic uncertainties.

`nCuts`, statistical outliers in the distributions were excluded from the training. The ranges of all variables in Tab. 8.1 have been restricted. Most notably the invariant mass of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ pair is required to be within $0\,\mathrm{GeV} < m_{\mathrm{MMC}} < 300\,\mathrm{GeV}$, excluding outliers up to 2 TeV. While this has no impact on the resulting separation power, the range restriction improves the agreement for the node splitting criteria between even and odd BDTs.

The modelling of correlations between input variables is demonstrated for selected variables in Fig. 8.11-8.12 and for the complete final variable set in Appendix B. Correlations are compared between Data and all background processes considered in the event model. A Kolmogorov-Smirnov test [169] was used to quantify the agreement between the presented correlations and verifies the satisfying modelling. The test is performed by removing signal sensitive regions.

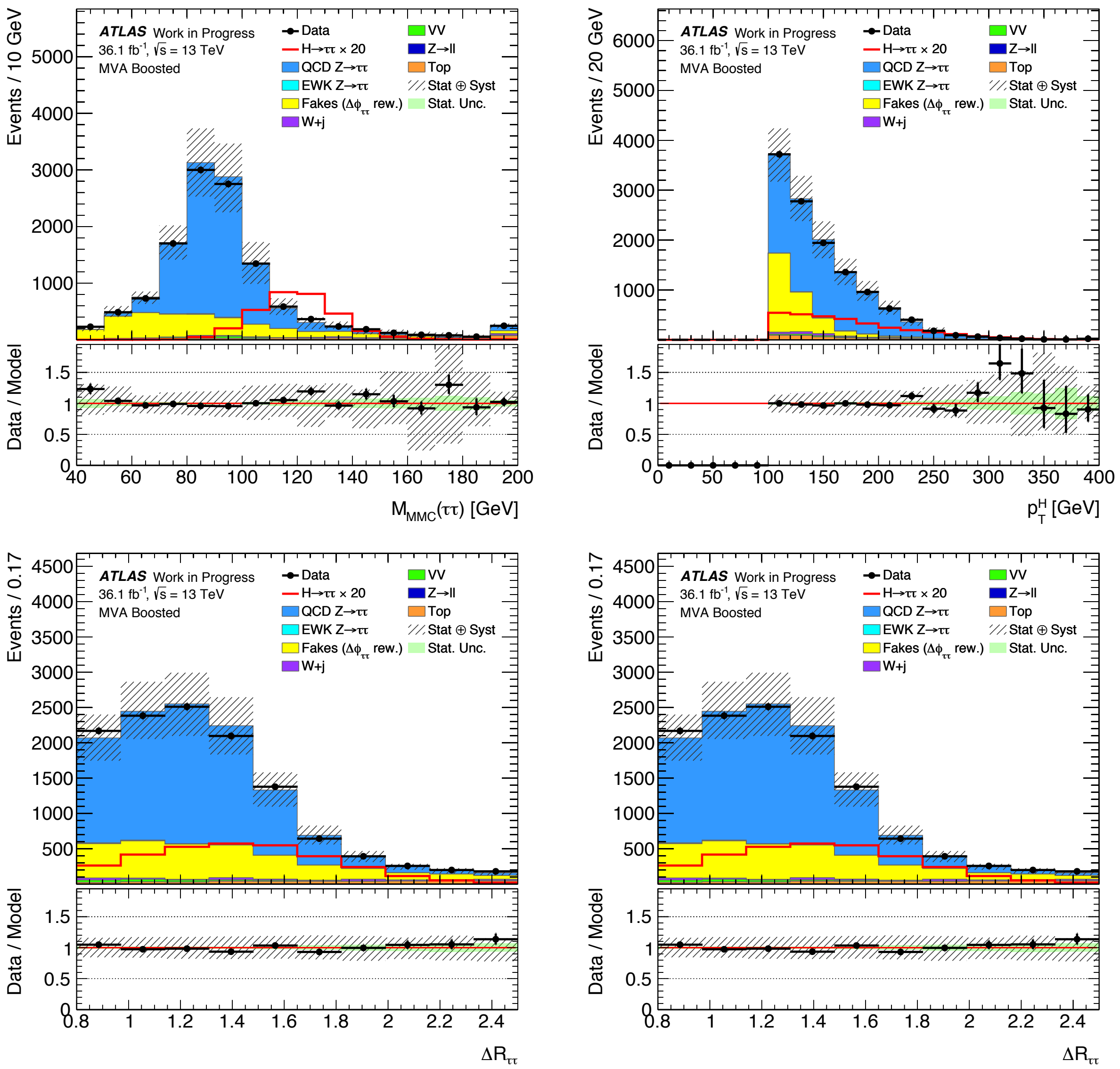


Figure 8.6: Modelling of selected input variables used in the $\tau_{\text{had}}\tau_{\text{had}}$ boosted BDT. The uncertainty band includes statistical and systematic uncertainties.

Training Statistics The limited number of training events in MVA VBF and MVA Boosted region leads to decrease in BDT performance and strong tendency to overfitting. While larger simulated samples could improve this situation, the data-driven Fake background estimate is intrinsically limited by the number of recorded events. In order to achieve stable result and avoid overfitting the selection is loosened to allow for additional events in the training of the BDTs. The kinematics and correlations of variables in the resulting event set resemble the set used in the statistical analysis closely, as detailed in Appendix B. By changing the identification criteria of the $\tau_{\text{had}}\tau_{\text{had}}$ pair as summarised in Tab. 8.2, the topological and kinematic characteristics of the events For each process the $\text{ID}(\tau_0, \tau_1)$ is loosened from (tight,tight) to:

- $Z \to \tau\tau$: ID(τ_0, τ_1) =(medium,medium)
- Fake background: ID(τ_0, τ_1) =(loose,loose)
- $ggH \to \tau\tau$ and VBF$H \to \tau\tau$: ID(τ_0, τ_1) =(loose,loose)

This has in particular a dramatic effect on the available Fake background events.

Region	Process	ID(τ_0, τ_1)=(tight,tight)		Looser ID reweighted	
VBF	$Z/\gamma^* \to \tau\tau$	952.27	(12216)	1373.96	(18006)
	VBF$H \to \tau\tau$	27.98	(11478)	51.93	(21198)
	Fakes	646.82	(928)	933.25	(13929)
Boosted	$Z/\gamma^* \to \tau\tau$	8649.99	(126659)	12571.77	(184323)
	$ggH \to \tau\tau$	117.75	(8444)	215.47	(15550)
	VBF$H \to \tau\tau$	22.68	(9444)	41.25	(17107)
	Fakes	3562.39	(4893)	5177.52	(70505)

Table 8.2: Number of training events available in the default tight and loosened identification scenario. The events are reweighted to preserve the relative normalisation.

In order to preserve the relative ratio between different processes and thereby avoiding a bias of the BDT classifier, the resulting events are reweighted to preserve the signal region background composition. Ratios are calculated by dividing the relevant processes for each identification scenario:

$$r^{tt} = {}^{\text{Fakes}^{tt}}/_{Z^{tt}} \qquad \text{and} \qquad r^{ll} = {}^{\text{Fakes}^{ll}}/_{Z^{ll}}$$

From these ratios, scale-factors are calculated and applied on the Fake background:

$$SF^{VBF} = \left.\frac{r^{tt}}{r^{ll}}\right|_{VBF} \qquad \text{and} \qquad SF^{Boost} = \left.\frac{r^{tt}}{r^{ll}}\right|_{Boost}$$

The resulting factors are summarised in Tab. 8.3.

Region	Processes	r^{tt}	r^{ll}	SF
VBF	Fakes/$Z \to \tau\tau$	0.68	7.09	0.096
Boosted	Fakes/$Z \to \tau\tau$	0.41	4.10	0.100
	ggH/VBFH	7.09	7.13	0.994

Table 8.3: Scale-factors applied on the Fake background events to preserve relative normalisation ratios.

8.2.2.3 Optimisation of BDTs

The decision tree specific parameters introduced in Sec. 8.2.1.3, `MaxDepth` and `MinNodeSize` together with the boosting-related parameters `nTrees`, `Shrinkage` and `nCuts` determine the performance of the trained BDTs. The tuning of these parameters is analysis-specific, as it depends on the available statistics and the BDT input variables. Therefore, the optimal parameter setting is not a-priori known but was adjusted with the following empirical strategy.

BDT Parameter Scans In order to find the optimal setting for the $\tau_{\text{had}}\tau_{\text{had}}$ analysis, a scan over the BDT parameter space was performed independently for the MVA VBF and Boosted category. For each parameter set, the even (odd) BDTs are trained and then evaluated on the odd (even) testing sets.
The scan strategy is depicted in Fig. 8.7. For the initial step, the full variable list presented in Tab. 8.1 was used as input to each BDT. The initial parameter grid was split into four sub-grids, each with varying granularity for the scan. A high granularity was chosen for the parameter grid spanned by `Shrinkage` $\in (0, 0.1)$ and `nTrees` $\in [20, 750]$, while the others

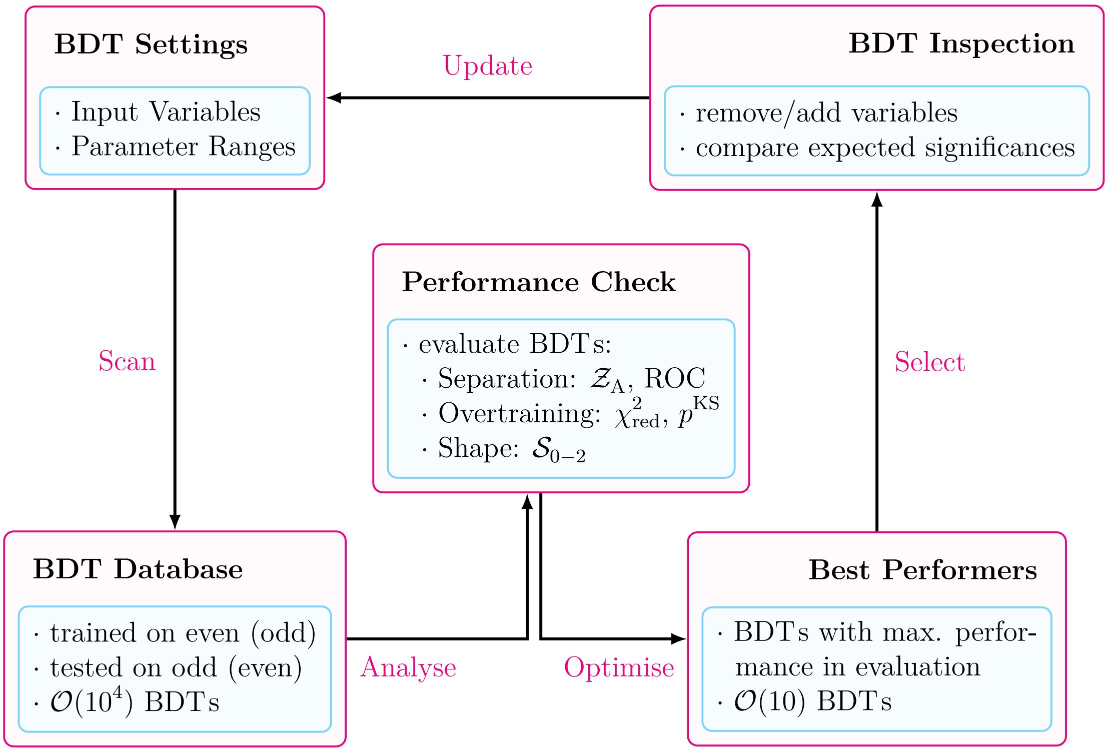


Figure 8.7: Sketch of the BDT parameter scan strategy. For a given set of input variables and parameter grid, a large performance database is created. This database contains several figure-of-merits describing the separation power, overfitting and BDT shape for each even (odd) BDT evaluated on training and testing set. Parameter settings which yield a high separation and show only small overfitting are chosen and manually inspected. Insights of this inspection are used to redefine the initial set and the procedure is repeated.

were scanned more broadly to save computing resources. For all sub-grids, the number of nodes for each decision tree `MaxDepth` was varied between $3, \dots, 6$. The minimum percentage of events in a node `MinNodeSize` was varied between $1\%, \dots, 10\%$. Finally, the intrinsic BDT algorithm granularity `nCuts` was set to 140 for all BDTs, corresponding to reasonable binning of the range for the variable with highest outliers.

For each grid point, the performance was evaluated by applying the BDT on training and testing set and inspecting the corresponding BDT score distributions. The following figure-of-merits (FOM) were chosen:

Separation Power: Quantifying the separation of signal and background events

- the area under the ROC[1]-curve, which is 1 for perfect separation and 0 for completely overlapping samples
- the Asimov significance $\mathcal{Z}_{\mathrm{A}}$ [170], an approximation to the expected significance introduced in Sec. 8.3.2. It is defined as

$$\mathcal{Z}_{\mathrm{A}} = \sqrt{2 \sum_{i=1}^{N} \left[\left(n_i^s + n_i^b \right) \ln \left(1 + {n_i^s}/{n_i^b} \right) - n_i^s \right]^2} \tag{8.3}$$

  for the BDT distribution with N bins, each containing n_i^s (n_i^b) signal (background) events.

Overfitting: Validating the agreement of BDT response between the BDT distributions of training and testing set

- a χ^2_{red} hypothesis test
- a Kolmogorov-Smirnov test [169] for binned histograms, with probability p^{KS}. Any parameter set yielding a $p^{\mathrm{KS}} < 0.05$ is rejected.

Shape: Criteria for the training and testing set, excluding large bin-wise fluctuations and double peaks arising from suboptimal splitting in the nodes.

For each category, the parameter set is selected which yields the maximal or minimal FOM, depending on the type. This optimisation is done simultaneously on odd and even BDTs and any parameter set with different performance on odd or even training set is rejected. This serves as additional validation against overfitting. Different performance categories were defined by combinations of the above FOM and are detailed in Appendix B.

The scan results for MVA VBF are visualised in Fig. 8.8, where the two FOM $\mathcal{Z}_{\mathrm{A}}$ and p^{KS} are shown in dependence on the `nTrees` and `Shrinkage` parameters. The respective contours reveal regions of stable performance. In particular it can be seen, that a large $\mathcal{Z}_{\mathrm{A}}$ does not necessarily correspond to an optimal setting, as areas with higher overfitting are observed, reflected in the corresponding p^{KS} for either signal or background.

The final selected parameter setting for the initial parameter scan is summarised in Tab. 8.4.

---

[1] Receiver Operator Characteristic, defined by signal efficiency versus background rejection

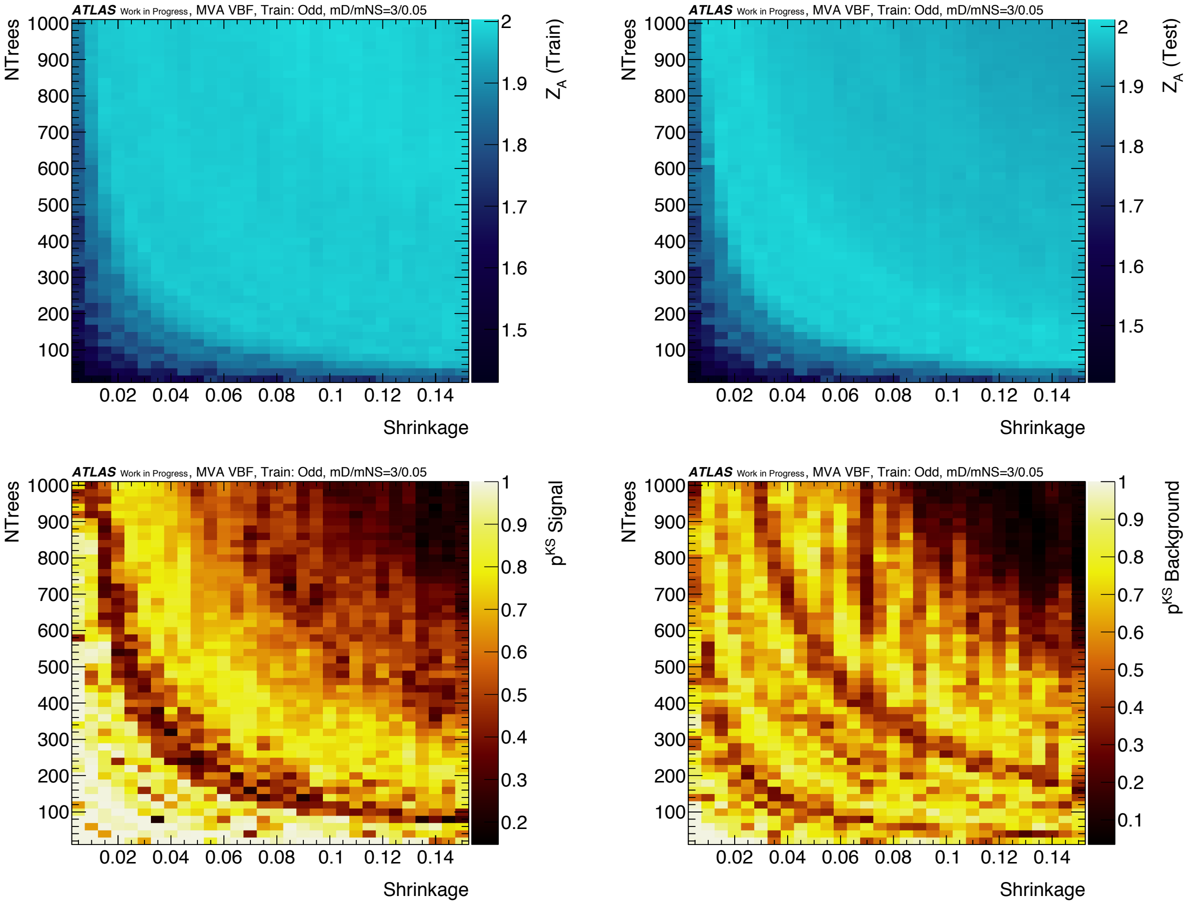


Figure 8.8: Asimov significance $\mathcal{Z}_\mathrm{A}$ and p^KS values per scan point for the BDT score distributions of training and testing sets in the MVA VBF category. For the figures on the left (right) the BDTs were trained on the even (odd) set. The additional BDT parameters MinimumNodeSize and MaxDepth vary between (1,5) and (3,4,5), respectively.

BDT Input Variable List Trimming The chosen BDTs were trained on the inclusive input variable list in Tab. 8.1 which is reduced with the following, iterative procedure hereafter referred to as trimming. In principle this full list could be used, as the discussed algorithm is insensitive to additional, potentially unused variables. Redundant variables however could introduce statistical noise and further dependencies on systematic uncertainties. Therefore, a minimum set of discriminating variables which does not lead to a decrease in $\mathcal{Z}_\mathrm{A}$ was determined.

The trimming is based on the importance ranking of the input variables. A variable's importance is calculated by counting how often it is used during the node splitting, as a result of having the highest separation power. A weight is applied on this number defined by the achieved separation gain during splitting and the number of events in the node [171]. The importance is calculated over all trees in the forest.

Starting with the initial variable set the trimming is performed by subsequently removing the lowest ranked variable. For each step, $\mathcal{Z}_\mathrm{A}$ is evaluated and the trimming is stopped when only one variable is left. The final variable set is then determined as the minimal set

	MVA	Boosting	`nTrees`	`MaxDepth`	`MinNodeSize`	`Shrinkage`
Trimming	VBF SF	gradient	140	4	5	0.12
	boosted	gradient	980	3	5	0.07
Final	VBF	gradient	500	0.005	3	0.05
	boosted	gradient	840	0.2	4	0.04

Table 8.4: Parameter settings selected with the scan strategy. Two scans were performed, first to estimate well performing parameter configurations and perform variable trimming and second to determine the optimal configuration for the final BDTs.

yielding a $\mathcal{Z}_\mathrm{A}$ close to the maximum.
In Fig. 8.9, the resulting trimming curve for the training of the even MVA VBF BDT is depicted. The result for the Boosted category can be found in Appendix B. For each MVA category, the final list of chosen variables is indicated by ✓ in Tab. 8.1. Each variable set represents characteristic features of the respective decay:

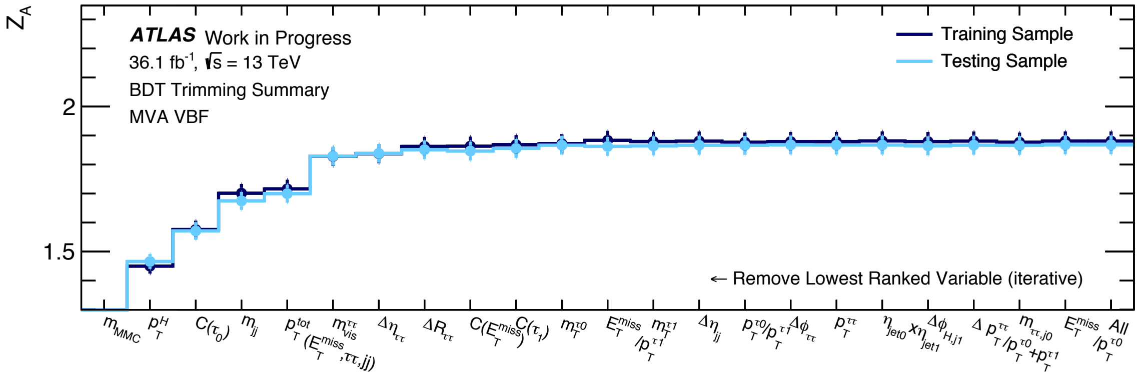


Figure 8.9: Visualisation of the trimming result for the BDT trained on even MVA VBF events. Starting with the full initial variable set, the lowest ranked featured, indicated by the bin labels are removed and the BDT is retrained. For each variable setting, $\mathcal{Z}_\mathrm{A}$ is evaluated on the BDT score distributions. A drop in $\mathcal{Z}_\mathrm{A}$ indicates decreased separation power.

- **VBF** features jet-related variables and topological information about the relative position of the $\tau_\mathrm{had}\tau_\mathrm{had}$ pair, central inbetween the two VBF jets
- **Boosted** benefits in particular from variables describing or correlated to the large transverse momentum of the Higgs boson decay system.

In both cases, the m_MMC offers the highest separation power. The Boosted result in Appendix B furthermore reveals, that only the $\Delta R_{\tau\tau}$ and p_T^H, also used in the categorisation for the CBA offer significant additional separation power.

Final BDT Evaluation The final BDT setting was then optimised by rerunning the parameter scan in Fig. 8.7 with the updated variable list. For this scan, a finer scan granularity for the `MinNodeSize` and `MaxDepth` of the decision tree was chosen, while the

`Shrinkage` and `nTrees` ranges were restricted to previously determined, well-performing regions. Parameter settings for the final selected BDTs are summarised in Tab. 8.4.
In Fig. 8.10, the response of the chosen BDTs for VBF and Boosted on the respective training and testing event set is shown. A good agreement is observed in particular for high (low) scores for signal (background), quantified by a p^{KS} test. Large statistical uncertainties are observed for the low score tail of the signal processes, originating from the limited number of available, simulated events. The importance ranking of input variables for these final BDTs is shown in Tab. 8.5.

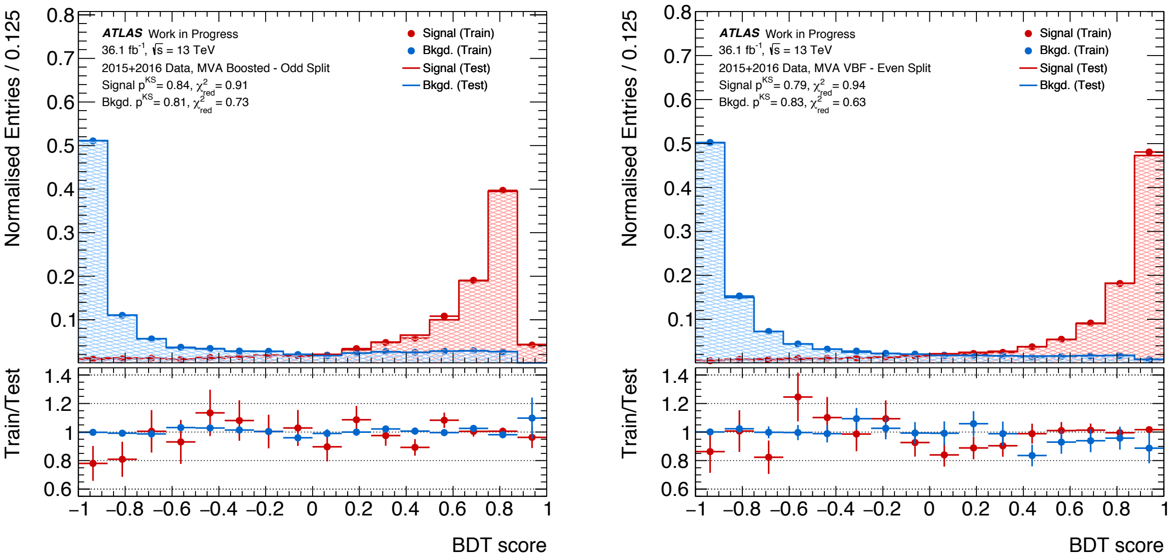


Figure 8.10: Overlay of the BDT response on the training and testing sets for MVA VBF (left) and MVA Boosted (right). The achieved agreement ensures the absence of overfitting.

MVA VBF		MVA Boosted	
Variable	Importance	Variable	Importance
m_{MMC}	0.25	m_{MMC}	0.46
$p_T(\tau_0, \tau_1, j_0, j_1, \mathrm{MET})$	0.14	m_{vis}	0.16
$\mathcal{C}_{jj}(\tau_0)$	0.11	$\Delta R_{\tau\tau}$	0.14
m_{jj}	0.10	p_{T}^{H}	0.13
$\mathcal{C}_{jj}(\tau_1)$	0.10	$\Delta\eta_{\tau\tau}$	0.11
$\mathcal{C}_{\tau\tau}(E_{\mathrm{T}}^{\mathrm{miss}})$	0.09		
$\Delta R_{\tau\tau}$	0.09		
$\Delta\eta_{\tau\tau}$	0.07		
p_{T}^{H}	0.06		

Table 8.5: Ranking of input variables for the selected BDTs.

8.2.3 Final Discriminants for Multivariate Approach

The BDTs are then applied on all signal region samples and studied.

Correlations Before evaluating the selected BDT on data, the correlations between input variables were studied. In Fig. 8.11-8.12, the correlation between selected variables are shown, including the BDT score itself. The good agreement is quantified with a Kolmogorov-Smirnov test. For the latter, the distributions are blinded in signal sensitive regions by requiring $m_{\mathrm{MMC}} \in [0, \ldots, 100\,\mathrm{GeV}) \wedge (150\,\mathrm{GeV}, 200\,\mathrm{GeV}]$ and removing the BDT score region with more than 30% signal contribution.

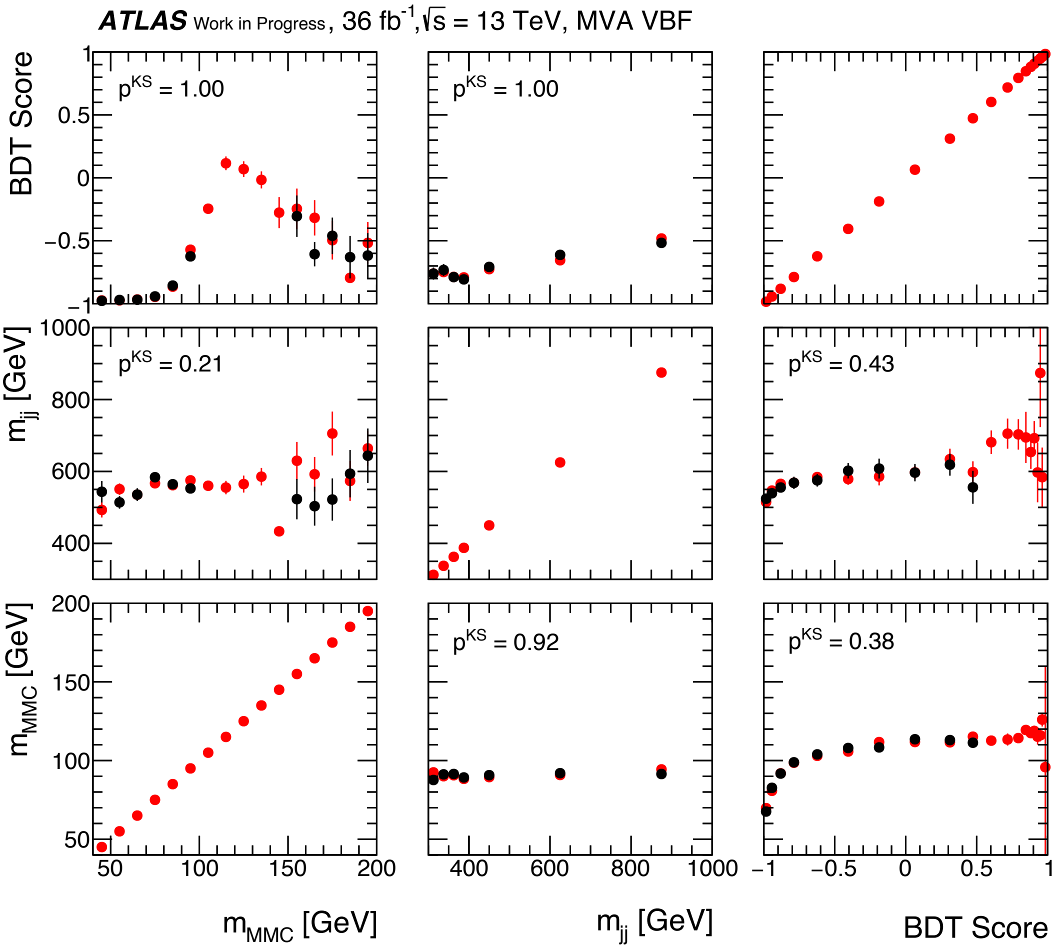


Figure 8.11: Correlation between selected BDT input variables and the BDT score used in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ VBF (left) and Boosted (right) categories. The agreement of the sum of all backgrounds (red) and data (black) is quantified with a Kolmogorov-Smirnov test, excluding particular signal-sensitive regions with of m_{MMC} and BDT score distributions. Only statistical uncertainties are considered.

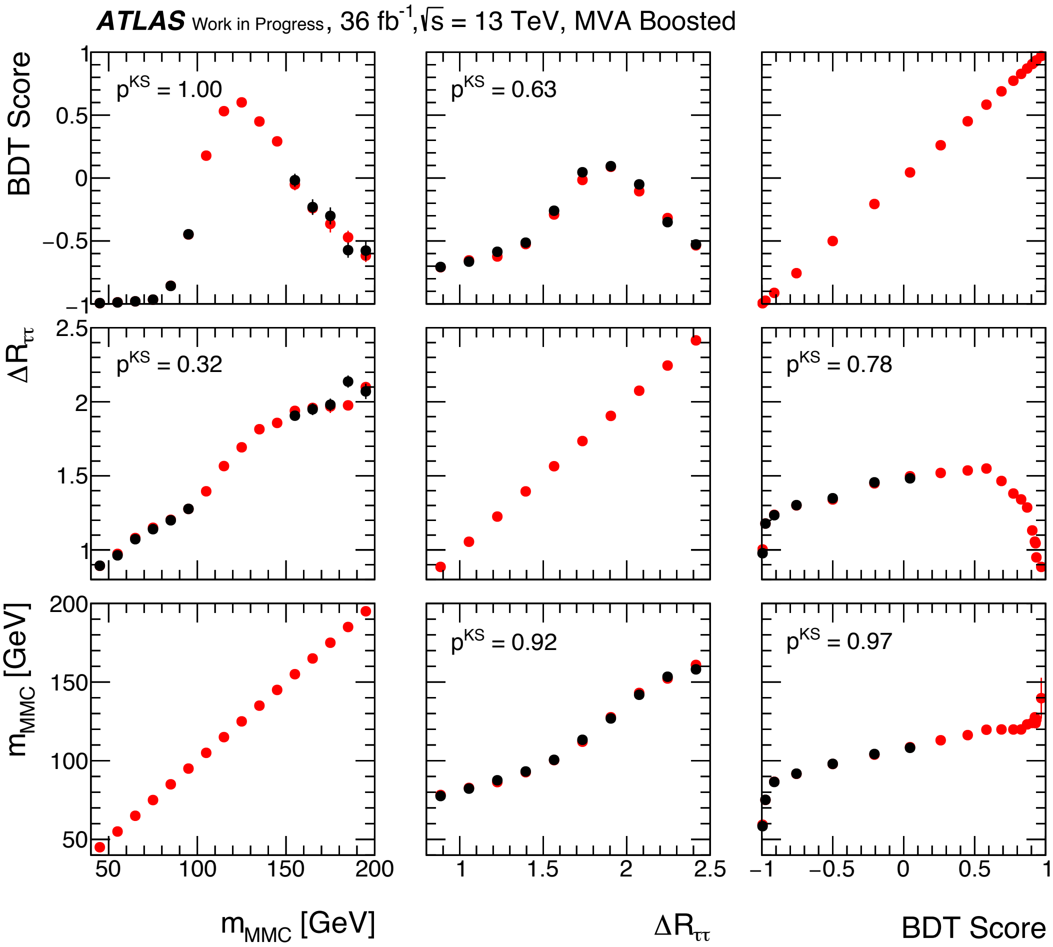


Figure 8.12: Correlation between selected BDT input variables and the BDT score used in the $\tau_{\text{had}}\tau_{\text{had}}$ VBF (left) and Boosted (right) categories. The agreement of the sum of all backgrounds (red) and data (black) is quantified with a Kolmogorov-Smirnov test, excluding particular signal-sensitive regions with of m_{MMC} and BDT score distributions. Only statistical uncertainties are considered.

BDT Modelling The unblinded pre-fit BDT score distributions for the MVA VBF and Boosted signal categories are shown in Fig. 8.13. A clear excess is observed for high score values. The region dominated by backgrounds are well-modelled, confirming the chosen BDT training strategy. From the distributions it is evident that the trained BDTs are excellent at distinguishing Fake background contributions from signal-like events, in particular for MVA VBF. $Z/\gamma^* \rightarrow \tau\tau$ backgrounds on the other hand contribute to signal dominated regions due to the similar event topologies.

High BDT Score Region As a first estimate of the BDTs separation power, a signal-enriched region is constructed by setting a lower bound on the BDT score. The threshold is tuned such that the integral of signal events of the targeted production mode - ggH for Boosted and VBFH for VBF - above the threshold exceeds 70% of the overall signal yield. In Fig. 8.14, the corresponding m_{MMC} distributions in such high-score regions are

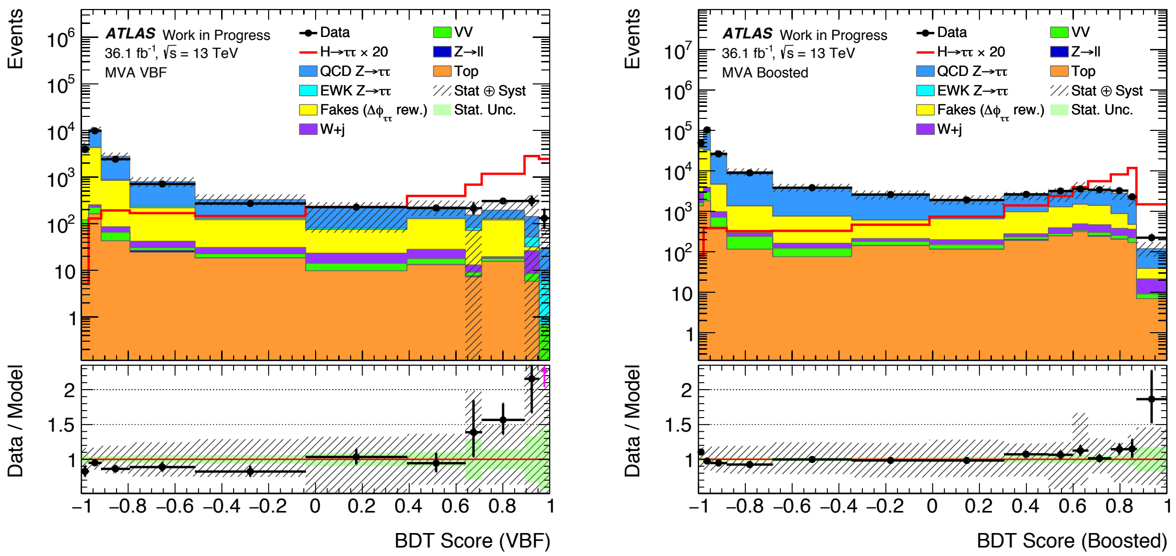


Figure 8.13: Modelling of the BDT score distributions for the MVA VBF (left) and MVA Boosted categories (right). The binning is chosen following the optimisation procedure in Sec. 8.3.5. The large uncertainties in the fourth, second and first bin of the VBF BDT distribution are caused by bin migration effects for backgrounds with a low number of events. These are treated with the strategy outlined in Sec. 8.3.4.

depicted for MVA VBF and Boosted. While the BDT is efficient at suppressing background contributions in the MVA VBF phase space, the distribution in the MVA Boosted category exhibits a sizeable background contamination, in particular arising from $Z/\gamma^{*} \rightarrow \tau\tau$ events. This indicates, that the kinematics and topology in events passing the Boosted selection is difficult to distinguish from background sources.

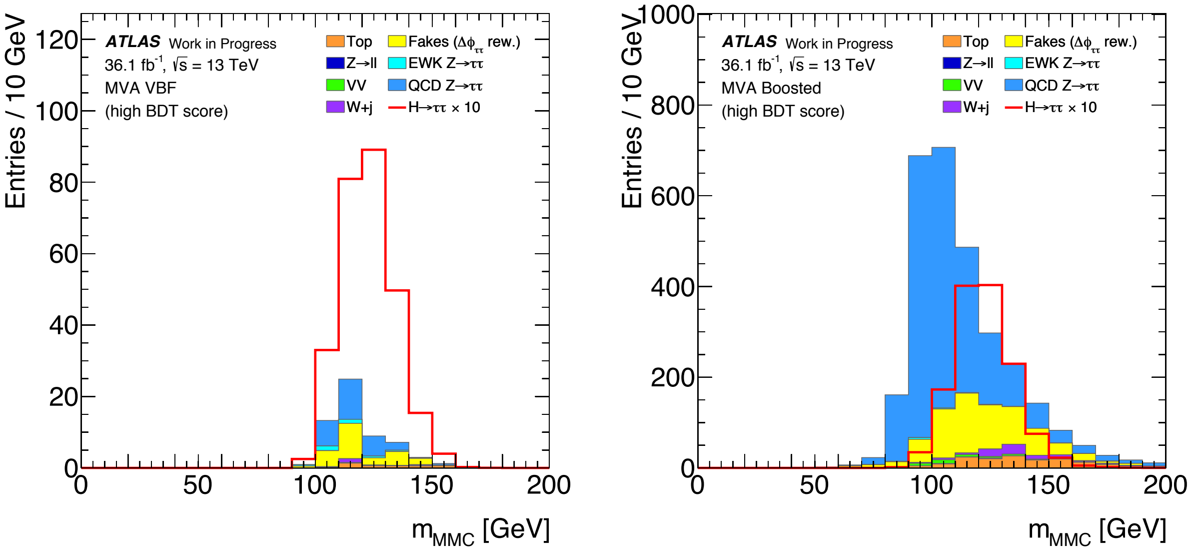


Figure 8.14: m_{MMC} distributions in the MVA VBF (left) and MVA Boosted (right) regions. A minimum cut is set on the BDT scores such that 70% of the signal events remain.

8.3 Statistical Model and Validation

The following section describes the statistical formalism used to determine significances of a deviation from a background-only hypothesis H_0 and to extract the signal strength μ. In simple terms, the statistical procedures deployed in this analysis test the agreement between the measured data and H_0 as well as the signal hypothesis H_1, which adds the expected signal contribution to H_0. A minimisation procedure yields the value for μ which fits H_1 best, yielding a measure of the $H \to \tau_{\text{had}}\tau_{\text{had}}$ cross section times branching ratio. The final discriminants presented in the previous chapter are used as input to a significance test based on a test statistic defined by a ratio of likelihood functions. This frequentist approach [172] to hypothesis testing is done independently for CBA and MVA. The parameters used to construct those likelihood functions are following the LHC Higgs Combination Group recommendations described in [173] and adapted to current RUN II conditions. The formal explanations in this section are based on [170].

8.3.1 Likelihood Function

To compare the measured data with the expectations, it is important to incorporate uncertainties on the prediction in the statistical model. The parameters of the prediction are grouped as nuisance parameters (NP). The nominal parameter settings are used to define the nominal predictions; their uncertainties result in variations of the prediction, as depicted in shaded bands of the final discriminant distributions. The true value of the parameters for the measurement is not a-priori known but assumed to lie within the 1σ uncertainty of each NP at a 68% confidence level. The statistical model therefore has to account for possible deviations from the nominal value of each NP. This is achieved by a fit to the measured data, as will be worked out in the following. Input to the statistical model comprises the expected event model, all NPs affecting normalisations and shapes of the distributions and the signal strength μ, the parameter of interest.

The input distributions are provided as binned histograms with bins n_i and $i = 0, \ldots, N$. The expectation value of each bin can be expressed by the predicted signal and background mean values per bin n_i^s and n_i^b as

$$n_i^{\text{exp}}(\mu, \theta) = \mu \cdot n_i^s(\theta) + n_i^b(\theta) \qquad . \tag{8.4}$$

The mean values $n_i^s(\theta)$ and $n_i^b(\theta)$ can be parameterised using PDFs, which, when integrated, determine the probability to find an event of the respective contribution in bin i. These PDFs depend on the NP of the event model and are denoted as θ. They are divided into three classes:

- the NPs encoding parameters of the event model and their systematic uncertainties θ_{syst}, comprised of the nominal values ν_α, the fit parameter p_α and the uncertainty σ_α,
- NPs θ_η with fit parameter p_η and uncertainty σ_η, introduced to allow floating normalisations in the fit for the background modelling samples,

- the bin-by-bin statistical uncertainty NP θ_γ, consisting of the nominal expected event yield ν_γ, statistical error σ_γ and the fit parameter p_γ.

In this analysis, the full parametrised PDFs used for the event model are composed of three different PDF types f_θ [173]:

- A Gaussian distribution G is used to model NP arising from systematic experimental sources. An example of such a NP is the τ-identification efficiency, with nominal value and uncertainties determined in a dedicated measurement as described in Sec. 5.4.3. The nominal value of each NP ν_α is varied in the fitting procedure by the floating parameter p_α[1]. This PDF is hence given as:

$$f_\alpha(\nu_\alpha|p_\alpha,\sigma_\alpha) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left(-\frac{(\nu_\alpha - p_\alpha)^2}{2\sigma_\alpha^2}\right) \tag{8.5}$$

- systematic NPs $\nu_{\tilde{\alpha}}$ originating from external theoretical predictions, such as cross sections, are modelled by defining a log-normal distribution

$$f_{\tilde{\alpha}}(\nu_{\tilde{\alpha}}|p_{\tilde{\alpha}},\sigma_{\tilde{\alpha}}) = \frac{1}{\nu_{\tilde{\alpha}}\sqrt{2\pi\sigma_{\tilde{\alpha}}^2}} \exp\left(-\frac{\ln({}^{\nu_{\tilde{\alpha}}}/{}_{p_{\tilde{\alpha}}})^2}{2\sigma_{\tilde{\alpha}}^2}\right) \tag{8.6}$$

- Finally, statistical uncertainties are also encoded as NP based on a generalised form of the Poisson distributions

$$f_\gamma(n|\nu_\gamma p_\gamma) = \frac{(\nu_\gamma p_\gamma)^n}{n!} \exp\left(-\nu_\gamma p_\gamma\right) \tag{8.7}$$

  for non-integer input, called Γ-distributions. The event yield n is decomposed in the nominal yield ν_γ and a relative factor p_γ parameterising the relative statistical uncertainty.

The HistFactory and RooStats software packages provide functionalities to construct such PDFs and were used in this analysis [174, 175]. The number of observed events in each bin n_i follow a Poisson distribution. The expected number of events n_i^{exp} from Eq. 8.4 depend on the μ and the NP θ. By summing over all event model processes p, the expected yield per bin i can be expressed as

$$n_i^{\text{exp}}(\mu,\theta) = \sum_s \lambda^s p_\gamma^{i,s} \cdot p_\eta^s(\theta) \cdot \sigma_\eta^s(\theta) \cdot \nu^{i,s}(\mu,\theta) \tag{8.8}$$

in each considered signal region. Here, λ^p denotes the integrated luminosity used to normalise simulated sample s. The $p_\gamma^{i,p}$ describe the relative statistical uncertainty for sample s in bin i. The normalisation parameters p_η^s are used to adjust the total background normalisations to observed data. They are introduced for the $Z/\gamma^* \to \tau\tau$ process and the Fake background templates in this analysis. Their uncertainties are parameterised with σ_η^s. Finally, ν^i is the yield in bin i and depends on the systematic uncertainties affecting

---

[1] In the fit the nominal value is $\nu_\alpha = 1$, as the measurements have been applied to the nominal event model. p_α is scaled such that the resulting Gaussian has a variance of 1.

the shapes. With this information, the statistical model can be expressed as a binned likelihood function with products over all considered categories c and input distribution bins i

$$L(\mu, \theta) = \prod_c \prod_i P(n_i | n_i^{\mathrm{exp}}) \cdot \prod_\theta \mathrm{PDF}_\theta(\nu_\theta | p_\theta) \tag{8.9}$$

where $P(n_i|n_i^{\mathrm{exp}})$ is the Poisson probability of observing n_i events in a bin i when expecting n_i^{exp} events. The fit is then performed using the RooFit software package [163] and Minuit function minimisation and error routines [176]. The parameter set $(\hat{\mu}, \hat{\theta})$ which maximises the likelihood L is called the Maximum Likelihood Estimator (MLE). The MLE can intuitively be understood as the set of parameters yielding the best agreement between measurement and model. When fitting the dataset against the signal hypothesis H_1, the $\hat{\mu}$ is therefore a measure of the cross section value. The uncertainty on this measurement is extracted by varying the nominal values of each nuissance parameters within $\pm 1\sigma$ and repeating the fit. The impact on $\hat{\mu}$ from this variation is the uncertainty on the measurement, σ_μ.

In practice, the likelihood-maximisation is achieved by minimising the negative logarithmic likelihood $\textsc{Nll} = -\ln L(\mu, \theta)$.

8.3.2 Likelihood Profiling and Statistical Testing

The evaluation of the statistical model includes profiling of the likelihood functions for CBA and MVA, testing with the observed dataset and quoting expected results. The aim of these steps is to exclude the background-only hypothesis H_0 in presence of a signal in the measurement and validate the signal model H_1.

Profile Likelihood Ratio Aside of the parameter of interest μ, the likelihood function in Eq. 8.9 depends on the multitude of NP θ. To facilitate the fitting procedure, the NP are expressed as a function of μ. This is called profiling, the NPs are absorbed during the fitting procedure.

For this, a NP-independent profile likelihood ratio is defined by

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} \tag{8.10}$$

with the aforementioned (unconditional) MLE and the conditional MLE $L(\mu, \hat{\hat{\theta}})(\mu)$. The latter uses the set of NP $\hat{\hat{\theta}}(\mu)$ which maximises the likelihood under a condition - in this case for a given μ. The likelihood ratio assumes values between $0 \leq \lambda \leq 1$, where larger values mean better agreement between observed data and model.

Statistical Testing The profile likelihood ratio is then modified, resulting in the so-called test statistic

$$t_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \geq 0 \\ -2\ln\lambda'(\mu) & \hat{\mu} < 0 \end{cases} \qquad \text{with } \lambda'(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(0, \hat{\hat{\theta}}(0))} \tag{8.11}$$

to account for negative $\hat{\mu}$. These can occur in the fit, as long as n_i^{exp} in Eq. 8.4 stays positive. t_μ is used to test the compatibility of hypotheses to the observed data. If a fit maximisation yields a MLE with $\hat{\mu} < 0$, the observed data agrees best with the signal-free hypothesis H_0 which motivates the definition of $\lambda'(\mu)$. This test statistic is used in the fit for measuring the signal strength μ in the MVA and CBA.

For rejecting the background-only hypothesis H_0 with $\mu = 0$, Eq. 8.11 is reduced to

$$q_0 = \begin{cases} -2\ln\lambda(\mu = 0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \tag{8.12}$$

In case the observed number of events does not show any deviation from or are even smaller than the background prediction, q_0 gives 0, corresponding to no deviation from the background model. If a higher number of events is observed than predicted by the background model alone, q_0 assumes larger values. This disagreement can be quantified by the p_0-value

$$p_0 = \int_{q_0^{\mathrm{obs}}}^{\infty} f(q_0|\mu = 0)\,\mathrm{d}q_0$$

with lower integral boundary given by the observed q_0. The PDFs $f(q_0|\mu = 0)$ are detailed in [170] and are derived after the approximation described in the following section. This p-value is often converted into an equivalent significance $\mathcal{Z}_{\mathrm{obs}}$, given by the quantile Φ^{-1} of the standard Gaussian distribution:

$$\mathcal{Z}_{\mathrm{obs}} = \Phi^{-1}(1 - p) \tag{8.13}$$

To claim a discovery of a particle[1] a commonly accepted threshold is set at $Z = 5$. This corresponds to a probability of $p_0 = 2.87 \times 10^{-7}$ to measure the observed data if H_0 is valid. A threshold of $Z = 3$ corresponds to $p_0 = 1.35 \times 10^{-3}$ and constitutes an evidence of a signal.

Expected Significances and Asimov Dataset An important measure for optimising an analysis is the sensitivity to a signal hypothesis, quantified as expected significance $\mathcal{Z}_{\mathrm{exp}}$ for a search. Due to the non-linearity of $\mathcal{Z}_{\mathrm{obs}}$ as defined in Eq. 8.13 in p, the expected significance is quoted as median of $\mathcal{Z}_{\mathrm{obs}}$ [170]. Categorisations, cut and binning optimisations as presented in Sec. 8.3.5 ultimately all aim for a maximal sensitivity of a given analysis.
The basic idea for quantifying $\mathcal{Z}_{\mathrm{exp}}$ is to construct an artificial dataset called Asimov

---

[1] Or an observation of a process.

dataset which fulfils the signal hypothesis H_1 and to extract $\mathcal{Z}_{\text{exp}}$ by testing the level of agreement of this dataset to H_0. The Asimov dataset is defined such that its MLE $(\hat{\mu}_A, \hat{\theta}_A)$ assume the values of H_1. This means, that signal hypothesis holds as defined in the event model, in particular that the signal strength is $\mu = 1$. Furthermore, the parameters of each NP $\hat{\theta}_A$ assume their original input values. The Asimov dataset therefore corresponds exactly to the event model prediction assuming the presence of a signal.
Under the asymptotic approximation[1] [177, 178] and physical constraints, such as $\hat{\mu} > 0$, the median significance is given by $\mathcal{Z}_{\text{exp}} = \sqrt{-2 \ln \lambda(\hat{\mu}_A, \hat{\theta}_A)}$. Hence, testing the Asimov dataset to the background-only hypothesis defines the expected significance.
The expected significance was used in this thesis to optimise the sensitivity of the signal extraction procedures. This is equivalent to minimising the uncertainty on the the MLE $\hat{\mu}$ and thereby the uncertainty on the cross section times branching ratio measurement for $H \to \tau_{\text{had}}\tau_{\text{had}}$.

8.3.3 $\tau_{\text{had}}\tau_{\text{had}}$ Fit Model

The model introduced in the previous section is fitted to the observed, combined dataset recorded in 2015 and 2016.

In Fig. 8.15, the $\tau_{\text{had}}\tau_{\text{had}}$ specific fit models for the independent cut-based and multivariate analysis are sketched. The three (two) CBA VBF (Boosted) signal regions are combined to an inclusive category. In each category, the normalisation for the $Z/\gamma^* \to \tau\tau$ process is left free floating in the fit, correlated across the individual signal regions. To improve the precision of the Fake background normalisation in the VBF region, a correlated normalisation nuisance parameter is introduced between both inclusive regions.

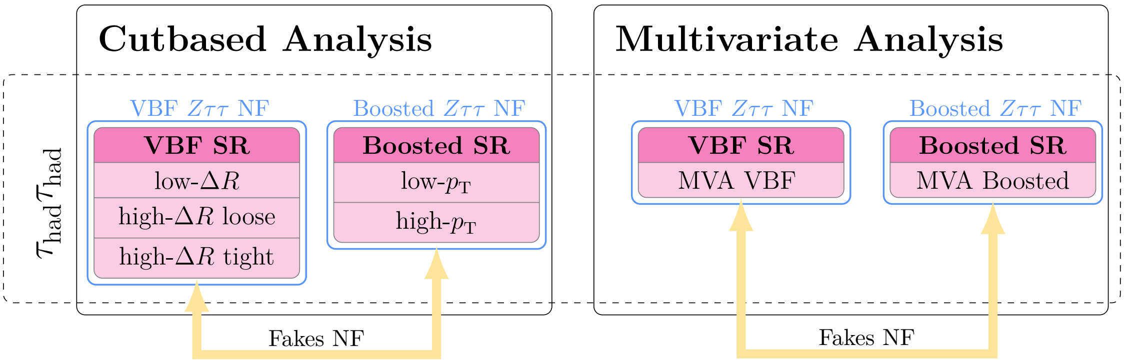


Figure 8.15: Pictogram of the $\tau_{\text{had}}\tau_{\text{had}}$ standalone fit-model. Individual signal regions are combined to inclusive CBA VBF and Boosted regions respectively, each using an independent normalisation nuisance parameter for the $Z/\gamma^* \to \tau\tau$ process. The Fake background normalisation is treated fully correlated between the regions.

The distributions of the final discriminant are considered as input for each signal category, specifically:

---

[1] Valid for a large number of events.

- three (two) binned m_{MMC} histograms for the CBA VBF (Boosted) signal regions
- one binned BDT-score histogram for each, MVA VBF and MVA Boosted signal region

The binning of these distributions is determined as detailed in Sec. 8.3.5. The normalisations are treated as follows:

- the normalisation of the Fake background templates are free floating and correlated between the event topologies. Due to the small amount of Fake background events in the VBF categories, the additional constraining power from the Boosted categories is exploited.
- the $Z/\gamma^* \to \tau\tau$ normalisations are free floating within each inclusive category[1] to account for phase space specific effects.

The fit-model used for the combination differs in particular for the normalisation strategy, as additional control regions from the other final states can be used to set constraints. This is described in Sec. 9.3.

8.3.4 Fit-Model Simplification

Before the measurement of the signal strength, the fitting procedure is simplified and validated. This is in particular important for the combination of all three analysis channels, due to the large number of fit parameters and resulting complexity of the fit. The procedures are common across channels and were verified individually and for the channel combination. As an example, the combined effect of the procedures below is depicted in Fig. 8.16 for two of the most important nuisance parameters in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ analysis.

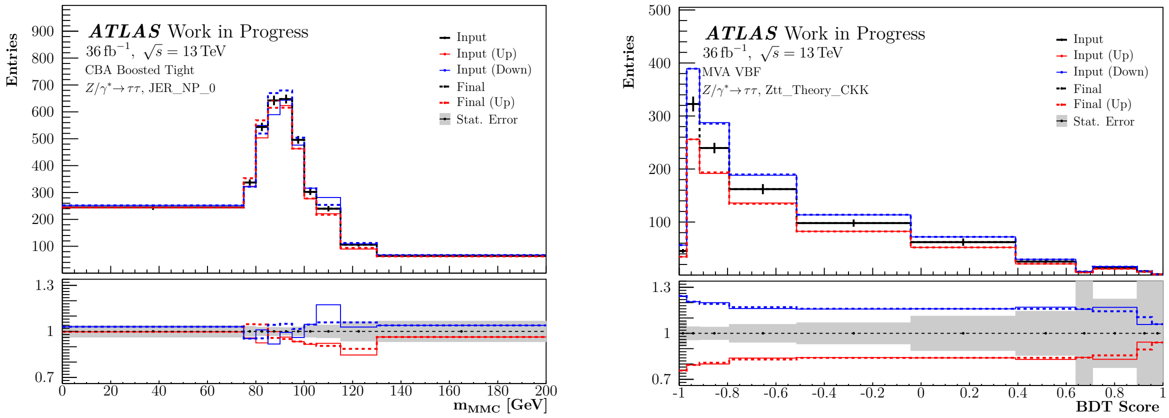


Figure 8.16: The presented distributions reflect the effects of the simplification procedures on a NP describing the jet energy resolution (left) and the CKKW-matching $Z/\gamma^* \to \tau\tau$ theory uncertainty (right) on the discriminants in CBA and MVA, respectively.

The simplification scheme is detailed in the following and based on a comparison between

---

[1] Meaning correlated between the 3 (2) combined CBA categories for VBF (Boosted).

the nominal fit input histograms to the systematically varied ones. For each step of the simplification, the impact on the expected significance was studied and validated.

Symmetrisation For each bin in each input histogram, the presence of systematic uncertainties with both $\pm 1\sigma$ variations in the same direction, called one-sided variations, is checked. These potentially introduce discontinuities and double minima for the MLE in the minimisation of the profile likelihood ratio. The parameter values estimated by the fit could then be ambiguous; their value not necessarily reflecting the parameter set describing the model best. The affected systematic uncertainties are symmetrised by using the maximum variation from the nominal value in each bin as symmetric uncertainty.

Pruning Technically, the nuisance parameters (NP) are split into components affecting the normalisation or the shape of a given histogram. Each component has to fulfil several conditions to be considered in the fit. Thresholds on these pruning criteria are chosen such that only negligible systematics are removed. This reduction is performed to avoid noise in the fitting procedure. NPs which affect only the event-by-event weight[1] are exempt from the pruning.

Firstly, all normalisation components of a given NP are removed if they have an impact smaller than 0.5% on the event yield or smaller than the statistical uncertainty of the respective sample.

Then, several criteria are applied to the shape components. These are affected by statistical fluctuations for regions with low population. Especially bin-by-bin migration can lead to large shifts in the bin-wise variations, overemphasising their actual impact. Hence, any shape component is pruned if their total statistical uncertainty exceed 10%.

A further reduction of shape components is achieved by comparing the varied histograms and rejecting those which do not yield a significant variation, quantified by a χ^2_{red}-test.

Lastly, a bin-by-bin significance is calculated from the varied and the nominal bins and the total statistical uncertainty of all background components in the respective bin. Shape components, which show a small overall significance are pruned.

None of the above step had a large impact on the expected uncertainty of the signal strength μ.

Smoothing Given the small yields in the signal regions, the fit-model is further simplified by smoothing variations arising from shape components. This smoothing relies on a quadratic extrapolation [102] and is applied on the ratio of varied and nominal histograms to ensure only effects from statistical noise are treated.

8.3.5 Binning Optimisation

The presented maximum profile likelihood fit is performed on binned input distributions. In the limit of infinite event numbers in the signal categories, the shapes of signal and

---

[1] Such as the normalisation weights arising from simulation.

background distributions are continuous curves with maximal discrimination power. Given the statistical limitations of this analysis, a compromise between fine binning for separable signal and background shapes and statistical uncertainty on the background model is determined with the strategies explained in the following. This is in particular important for the robustness of the fitting procedure and background estimation. A highly granulated binning increases the complexity of the profile likelihood function and potentially results in empty bins, given the limited event numbers of the background model. The latter situation would lead to an overestimation of the expected significance.

For each, the cut-based and multivariate analyses, dedicated algorithms were designed and applied independently for each signal category.
For both discriminants, the chosen algorithms result in a non-uniform binning with fine bins in signal- or background-enriched regions and coarser binning in lowly populated areas.

CBA The m_{MMC} distribution binning optimisation in CBA is based on a maximisation of the expected significance $\mathcal{Z}_{\mathrm{exp}}$ including all systematic and statistical uncertainties.
Initially, a fine binning of 5 GeV is chosen for all five CBA signal regions. For each of the three (two) CBA VBF (Boosted) signal regions, the binning is varied and $\mathcal{Z}_{\mathrm{exp}}$ is evaluated on the combined respective category.
An inspection of Fig. 8.1 reveals that the bulk of the signal populates the region $100\,\mathrm{GeV} < m_{\mathrm{MMC}} < 150\,\mathrm{GeV}$ with tails of the background contributions extend to the sideband regions. Given the low population of the latter, the optimisation algorithm divides the full considered spectrum $m_{\mathrm{MMC}} < 200\,\mathrm{GeV}$ in three regions. Dedicated scans, varying either bin edges or numbers in each region and estimating $\mathcal{Z}_{\mathrm{exp}}$ for each scan point are performed in the following order:

1. In the low-mass region $m_{\mathrm{MMC}} < 100\,\mathrm{GeV}$, the width of the lowest bin is increased in 5 GeV steps until a decrease in $\mathcal{Z}_{\mathrm{exp}}$ is observed.
2. Similarly, in the mass sideband $150\,\mathrm{GeV} < m_{\mathrm{MMC}} < 200\,\mathrm{GeV}$, the highest bin's size is successively increased until $\mathcal{Z}_{\mathrm{exp}}$ drops.
3. The binning of the remaining, signal-populated region is optimised by simultaneously increasing the bin width of all bins in succession. An optimum is determined from the widest possible width not significantly impacting $\mathcal{Z}_{\mathrm{exp}}$.

Each step subsequently uses the result of the previous scan. Background modelling and fit stability were validated for the resulting binning, which is applied on the distributions in Fig. 9.7-9.8. For CBA VBF, the optimal binning was found to be almost identical between all three signal regions and was hence harmonised.

MVA A characteristic of the BDT score distributions in the multivariate approach is the dense population of signal (background) events at high (low) values. While the background dominated region provides valuable information about nuisance parameters and normalisations, regions with high scores are key to the overall sensitivity.
A two-stage algorithm was designed to determine a binning yielding a significance close to

the maximum but respecting certain requirements on the background model robustness. Starting from a binning fine enough to resolve single events, the first stage applies certain criteria on the background distributions binning. The algorithm starts from the maximum score and combines the initial, creates a new bin from the initial fine bins if the following criteria are met:

- the relative statistical uncertainty per bin for $Z/\gamma^* \to \tau\tau$, Fake and Others background contributions is smaller than 100%
- the expected total background yield per bin exceeds the one in the previous bin
- two criteria targeting for the high/low score regions:
  - the ratio of signal to total background in a new bin is at least 90% of the ratio in the previous bin or
  - the expected total background in the new bin is 50% higher than in the previous one.

Combined, these criteria ensure a stable fit and facilitate the final background normalisation estimation. It was however observed, that the sensitivity is artificially decreased in this approach, as in particular a finer binning at high BDT score is possible without affecting the fit performance.

This motivated the second stage of the optimisation procedure. Fixing the binning to the result of the first stage as starting point, the region with $> 70\%$ of the expected signal events is determined. All edges of the initial binning in this region are subsequently varied, namely three (four) for Boosted (VBF) categories, respectively. For each variation, $\mathcal{Z}_{\text{exp}}$ is evaluated by performing the fit including all systematic uncertainties. The optimal binning is chosen from scan plateaus close to the maximum value for each bin edge.

Background model and fit stability were verified for the chosen binning, shown in Fig. 8.13 and Fig. 9.6, in the MVA VBF and Boosted categories separately

CHAPTER 9

Results

This chapter summarises the results for the measurement of the signal strength parameter μ introduced in Eq. 8.1 and the cross section $\sigma_{H\tau_{\mathrm{had}}\tau_{\mathrm{had}}}$.
First, the binned maximum likelihood fit is performed on the Asimov dataset introduced in Sec. 8.3.1. This serves as validation of the fit-model and provides valuable insights into the expected sensitivity and the impact of nuisance parameters (NP).
Thereafter, the compatibility between background-only hypothesis H_0 and observed data is tested by performing a fit under the assumption of $\mu = 0$ to the recorded dataset.
Finally, an unconditional fit is performed to measure the best-fit signal strength $\hat{\mu}$ which maximises the likelihood function for the Higgs boson mass hypothesis $m_H = 125\,\mathrm{GeV}$.

Both the cut-based and multivariate analysis approaches are presented in parallel, providing an additional cross-check of the measurement's validity.

9.1 Fit-Model Validation and Expected Results

Extensive tests of the minimisation procedure and fit model are performed, in particular probing for multiple minima. These would indicate degenerate model parameter settings maximising the likelihood function and lead to ambiguous measurement results.
Furthermore, the fit-model is scrutinised for

- impact of each NP on the expected signal strength, to understand whether irrelevant uncertainties have a large impact,
- constraints on the uncertainty of a given NP,
- unexpected correlations between the NPs,
- convergence of the fit and fitted normalisations in signal depleted regions.

All tests confirmed a well-behaving fit and are outlined in the following.

9.1.1 Likelihood Function Dependencies

To exclude multiple minima and validate the parabolic shape of the Likelihood function, the profile of minimal negative log-likelihood (NLL) values is estimated for each parameter. Using the Asimov Dataset, the nominal value for a given parameter is fixed to a value within $\pm 2.5\sigma$ of its input uncertainty σ. For each point in this grid, the maximum likelihood estimator (MLE) is estimated and the difference of the resulting NLL value to the original, unshifted result is determined. This results in parabolas as presented in Fig. 9.1, centred at the unconditional MLE. Two examples of parameter scans are presented here, the dependence on the signal strength μ and on two important NPs - the uncertainty on the jet energy resolution for the CBA and the NP encoding the uncertainty

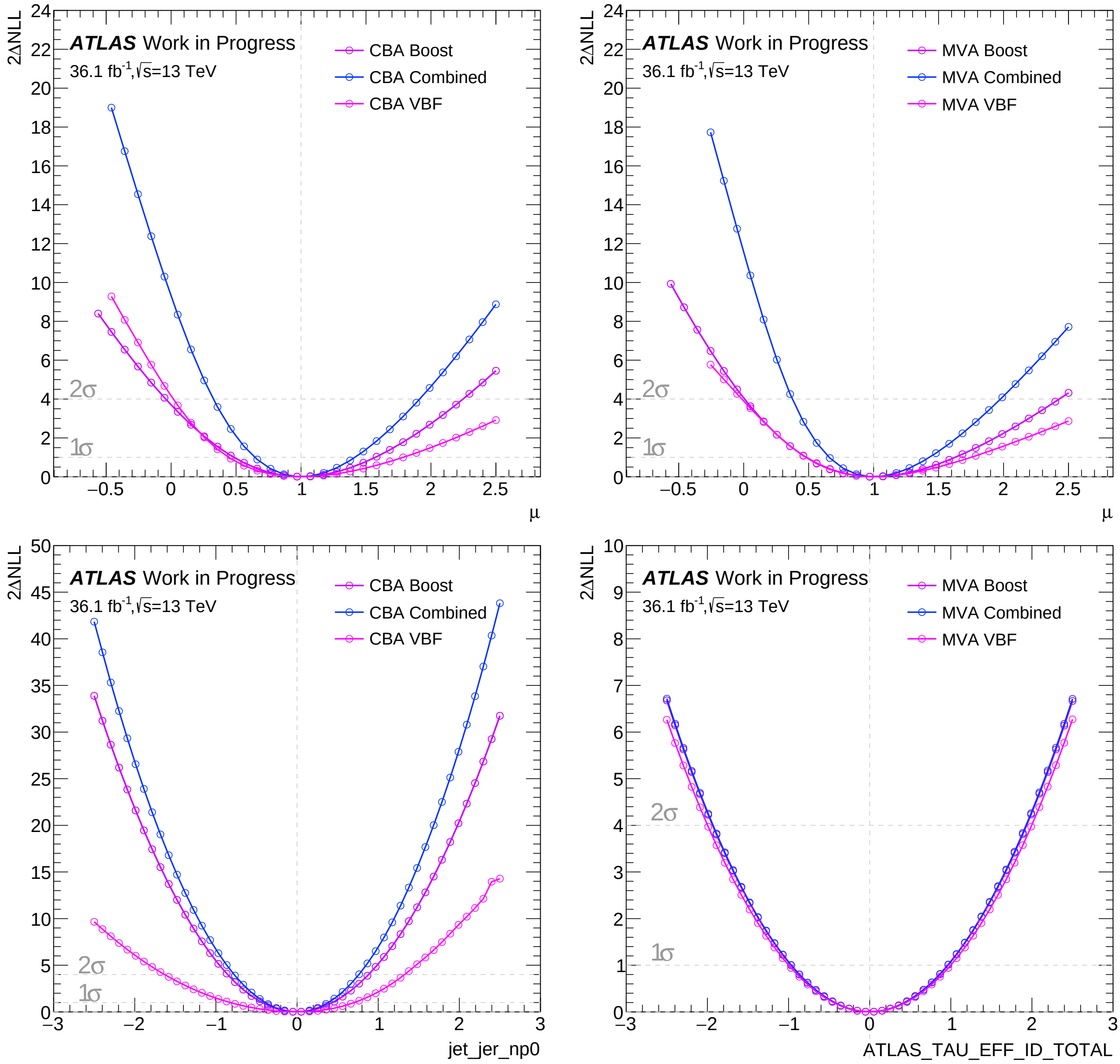


Figure 9.1: Selected negative log-likelihood NLL distributions for CBA and MVA. Smooth parabolic shapes are observed around the minimum value.

on the identification efficiency for τ_{had} for MVA.
The profiles exhibit a parabolic shape near the global minimum.

9.1.2 Nuisance Parameter Constraints

A similar test is performed to estimate the impact of each NP on the signal strength μ. The initial nominal value of a given NP θ_0 is shifted to its $\pm 1\sigma$ uncertainty, and the minimisation is repeated for all other parameters of the model. The difference of the resulting $\hat{\mu}$ to the unconditional MLE $\Delta\mu$ determines the effect of a given uncertainty on the measurement and is used to rank NPs according to their importance for the measurement.
Furthermore, it is useful to display the measured shift of each NP for the unconditional MLE to its nominal value $(\hat{\theta} - \theta_0)$ normalised to the nuisance parameter's pre-fit uncertainty σ_θ. This is referred to as pull. This shift reveals whether the fitted background model prefers NP values differing from the initial one to describe the observed dataset. In addition, the fit provides insight on the uncertainty of a NP. With a NP shifted to $\pm 1\sigma_\theta$, the impact of its uncertainty is reflected in the shift of the model response. In case of an unconstrained uncertainty, as shown in the bottom right of Fig. 9.1, the $\pm 1\sigma$ uncertainty corresponds to a $\pm 1\sigma$ difference in the model response, i.e. the NLL minimum. If a constraint is present, the $\pm 1\sigma_\theta$ uncertainty leads to larger shifts in the model response, as for example shown in the bottom left of Fig. 9.1. In simple terms this means, that the pre-fit uncertainty estimate of a given NP is overly conservative for the fitted dataset, i.e. that a higher precision for the NP in question can be achieved with the measurement at hand. This can in particular happen for experimental uncertainties included in the Likelihood function relying on external measurements.

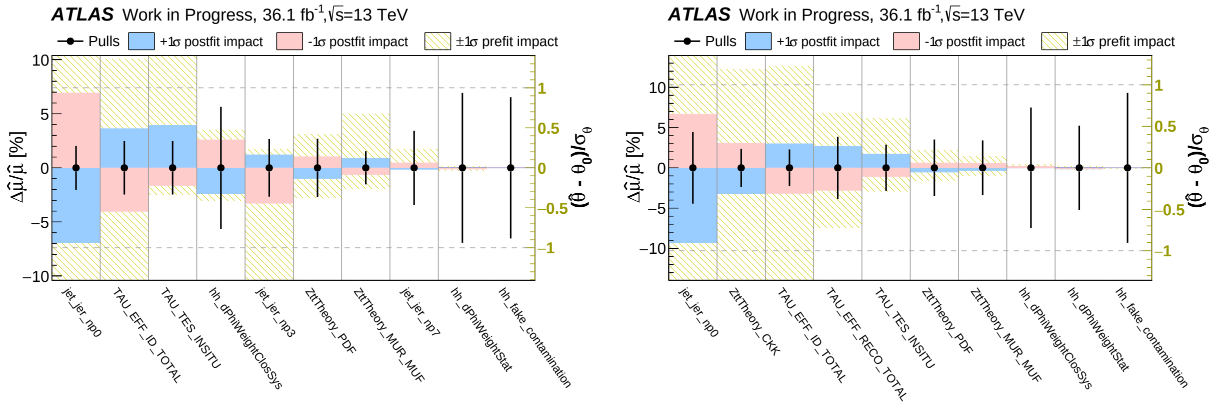


Figure 9.2: Visualisation of the seven most constrained nuisance parameters in CBA and MVA. In addition, the introduced systematic uncertainties on the Fake background model are shown.

In Fig. 9.2, the 7 NP with highest constrained uncertainties and, in addition, the uncertainties on the Fake background model are shown for CBA (left) and MVA (right) for a fit to Asimov data. The NPs are ranked from left to right after decreasing $\Delta\mu$ and hashed yellow bands demonstrate the impact of $\pm\sigma_\theta$ on the signal strength. Blue and red boxes similarly show the impact on μ but for the post-fit uncertainty $\pm\hat{\sigma_\theta}$, which, in case of the selected, constrained NPs, is smaller than σ_θ. The black markers represent the pulls, in case of the presented Asimov Dataset fixed at 1 by construction. The latter are shown with the post-fit uncertainty (black line) which can be compared to the pre-fit $\pm\sigma_\theta$ uncertainty (black hashed horizontal line). It can be seen, that the independent fits for CBA and MVA show similar NP constraints, indicating that these originate from conservative σ_θ.

For both, CBA and MVA the uncertainties on the jet energy resolution are constrained. Detailed checks documented in [125] have shown that this constraint originate from sizeable statistical uncertainties of the corresponding auxiliary measurement [115].

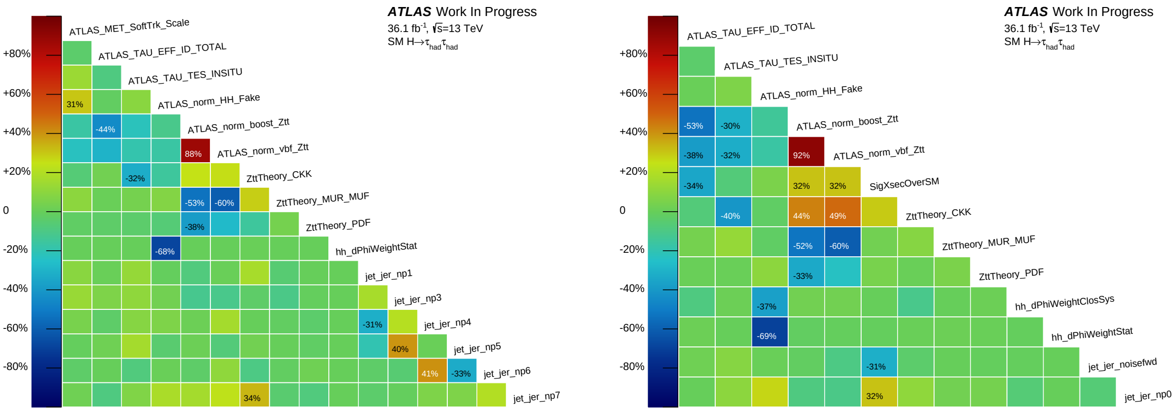
The uncertainties on the τ_{had} identification, reconstruction and calibration are constrained as well. Measurements of the latter are based on the 2015 dataset [12]. As the $\tau_{\text{had}}\tau_{\text{had}}$ channel probes very specific phase spaces in the VBF and Boosted categories for the full dataset, these constraints are expected.

Finally, constraints on uncertainties arising from the prediction of the $Z/\gamma^*\!\to\tau\tau$ background are caused by the maximisation approach when combining the individual components, as was discussed in Sec. 7.6.

The NPs constructed for the $\tau_{\text{had}}\tau_{\text{had}}$ Fake background model are also presented in Fig. 9.2 and do not show strong constraints.

9.1.3 Correlations of Nuisance Parameters

The HistFactory framework provides the possibility to estimate the correlations between individual NPs during the minimisation. A large (anti-)correlation between two given NPs indicates that they describe a similar physical effect on the measurement. In Fig. 9.3, the matrix for all NPs which are correlated to at least one other NP by more than 30% is



CBA MVA

Figure 9.3: Correlations between nuisance parameters for CBA (left) and MVA (right). Shown are the NP of the model, which are at least more than 30% correlated to any other NP. The NP naming is explained in Appendix C.

presented. Here, normalisation, theoretical and experimental NPs are considered as well as the parameter of interest, μ. In general, NPs with a large impact on the simulated $Z/\gamma^* \to \tau\tau$ process normalisation and shape are correlated, such as the τ_{had} uncertainties and the theoretical NPs. A high correlation is observed only for the individual $Z/\gamma^* \to \tau\tau$ normalisation factors in the $\tau_{\text{had}}\tau_{\text{had}}$ fit-model for the VBF and Boosted categories, respectively. This is evident, as an underestimation of the $Z/\gamma^* \to \tau\tau$ contribution affects both categories similarly.

9.1.4 Fit Convergence in a Signal-depleted Region

A further validation fit was performed for the CBA and relies on the observed dataset in the mass sidebands MMC $< 100\,\text{GeV}$ and $m_{\text{MMC}} > 150\,\text{GeV}$. This region is hence depleted of signal and dominated by background contributions. The unconditional MLE confirmed a signal strength of $\mu \approx 0$, as expected for the background-only hypothesis. Pulls in this region, in particular for the jet energy resolution parameters, closely reflect the situation for the unconditional fit on the complete signal regions, which is discussed in the following section. In particular, the free-floating normalisation NPs for Fake background and $Z/\gamma^* \to \tau\tau$ processes preferred in these regions are determined and found to be compatible with 1 within their uncertainties. This validates their chosen starting values estimated as described in Sec. 7.4.

9.2 $H \to \tau_{\text{had}}\tau_{\text{had}}$ Signal Strength Measurement

In order to determine the signal strength parameter μ, an unconditional NLL minimisation is performed to the full observed dataset. This fit adjusts all normalisation factors and systematic uncertainties to the preferred values in the selected dataset, separately using the full m_{MMC} distributions for the CBA and the two BDT score distributions for the MVA signal regions. Results are quoted for a combined fit on VBF and Boosted categories as well as for a standalone fit in each inclusive signal region.

As the analysed events for the CBA are a subset of the MVA, the results of both methods are correlated. Due to the different analysis strategies however, the results provide an independent cross-checks, as will be emphasised in the following sections. The resulting best estimate for the parameter of interest, $\hat{\mu}$, assuming the signal hypothesis H_1 with $m_H = 125\,\text{GeV}$ is presented in Sec. 9.2.4.

9.2.1 Impact of Nuisance Parameters

To better understand the adjustment of the model in the fit, the impact and pull of all nuisance parameters is investigated with the procedure described in Sec. 9.1.2.

In Fig. 9.4, the 25 NP with highest impact on $\hat{\mu}$ are visualised. On top of the already discussed plot entries, green markers indicate the normalisation factors for the Fake background and $Z/\gamma^* \to \tau\tau$ processes. The pre-fit normalisation error presented in Sec. 7.4

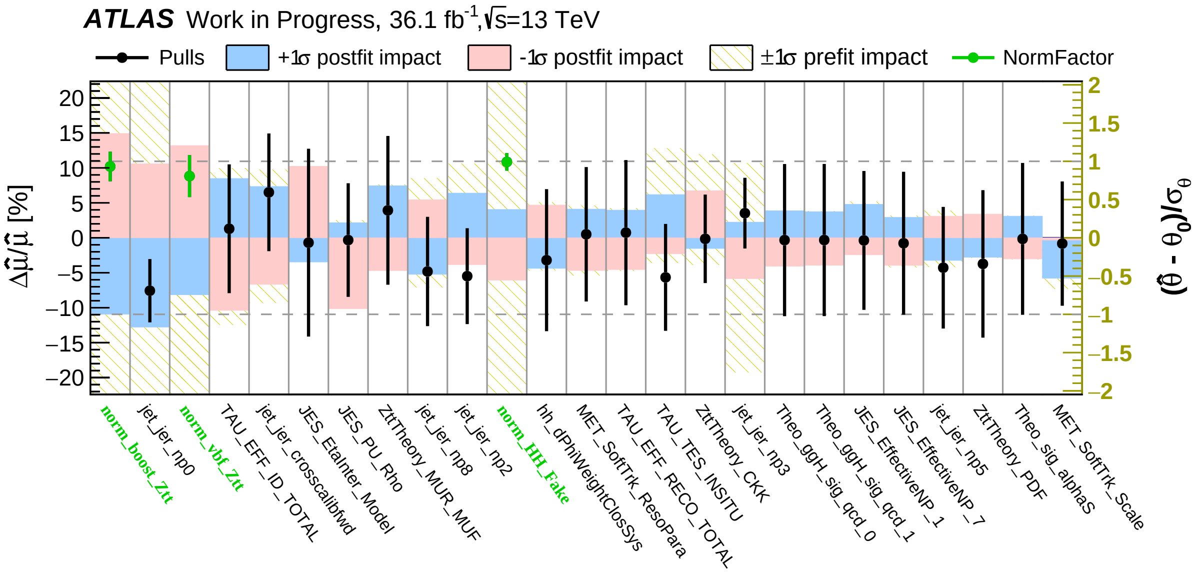


Figure 9.4: Ranking of the 25 NP with highest impact for CBA. The nuisance parameter naming is explained in Appendix C.

is not taken into account in the fit, to allow for full freedom in adjustment of the normalisation NPs. The constraint seen on the post-fit error does therefore not represent any real effect, as no physical $\pm 1\sigma_\theta$ pre-fit variation was defined.
The $Z/\gamma^* \to \tau\tau$ normalisation in both regions exhibit a large impact between $10 - 15\%$ on $\hat{\mu}$, indicating the presence and importance of $Z/\gamma^* \to \tau\tau$ events in the signal mass region. These events are also strongly affected by the jet energy resolution parameters and the τ_{had} identification efficiency, in particular in the VBF categories. Based on the aforementioned limitations of the auxillary measurement of this parameter, the observed pull reflects the optimal adjustments in the $\tau_{\text{had}}\tau_{\text{had}}$ phase space.

Similarly, the NP ranking for the MVA is presented in Fig. 9.5, exhibiting a $\sim 10\%$ impact of the two aforementioned experimental NP. In contrast to the CBA, the jet energy resolution parameter is not strongly constrained in the MVA. This NP acts strongly on the shape of the m_{MMC} distribution, in particular in the tails of the $Z/\gamma^* \to \tau\tau$ background. As the BDT score itself does not carry any shape information of the invariant $\tau_{\text{had}}\tau_{\text{had}}$ mass, the constraint is smaller, as the fit lacks the constraining shape information present in the CBA observable.
The smaller impact of normalisation factors in MVA can also be explained by the different observables in each analysis. For the MMC, the signal is spread out over a larger $Z/\gamma^* \to \tau\tau$ contaminated range. A small shift in the normalisation then affects multiple signal sensitive bins strongly. For the BDT score, the majority of signal populates the last two bins. These exhibit only a very small background contamination, therefore small normalisation changes do not affect the measured signal strength in these strongly.
The post-fit normalisations for the unconditional MLE are summarised in Tab. 9.1 and agree with the pre-fit values and across the analysis approaches within uncertainties. Differences for the Fake background normalisation are caused by the variation of the

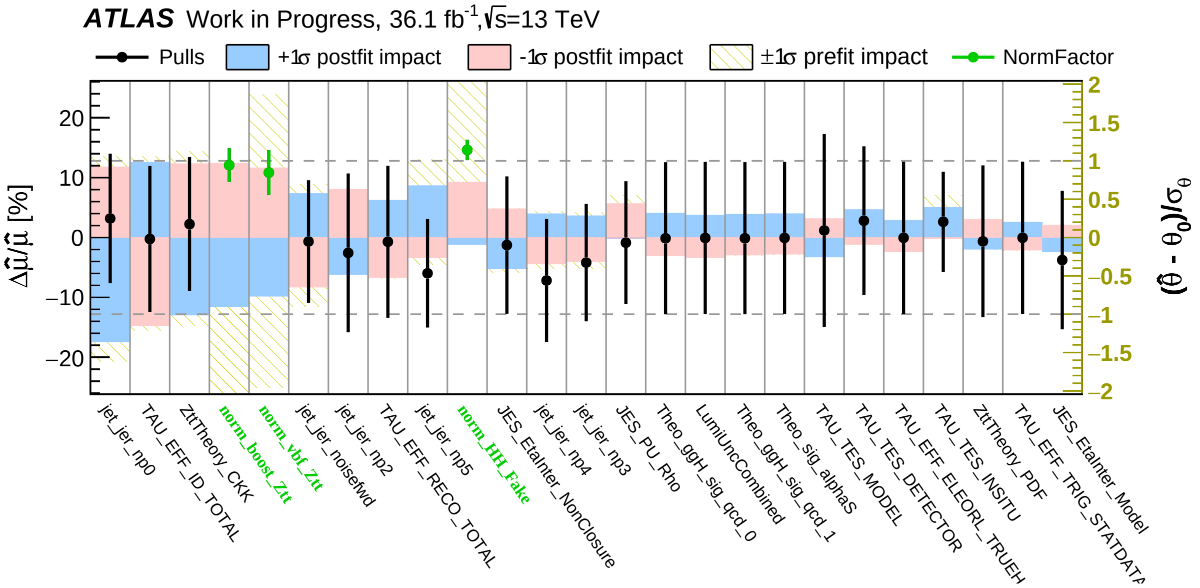


Figure 9.5: Ranking of the 25 NP with highest impact for MVA. The nuisance parameter naming is explained in Appendix C.

hh_dPhiWeightClosSys, one of the Fake background template reweighting uncertainties, as can be seen in the middle part of Fig. 9.4. This uncertainty was derived based on the m_{MMC} distribution as explained in Sec. 7.3.2.3.

Normalisation Factor	CBA	MVA
$\hat{r}_Z^{\mathrm{VBF}}$	0.81 ± 0.27	0.85 ± 0.29
$\hat{r}_Z^{\mathrm{Boost}}$	0.93 ± 0.19	0.94 ± 0.22
$\hat{r}_{\mathrm{QCD}}$	0.99 ± 0.11	1.14 ± 0.13

Table 9.1: Normalisation Factors for $Z/\gamma^* \to \tau\tau$ in VBF and Boosted categories respectively and for the Fake background for both CBA and MVA estimated in the Likelihood maximisation.

9.2.2 Compatibility to Background-only Hypothesis

As was discussed in Sec. 8.3.2, the compatibility between a given dataset and the background-only hypothesis H_0 is estimated using the test statistic q_0. The resulting compatibility is quantified by the probability p_0 and the equivalent excess significance $\mathcal{Z}$. In Tab. 9.2, the following three figure-of-merits are shown:

- Asimov significance $\mathcal{Z}_{\mathrm{A}}$ introduced in Eq. 8.3, calculated on the Asimov dataset with nuisance parameters fixed to their nominal value,
- expected significance $\mathcal{Z}_{\mathrm{exp}}$, including the full set of NPs and evaluated in the conditional fit with $\mu = 1$,

- observed significance $\mathcal{Z}_{\mathrm{obs}}$, determined on the selected dataset in an unconditional fit and including all NPs.

Both $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ analysis approaches were optimised independently for each inclusive category[1] to yield maximal values for $\mathcal{Z}_{\mathrm{A}}$ and $\mathcal{Z}_{\mathrm{exp}}$.

	$\mathcal{Z}_{\mathrm{A}}$		$\mathcal{Z}_{\mathrm{exp}}$ $[\sigma]$		$\mathcal{Z}_{\mathrm{obs}}$ $[\sigma]$	
Region	CBA	MVA	CBA	MVA	CBA	MVA
VBF	3.81	4.88	2.04	2.54	1.74	2.83
Boosted	3.61	4.39	1.92	2.01	1.14	2.35
Combined	5.25	6.56	3.05	3.56	**2.25**	**4.15**

Table 9.2: Summary of the estimated Asimov significances and the observed (expected) excess significances expressed in units of standard deviations for CBA and MVA.

The Asimov significance gives a first indication on the sensitivity of the analyses. The combined CBA result is improved by $\sim 25\%$ using the multivariate approach. Individual results for the inclusive categories are close to each other in each analysis, albeit the VBF categories yield a slightly higher $\mathcal{Z}_{\mathrm{A}}$ due to the larger expected signal contribution.

Revealing the sizeable impact of uncertainties on the sensitivity, the expected significances are smaller than $\mathcal{Z}_{\mathrm{A}}$. However, they still exceed the threshold of claiming an evidence of a signal, 3σ - for both, CBA and MVA. The combined $\mathcal{Z}_{\mathrm{exp}}$ for MVA shows a $\sim 15\%$ improvement over the CBA result. This is mainly driven by the improvement in the VBF category, for the Boosted, both CBA and MVA significances are close. This indicates that the categorisation in CBA Boosted, relying on $\Delta R_{\tau\tau}$ and p_{T}^{H}, together with the m_{MMC}-sensitive fit is close to the optimal boosted BDT configuration chosen for the MVA. In Appendix B, it can be seen that the improvement from additional variables used in the BDT-training is minimal. The full separation information in the Boosted categories is contained in the mass and angular distance of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ system and the transverse momentum of the Higgs boson, all of them being correlated to each other and related to the boost of the decay system.
The unconditional fit yields a significance of 4.15σ for MVA and 2.25σ for CBA, corresponding to a compatibility of H_0 to the observed dataset of

$$p_0^{\mathrm{MVA}} = 1.65 \times 10^{-5} \quad \text{and}$$
$$p_0^{\mathrm{CBA}} = 0.01 \ .$$

This result is a direct evidence for the Yukawa coupling in $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ decays.

---

[1] CBA/MVA VBF and Boosted, respectively.

9.2.3 Post-fit Modelling and Yields

The unconditional MLE adjust the fitted model to optimally describe the observed dataset. Shapes and normalisations for all event model processes are adjusted and the resulting nuissance parameter settings reflect the observed data the best.
Fig. 9.7-9.8 shows the final, post-fit distributions of the m_{MMC} mass in each CBA signal region. Signal, $Z/\gamma^* \to \tau\tau$ and Fake background contributions are presented as in the pre-fit input distributions in Sec. 8.1. Other background contributions are summarised, following the $\tau_{\text{had}}\tau_{\text{had}}$ fit strategy. Furthermore, the H_1 pre-fit model prediction is depicted as hashed green line.
The hashed band in the top and bottom panel visualise the $\pm 1\sigma$ combined post-fit uncertainties per bin. In the bottom panel, the ratio of observed data to the fitted signal-plus-background model is shown as marker points together with the statistical uncertainty on the model.
All m_{MMC} distributions exhibit a good agreement to the H_1 hypothesis across all regions, any fluctuation is covered well within the uncertainties. The latter range between $20 - 50\%$ for the VBF categories in mass sensitive m_{MMC} regions. This is mainly driven by the statistical uncertainties due to the limited number of events passing the VBF selection. For the Boosted categories, the uncertainties are $\sim 10\%$, systematic and statistical errors of similar size along the mass sensitive bins.

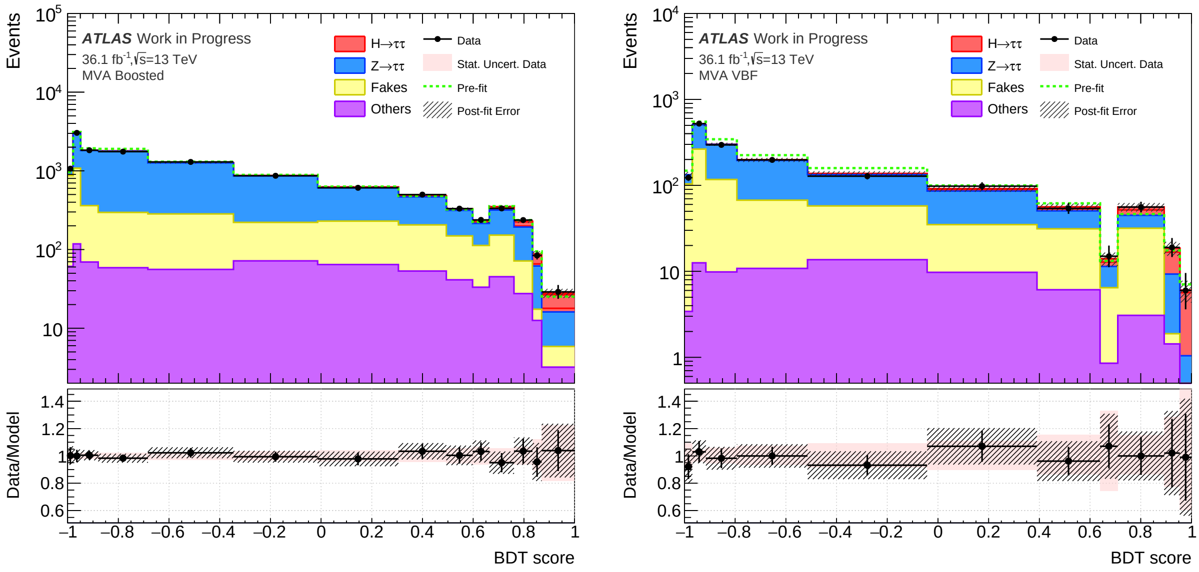


Figure 9.6: Post-fit distributions of the BDT score used as final discriminant in the MVA.

In Fig. 9.6, the two BDT score distributions for the MVA VBF and Boosted categeories are shown. A satisfying modelling similar to the input distributions in Sec. 8.2.3 is observed with uncertainties $< 10\%$ for background dominated regions. Uncertainties in the lowly populated high BDT score regions range up to 20% (60%) for the Boosted (VBF) category. These are dominated by the statistical uncertainty on the model prediction.

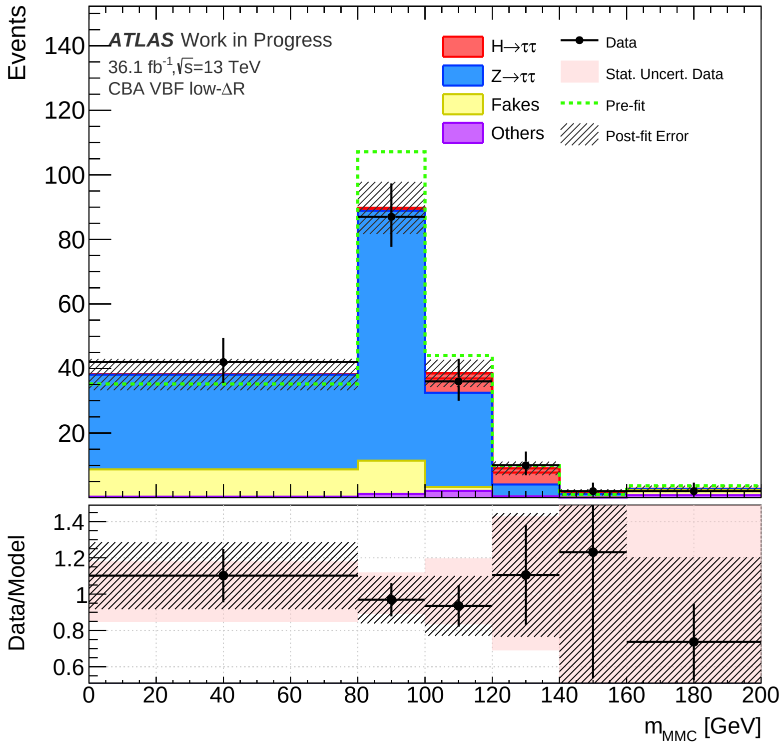


Figure 9.7: Post-fit m_{MMC} distribution for the CBA signal regions (continued in Fig. 9.8). The statistical uncertainties on the Data points per bin are indicated with lines in the top panel and solid band in the ratio. The green line depicts the pre-fit prediction for H_1. The hashed band represents the total post-fit uncertainty on the event model. Markers in the bottom panel compare the post-fit model prediction for H_1 with the observed Data.

A clear excess for both approaches can be seen in the signal-sensitive region. The excess agrees well with the SM prediction as will be shown in the following sections.

	Region	$Z \to \tau\tau$	Fakes	Others	$H \to \tau\tau$	Total	Data
CBA	VBF Low ΔR	92.2±12.5	106.1±11.2	12.1±2.3	7.1±2.7	217.6±12.2	220.0±14.8
	VBF High ΔR Tight	141.0±13.9	21.6±4.2	4.7±1.1	12.4±4.3	179.8±13.0	179.0±13.4
	VBF High ΔR Loose	61.1±10.1	49.5±6.5	4.8±1.8	3.0±1.1	118.4±9.7	121.0±11.0
	Boosted Tight	3161.5±160.1	1924.6±169.6	325.2±36.3	49.4±16.6	5460.7±78.0	5455.0±73.9
	Boosted Loose	3569.3±93.2	334.6±61.1	132.2±17.0	59.7±19.9	4095.8±70.9	4103.0±64.1
MVA	VBF	755.2±74.8	645.5±56.7	71.7±17.2	47.8±10.9	1520.2±42.3	1516.0±38.9
	Boosted	7960.9±307.0	3342.9±290.5	715.9±88.5	187.0±44.6	12206.7±129.9	12211.0±110.5

Table 9.3: Post-fit event yields for MVA and CBA. The total sum corresponds to the post-fit H_1 hypothesis.

A further quantification of this agreement is provided in Tab. 9.3, where the post-fit event yields are summarised.

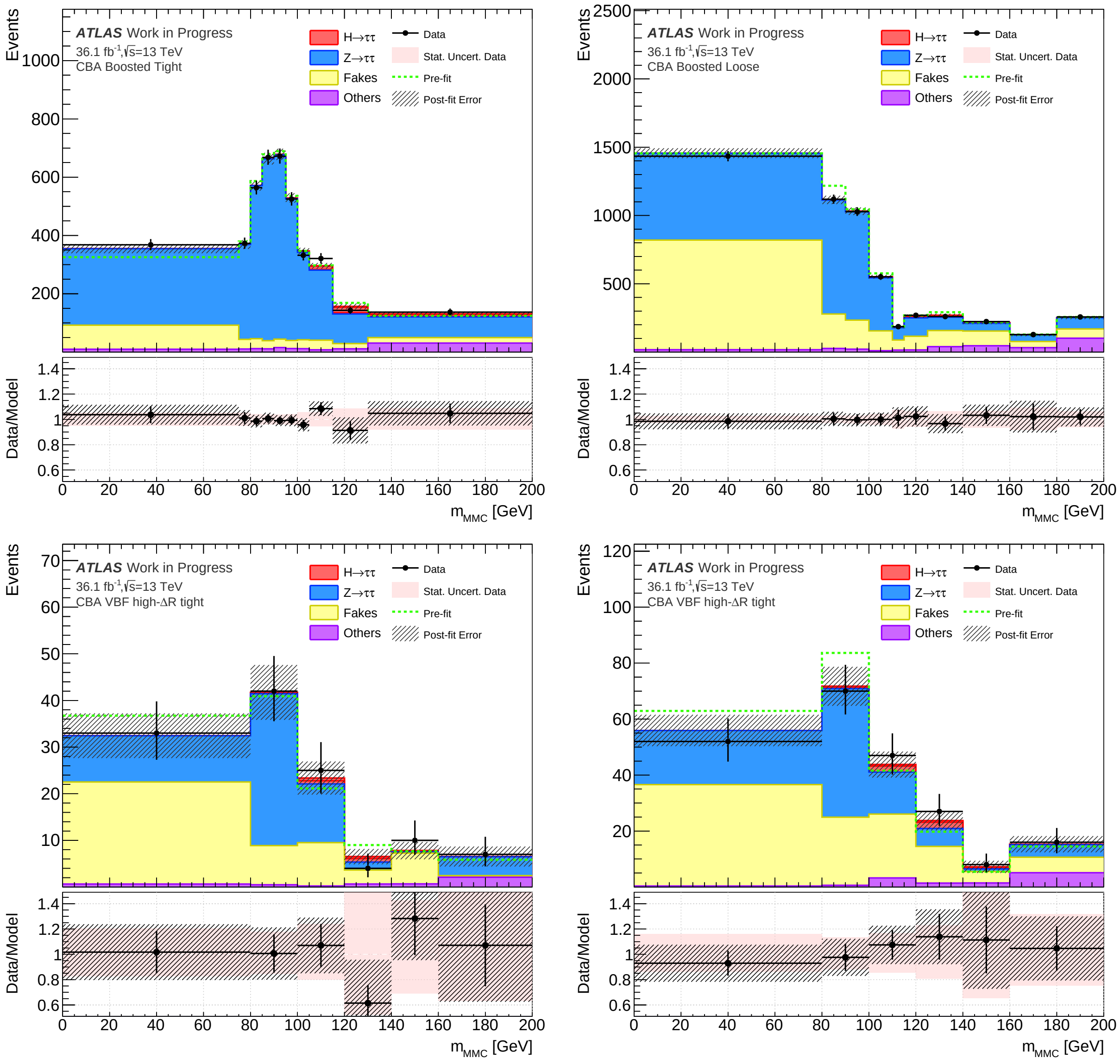


Figure 9.8: Post-fit m_{MMC} distributions for the CBA continued from Fig. 9.7.

9.2.4 Signal Strength and Cross Section Measurement

The unconditional MLE provides the best estimate for the signal strength $\hat{\mu}$ and therefore a measurement of the Standard Model Higgs boson cross section times branching ratio in $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ decays for a mass hypothesis of $m_H = 125\,\mathrm{GeV}$.

For both analysis approaches, the signal strength measurement is performed for the combined categories and in the VBF and Boosted inclusive categories independently, allowing for a cross check of the respective results.
The measured values for the combined categories in cut-based and multivariate analysis are

$$\hat{\mu}^{\mathrm{MVA}} = 1.15 \pm 0.39$$
$$\hat{\mu}^{\mathrm{CBA}} = 0.73 \pm 0.34$$

quoted with the total, combined statistical and systematic uncertainty. The values and uncertainties are well compatible with each other and with $\mu = 1$, confirming the Standard Model prediction. The measurements in the individual categories for both analyses are summarised in Fig. 9.9, revealing no large fluctuation per category.

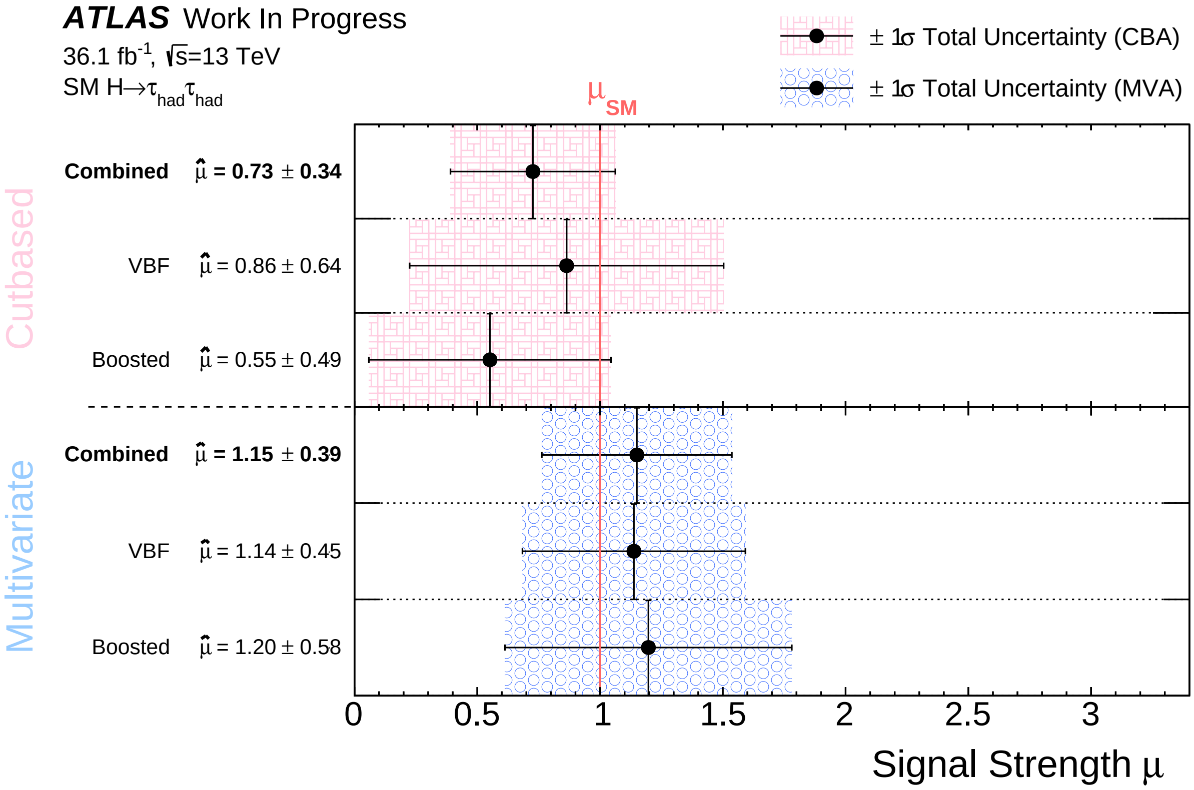


Figure 9.9: Measured $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ signal strengths in cut-based and multivariate analysis.

Cross Section Measurement In order to estimate the total Higgs boson production cross section times branching ratio of all production modes in the $\tau_{\text{had}}\tau_{\text{had}}$ final state, the measurement of the signal strength muhat is reinterpreted. By following Eq. 8.1, the multiplication of $\hat{\mu}$ with the SM prediction for the cross section times branching ratio yields

$$\begin{aligned}\sigma^{\text{tot,obs}}_{H\tau_{\text{had}}\tau_{\text{had}}} &\equiv \left[\sigma^{\text{tot}}(pp \to H) \times \text{BR}(H \to \tau_{\text{had}}\tau_{\text{had}})\right]^{\text{obs}} \\ &= \hat{\mu}\left[\sigma^{\text{tot}}(pp \to H) \times \text{BR}(H \to \tau_{\text{had}}\tau_{\text{had}})\right]^{\text{SM}} .\end{aligned}$$

The theoretical values for the cross section of the simulated signal samples are summarised in Tab. 9.4 and based on the calculations described in Sec. 7.2. Higgs decays into a pair of τ leptons have a branching ratio of $BR(H \to \tau\tau) = 0.06256$ [54]. Finally, τ leptons decay hadronically with a branching ratio of 0.6479 [27] as introduced in Sec. 5.4.1.

Production Modes		$\sigma^{\text{SM}}_{H\tau\tau}$ [pb]	$\sigma^{\text{SM}}_{H\tau_{\text{had}}\tau_{\text{had}}}$ [pb]
targeted	ggH	3.035	1.274
	VBFH	0.236	0.099
other	WH	0.086	0.036
	ZH	0.055	0.023
	$t\bar{t}H$	0.053	0.022
total	$pp \to H$	3.465	1.454

Table 9.4: Summary of the cross sections of the Higgs boson production modes times branching ratio of Higgs decays into a pair auf τ leptons considered in this thesis. The second column shows the values for subsequent hadronically decaying tau pairs $\tau_{\text{had}}\tau_{\text{had}}$. *Targeted* indicates the two production modes for which the presented analyses were optimised, *other* the additional considered signal contributions.

With these ingredients and the results for $\hat{\mu}$, the measured cross sections times branching ratios for the cutbased analysis are

$$\sigma^{\text{CBA}}_{H\tau_{\text{had}}\tau_{\text{had}}} = 1.06 \pm 0.49\,\text{pb}$$

and for the multivariate analysis approach

$$\sigma^{\text{MVA}}_{H\tau_{\text{had}}\tau_{\text{had}}} = 1.67 \pm 0.57\,\text{pb} .$$

Both results are quoted with the total uncertainty, inluding all respective statistical, systematic and theoretical components. The resulting values are consistent with the SM prediction and compatible with each other.

9.3 Combination

As discussed in Sec. 6.1, the measurement of the $H \to \tau\tau$ signal strength in ATLAS is performed in a combination of the three analysis channels

- H $\to \tau_\ell\tau_\ell$, targeting leptonic-leptonic decays of the $\tau\tau$ pair produced in Higgs boson decays with branching ratio $\text{BR}(\tau_\text{lep}\tau_\text{lep}) = 12.4\%$,
- H $\to \tau_\ell\tau_\text{had}$, targeting leptonic-hadronic decays with $\text{BR}(\tau_\text{lep}\tau_\text{lep}) = 45.6\%$,
- $H \to \tau_\text{had}\tau_\text{had}$, the fully hadronic decay channel presented in this thesis with $\text{BR}(\tau_\text{lep}\tau_\text{lep}) = 42.0\%$.

Therefore, the leptonic-hadronic final state provides the highest rate of events amongst the three channels, closely followed by $\tau_\text{had}\tau_\text{had}$.
Decays involving leptons $\ell = e, \mu$ benefit from the higher trigger and identification efficiencies compared to the fully hadronic channel. As discussed in Sec. 5.6, the presence of an additional neutrino ν_ℓ worsens the MMC-algorithm based $\tau\tau$ mass resolution. The presence of a lepton ℓ furthermore implies additional background sources to the irreducible $Z/\gamma^*\to\tau\tau$ contribution present in all channels.. The $\tau_\text{lep}\tau_\text{lep}$ channel involves final states with considerable $Z/\gamma^* \to \ell\ell$, W boson and top-quark related background contributions. The latter two also play an important role for the $\tau_\text{lep}\tau_\text{had}$ background composition. Independent background estimates were designed for each channel and are detailed in [125].
Each channel provides distinct signatures related to the VBFH and ggH Higgs boson production modes. Different sets of variables are used to define the VBF and Boosted signal regions. For each channel, discriminants in the signal regions are built separately, in particular for the MVA. The BDT training in each channel is based on the same initial variable list. This list is reduced independently using similar strategies to Sec. 8.2.2.3.

9.3.1 Combined Fit-Model

During the event selection orthogonality between the analysis is ensured, facilitating a combined binned likelihood fit. The same approach to perform the statistical testing is chosen, but with modifications to the fit-model in order to fully exploit the available, combined information. In Fig. 9.10, the combined fit-model is presented for the cut-based approach. The MVA shares the same strategy but only uses the inclusive signal regions. The $\tau_\text{lep}\tau_\text{had}$ and $\tau_\text{lep}\tau_\text{lep}$ background compositions allow for construction of dedicated control regions, for example enriched in top-quark events by inverting a b-jet veto requirement. This allows to set constraints on the normalisation of these background components.
The advantage of the combined fit becomes apparent when considering the normalisation of the $Z/\gamma^*\to\tau\tau$ process correlated across channels. In combination, this normalisation is constrained across channels which allows for a more precise estimation.

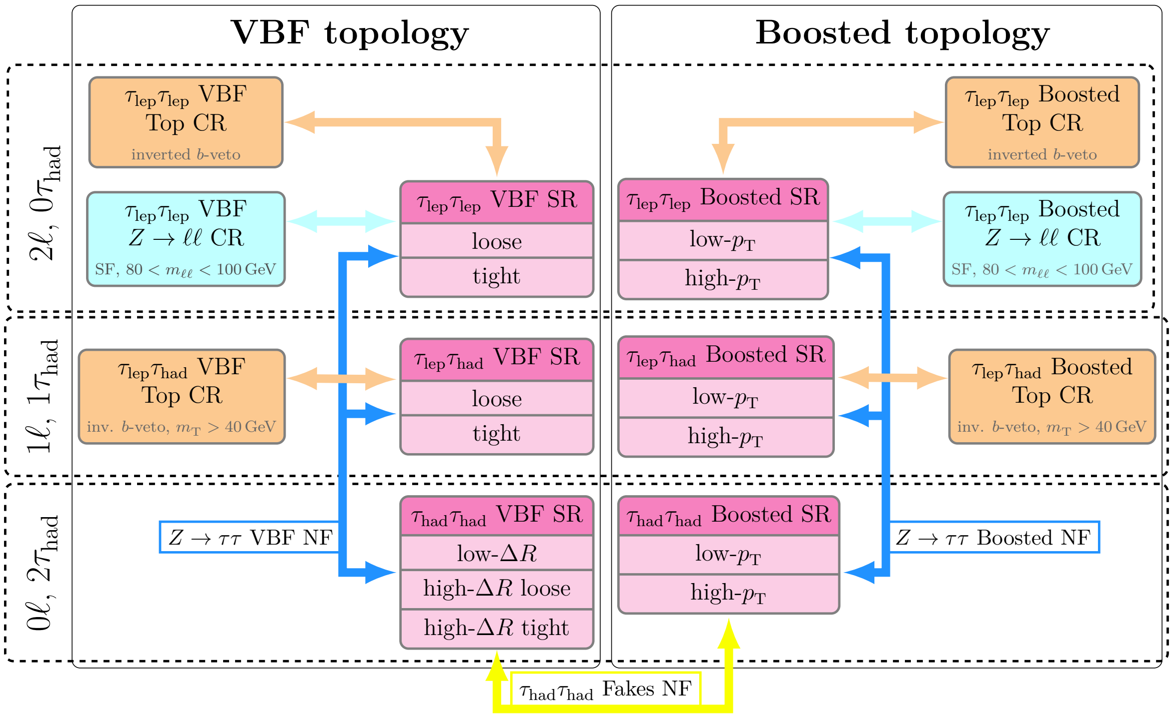


Figure 9.10: Schematic overview of the cut-based analysis fit-model. All considered regions are grouped by topology and channel. The coloured arrows represent free floating normalisations for specific background processes correlated across regions. Figure adapted from [125].

9.3.2 Expected Results

The discussion in this section is based on preliminary results in Janunary 2018.

The Asimov significance $\mathcal{Z}_{\rm exp}$ is determined in combination of the three channels and presented in Tab. 9.5. The highest sensitivity is reached in the $\tau_{\rm had}\tau_{\rm had}$ channel, with significances close to the standalone fit presented in this thesis. Currently, the $\tau_{\rm lep}\tau_{\rm had}$ channel provides comparable significances for CBA and MVA, similarly to the $\tau_{\rm lep}\tau_{\rm lep}$ channel.

	Combination	$\tau_{\rm lep}\tau_{\rm lep}$	$\tau_{\rm lep}\tau_{\rm had}$	$\tau_{\rm had}\tau_{\rm had}$
CBA $\mathcal{Z}_{\rm exp}$	4.38	1.00	2.69	2.96
MVA $\mathcal{Z}_{\rm exp}$	5.01	0.78	2.61	3.83

Table 9.5: The expected significances in the combined fit as of January 2018, for both analyses approaches and split in individual channel contributions.

CHAPTER 10

Summary and Outlook

The Standard Model of particle physics is an immensely successful theoretical framework, describing the interactions and properties of elementary particles. One of its core features is the spontaneous symmetry breaking in the electroweak sector which causes the emergence of mass terms for weak gauge bosons. The underlying mechanism predicts the existence of the Higgs boson and thereby motivates large-scale experimental efforts over the last decades.

The pinnacle of these efforts is the proton-proton collider LHC at CERN, at which four independent experiments investigate the very structure of our universe. In 2012, the discovery of a new scalar boson at an approximate mass of 125 GeV was announced by the ATLAS and CMS collaborations, marking the start of a new era in modern particle physics. All measurements performed since this first discovery are in agreement with the Standard Model nature of this Higgs boson, in particular its scalar $J^P = 0^+$ properties and coupling strengths. Searches in various decay channels and event topologies lead to observations of gluon-gluon fusion and vector-boson fusion production modes, as well as evidence of the Higgs-strahlung and associated production with top-quarks.

The majority of these measurements are driven by investigating bosonic decay channels, due to the superior background suppression in comparison to relevant fermionic decays. In the Standard Model, the latter are described by Yukawa couplings between fermions and the Higgs boson. Measurements of this Yukawa coupling are essential to validate the mechanism introducing fermion masses to the theory.

Higgs boson decays into τ lepton pairs constitutes the most sensitive channel for coupling measurements in the fermion sector and offers unique insights to the Higgs boson coupling to leptons. By exploiting characteristic jet topologies and τ decay kinematics, this channel provides the possibility to measure the Higgs boson production rates in vector-boson fusion. A direct determination of the $H \to \tau\tau$ coupling strength is complementary to top-quark related measurement, as the latter are up-type fermions.

The final states in $H \to \tau\tau$ decays are a consequence of the individual τ lepton decay modes, namely leptonically to electrons and muons and hadronically, resulting in characteristic, collimated jets. This thesis presents the measurement of the Higgs boson coupling to τ leptons in the fully hadronic final state $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and was performed in context of the combined $H \to \tau\tau$ coupling measurement, including leptonic-hadronic and fully leptonic final states. The analysed dataset was recorded with the ATLAS detector in proton-proton collisions at a centre-of-mass energy of 13 TeV between 2015 to 2016, amounting to a total of $36.1\,\mathrm{fb}^{-1}$.

The dominant backgrounds in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel arise from $Z/\gamma^{*}\to\tau\tau$ processes, which exhibit an experimental signature close to the signal processes and are therefore difficult to suppress. On top of that, multijet events introduce backgrounds from jets mimicking the experimental τ_{had} signature. By exploiting the characteristic kinematics and topologies of the signal process final states, in particular of the accompanying jets and the transverse momentum of the reconstructed Higgs boson candidate, effective handles on these backgrounds are constructed.
Two separate analysis approaches were explored, providing independent cross checks and complementary advantages. Both approaches categorise events according to the targeted production mode in VBF and Boosted signal regions, the latter enhancing contributions from ggF. As the production modes depend respectively on the couplings to weak bosons and top-quarks, the measurement in both categories provides complementary insights to the Higgs boson properties.

In the cut-based analysis (CBA) signal and background separation are optimised by using signal sensitive observables to divide the phase space according to the signal event signatures. The reconstructed $\tau\tau$ mass is chosen as final discriminator between signal and background, offering an intuitive separation due to the resonant structure of the Higgs boson decay. The cut-based analysis is hence sensitive to the Higgs boson mass within the given experimental resolution.

In order to fully explore and optimally use the complete available event information, a multivariate analysis (MVA) is designed based on a variety of kinematic and topological variables. This approach is based on boosted decision trees, which partition the available phase space to maximise signal and background separation. The analysis is optimised for a cross section measurement at the mass hypothesis of $m_H = 125\,\mathrm{GeV}$ and combines the information from many input variables into a single classifier. A higher sensitivity is reached with this strategy, in particular due to an efficient multijet background suppression.

The final, binned discriminants in each constructed signal region are independently evaluated with a statistical test based on an unconditional maximum profile likelihood fit. In this fit, all model parameters and uncertainties are adjusted to the observed data and the best estimator for the event model parameters is determined by the maximum value of the likelihood function. The parameter of interest in this fit is the signal strength μ, defined by the measured cross section times branching ratio normalised to the Standard Model prediction. This observable directly quantifies the compatibility of the measurement with the expectations and is used to measure the cross section of $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ decays.

An excess of data over the predicted background is observed in both analyses. For the CBA, the excess reaches an observed significance of 2.25σ while 3.05σ are expected. The observed (expected) excess significance in the MVA reaches $4.15\ (3.56)\sigma$, exceeding the threshold of three standard deviations and therefore constituting a direct evidence for the Higgs boson coupling to hadronically decaying τ pairs. The sensitivity in both approaches

is driven by the VBF categories, albeit closely followed by the Boosted categories. As the increased collision energy in RUN II generally leads to higher energetic decay products, the importance of the Boosted categories is increased with respect to RUN I results.
The signal strength measurement, main result of this thesis, confirms the Standard Model prediction for $m_H = 125\,\text{GeV}$ within uncertainties and is consistent between the two approaches:

$$\begin{aligned} \mu^{\text{MVA}} &= 1.15 \pm 0.39 \\ \mu^{\text{CBA}} &= 0.73 \pm 0.34 \ . \end{aligned}$$

The quoted total uncertainties are approximately equally composed of statistical and systematic uncertainty sources. Main systematic uncertainties arise from the auxiliary jet energy resolution and τ identification measurements. The similar sized uncertainties exhibited in the results indicate a similar sensitivity of both analyses. The measurement of the signal strength in VBF and Boosted categories separately validates the consistency of the result.
The corresponding ATLAS RUN I results for $\hat{\mu}$ in MVA and CBA are $\mu = 1.77^{+0.93}_{-0.71}$ and $\mu = 1.64^{+0.90}_{-0.74}$, respectively [9].
This measured value of the signal strength is reinterpreted as total cross section times branching ratio by multiplying the SM expectation. Combining all production modes and channels in each analysis approach, the values are

$$\begin{aligned} \sigma^{\text{CBA}}_{H\tau_{\text{had}}\tau_{\text{had}}} &= 1.06 \pm 0.49\,\text{pb} \\ \sigma^{\text{MVA}}_{H\tau_{\text{had}}\tau_{\text{had}}} &= 1.67 \pm 0.57\,\text{pb} \ , \end{aligned}$$

quoted with a total uncertainty composed as previously described. Currently, preparations for a combined analysis of all three $H \to \tau\tau$ decay channels are ongoing. Such a combination provides valuable additional information across channels, as common features such as normalisations and systematic uncertainties can be correlated between the channels, improving the predictive power in each channel. In addition, the combined event yields reduce statistical uncertainties on the measurement.
In the current estimate, the $\tau_{\text{had}}\tau_{\text{had}}$ channel is the most sensitive amongst the three in both analysis approaches, closely followed by the leptonic-hadronic channel.

The presented result constitutes the most precise standalone measurement of the $H \to \tau_{\text{had}}\tau_{\text{had}}$ signal strength to date and confirms the Yukawa couplings to τ leptons.

Outlook

The presented analysis will be combined with the di-leptonic and leptonic-hadronic final state analyses. These additional channels provide important possibilities to improve the measurement. Various control regions can be used to set constraints on parameters in the fit, improving the precision of the estimation. Specifically, the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel will benefit from the normalisation information for the $\mathrm{Z} \to \tau\tau$ background component provided by correlating the Z-normalisation factor in boosted and VBF categories respectively. Correlating other external experimental parameters across the channels can furthermore improve the robustness of the fitting procedure. Due to the additional, independent events, statistical uncertainties will also decrease. The analyis strategy presented in this thesis allows for additional measurements of various important quantities and re-interpretation of the observed results.
Instead of the signal strength, the cross section $\sigma_{H\tau\tau}$ can directly be used as parameter of interest in the maximum likelihood fit. By introducing a second parameter of interest, the likelihood fit can be re-evaluated to measure the relative contribution of ggF and VBF production modes to $\sigma_{H\tau\tau}$. This also facilitates the combination for such a production mode dependent cross section measurement with other Higgs decay channels.
The unbiased sensitivity of the cut-based analysis to the Higgs boson mass can be exploited by repeating the fit for different mass hypotheses. The most likely result then determines the best estimate of m_H.
A detailed comparison between cut-based and multivariate approach can be estimated by using re-sampling methods like bootstrapping.

In general, this analysis will benefit greatly from the inclusion of the complete RUN II dataset and updated measurements on auxiliary inputs.
The $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ analysis presented in this thesis can be improved by updating the background estimate for the multijet background, decreasing systematic uncertainties and including an updated measurement of the τ_{had} identifications efficiencies.
Different $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ mass reconstruction algorithms are in development to improve the good performance of the MMC algorithm. These use machine learning methods like neural networks and boosted regression trees.
A re-training of the presented BDT on samples with a larger amount of events - both, simulated and data-driven - will improve the stability and robustness of the node decisions.

The total RUN II dataset is expected to more than triple the analysed dataset in this thesis, promising exciting times ahead.

Acknowledgements

My utmost gratitude belongs to Prof. Dr. Arnulf Quadt, *der weiß wie der Hase läuft* and whose constant encouragements made this thesis possible. In particular, I would like to thank you for your integrity, the countless unique opportunities I was granted - like the prolonged research stay at Cern and the organisation of the Hasco summer school - and the invaluable lessons not only in physics but also for my social life. Thank you for letting me look beyond the keyboard, explore the universe of particle physics and acquire a vast palette of expertises.

A great bouquet of gratefulness goes to Dr. Ulla Blumenschein, who, over all the enjoyable years of collaboration, never seized to amaze me with her in-depth physics knowledge and likewise never failed to put a smile on my face. Thank you for the many interesting discussions and valuable advices.

A *plethora* of gratefulness belongs to Zinonas *"ZZ"* Zinonos, who is simply the kindest and coolest person I had the pleasure to collaborate with and with whom I celebrated many *major successes* and *big steps forward*. You are an inspiring example to me, especially in times of great stress and I thank you deeply for all the selfless support, the invaluable lessons in coding and making work exciting and enjoyable. Εύχομαι σε εσένα και στην οικογένειά σου τα καλύτερα κι ελπίζω να συνεχίσουμε να συνεργαζόμαστε και στο μέλλον.

Special thanks belong to my *Sunbro* Dr. Michel Janus, who is a *pillar of a*wesome discussions and whose patient mentoring, especially over the last months of this thesis, deepened my physics understanding significantly. Thank you for the open, understanding ears and being a reliable friend in tough times.

I am furthermore deeply grateful for the patient and relentless support I received from Dr. Christian Grefe, with whom I had many HAPPy discussions. It was a great and instructive pleasure to collaborate (and have lunch) with you during my time at Cern.

I would also like to thank Prof. Dr. Stan Lai very much for his many kind and encouraging words, helpful advices and sharing of expertise in our enjoyable group meetings.

Many others have contributed to the results in this thesis and the joyful bliss of the last years, and I thank each and every one of the kind and inspiring people I met over the course of this thesis. Specifically, I would like to thank the $H \to \tau\tau$ analysis team, including Antonio, Daniel, Elias, Pier-Olivier, Quentin and many more. I am looking forward to all the exciting results to come in the course of the next months and years.
Very special thanks go out to Dr. Christian Schillo, who, in the very early days of my PhD, inspiringly demonstrated how to work in a collaboration.
Furthermore, I would like to thank all my friendly and helpful colleagues at the II. In-

stitute of Physics in Göttingen for the countless nice chats and exciting group outings. Special thanks go to Lucie *"Sunshine"* Hamdi for the many laughter-filled hours of joyful socialising.
A particular shout-out to my friend Mary Jane for soothing frailed nerves and the countless exhilarating, creative ideas.

This thesis would have hardly been possible without the support of *ma belle* Eva Tolosa, my beacon in dark and stormy tides. Thank you for being my loving, reliable muse especially over the final months of this thesis. *Tu m'as inspiré par ton caractère dévoué, pragmatique et aimant et je suis impatient d'explorer le monde avec toi!*

Mein Studium und diese Arbeit wäre ohne die fortwährende mentale und finanzielle Unterstuetzung meiner Familie nicht möglich gewesen. Meine tiefste Dankbarkeit gebuehrt meiner lieben Mutter Katrin fuer den unentwegten Rueckhalt und den sicheren Heimathafen.

Ich widme diese Arbeit meiner Großmutter Barbara, die mir mit ihrem Durchhaltevermögen und ihrer Stärke ein einzigartiges Vorbild ist und meinem Großvater Dieter, welcher die Weichen fuer mein physikalisches Verständnis stellte.

Bibliography

[1] D. Griffiths, *Introduction to Elementary Particles*, 2^{nd} Edition, Wiley-VCH, Weinheim 2008.

[2] M. Thomson, *Modern Particle Physics*, Cambridge University Press, New York 2013.

[3] ALEPH, DELPHI, L3 and OPAL Collaborations, *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP*, Phys. Rept. **532** (2013) 119–244.

[4] CDF and DØ Collaborations, *Combined CDF and DØ search for Standard Model Higgs boson production with up to 10.0 fb^{-1} of Data*, FERMILAB-CONF-12-065-E, 2012.

[5] H1 Collaboration, *A Measurement and QCD analysis of the proton structure function $f_2(x, q^2)$ at HERA*, Nuclear Physics **B470** (1996) 3–40.

[6] O. S. Brüning et al., *LHC Design Report*, CERN-2004-003-V-1, 2004.
URL: `http://cds.cern.ch/record/782076`

[7] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1–29.

[8] S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61.

[9] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) 117.

[10] ATLAS and CMS Collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, JHEP **08** (2016) 045.

[11] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015.
URL: `https://cds.cern.ch/record/2064383`

[12] ATLAS Collaboration, *Measurement of the τ lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2017-029, 2017.
URL: `http://cds.cern.ch/record/2261772`

[13] ATLAS Collaboration, *Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb*$^{-1}$ *of pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **01** (2018) 055.

[14] F. Halzen and A. D. Martin, *Quarks and Leptons: An introductory course in modern particle physics*, Wiley, New York 1984.

[15] E. Drechsler, *Search for the Higgs boson in vector-boson fusion in* $\tau^+\tau^-$ *final states with the ATLAS experiment*, II. Physik-UniGö-MSc-2013/04, 2013. URL: `https://www.uni-goettingen.de/de/msc/medu+theses/550897.html`

[16] S. L. Glashow, *Partial symmetries of weak interactions*, Nuclear Physics **22** (1961) 579–588.

[17] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964) 585–587.

[18] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[19] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[20] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[21] C.-N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954) 191–195.

[22] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Advantages of the color octet gluon picture*, Phys. Lett. **47B** (1973) 365–368.

[23] D. J. Gross and F. Wilczek, *Ultraviolet behavior of nonabelian gauge theories*, Phys. Rev. Lett. **30** (1973) 1343–1346.

[24] H. D. Politzer, *Reliable perturbative results for strong interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349.

[25] E. Noether, *Invariant variation problems*, Transp. Theory and Stat. Phys. **1** (1971) 186–207.

[26] M. Robinson, *Symmetry and the Standard Model: Mathematics and particle physics*, Springer, New York 2011.

[27] P. D. Group, *Review of Particle Physics*, Chin. Phys. C **40** (2016) 100001.

[28] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, *QCD and Resonance Physics. Theoretical Foundations*, Nucl. Phys. **B147** (1979) 385–447.

[29] G. P. Lepage and P. B. Mackenzie, *On the viability of lattice perturbation theory*, Phys. Rev. **D48** (1993) 2250–2264.

[30] LHCb Collaboration, *Observation of the resonant character of the* $Z(4430)^-$ *state*, Phys. Rev. Lett. **112** (2014) 222002.

[31] LHCb Collaboration, *Observation of* $J/\psi p$ *Resonances Consistent with Pentaquark States in* $\Lambda_b^0 \to J/\psi K^- p$ *Decays*, Phys. Rev. Lett. **115** (2015) 072001.

[32] A. Salam, *Weak and electromagnetic interactions*, Conf. Proc. **C680519** (1968) 367–377.

[33] J. Goldstone, A. Salam and S. Weinberg, *Broken symmetries*, Phys. Rev. **127** (1962) 965–970.

[34] M. Kobayashi and T. Maskawa, *CP Violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[35] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531–533.

[36] S. L. Glashow, J. Iliopoulos and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. **D2** (1970) 1285–1292.

[37] F. Gabbiani et al., *A complete analysis of FCNC and CP constraints in general SUSY extensions of the Standard Model*, Nuclear Physics **B477** (1996) 321–352.

[38] Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562–1567.

[39] UA2 Collaboration, *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, Phys. Lett. **B122** (1983) 476–485.

[40] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at* $\sqrt{s} = 540$ *GeV*, Phys. Lett. **B122** (1983) 103–116.

[41] UA2 Collaboration, *Evidence for* $Z^0 \to e + e-$ *at the CERN anti-p p Collider*, Phys. Lett. **B129** (1983) 130–140.

[42] G. 't Hooft, *Renormalization of massless Yang-Mills fields*, Nucl. Phys. **B33** (1971) 173–199.

[43] Y. Nambu, *Quasiparticles and gauge invariance in the theory of superconductivity*, Phys. Rev. **117** (1960) 648–663.

[44] J. Goldstone, *Field Theories with superconductor solutions*, Nuovo Cim. **19** (1961) 154–164.

[45] A. Djouadi, *The anatomy of electro-weak symmetry breaking. I: The Higgs boson in the Standard Model*, Phys. Rept. **457** (2008) 1–216.

[46] P. J. Mohr, D. B. Newell and B. N. Taylor, *CODATA recommended values of the fundamental physical constants: 2014*, Rev. Mod. Phys. **88** (2016).

[47] C. L. Bennett et al., *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: final maps and results*, Astrophys. J. Suppl. **208** (2013) 20.

[48] V. C. Rubin, N. Thonnard and W. K. Ford Jr., *Rotational properties of 21 SC galaxies with a large range of luminosities and radii*, Astrophys. J. **238** (1980) 471.

[49] S. P. Martin, *A Supersymmetry primer*, Adv. Ser. Direct. HEP **18** (1998).

[50] W. Buchmuller, P. Di Bari and M. Plumacher, *Cosmic microwave background, matter - antimatter asymmetry and neutrino masses*, Nuclear Physics **B643** (2002) 367–390.

[51] ATLAS and C. Collaboration, *Searches for Beyond the Standard Model Physics at the LHC: Run1 Summary and Run2 Prospects*, CMS-CR-2015-141, 2015.
URL: `https://cds.cern.ch/record/2039454`

[52] J. Ellis, *Outstanding questions: Physics beyond the Standard Model*, Phil. Trans. Roy. Soc. Lond. **A370** (2012) 818–830.

[53] ALEPH, DELPHI, L3 and OPAL Collaborations, *Precision Electroweak Measurements and Constraints on the Standard Model*, CERN-PH-EP-2009-023, 2009.
URL: `https://cds.cern.ch/record/1222431`

[54] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, FERMILAB-FN-1025-T, 2016.
URL: `https://cds.cern.ch/record/2227475`

[55] C. Anastasiou et al., *Higgs Boson Gluon-Fusion Production in QCD at Three Loops*, Phys. Rev. Lett. **114** (2015) 212001.

[56] C. Anastasiou et al., *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, JHEP **05** (2016) 058.

[57] S. Actis et al., *NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders*, Phys. Lett. **B670** (2008) 12–17.

[58] C. Anastasiou, R. Boughezal and F. Petriello, *Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion*, JHEP **04** (2009) 003.

[59] M. Ciccolini, A. Denner and S. Dittmaier, *Strong and electroweak corrections to the production of Higgs + 2jets via weak interactions at the LHC*, Phys. Rev. Lett. **99** (2007) 161803.

[60] M. Ciccolini, A. Denner and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC*, Phys.Rev. **D77** (2008) 013002.

[61] P. Bolzoni et al., *Higgs production via vector-boson fusion at NNLO in QCD*, Phys. Rev. Lett. **105** (2010) 011801.

[62] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys. Lett. **B579** (2004) 149–156.

[63] L. Altenkamp et al., *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, JHEP **02** (2013) 078.

[64] A. Denner et al., *Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with HAWK*, JHEP **03** (2012) 075.

[65] A. Djouadi, J. Kalinowski and M. Spira, *HDECAY: A Program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56–74.

[66] A. Bredenstein et al., *Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons*, Phys. Rev. **D74** (2006) 013004.

[67] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. **B726** (2013) 120–144.

[68] ILC Collaboration, *International Linear Collider reference design report Volume 2: Physics at the ILC*, FERMILAB-DESIGN-2007-04, 2007.
URL: `https://cds.cern.ch/record/1056637`

[69] ALEPH, DELPHI, L3 and OPAL Collaborations, *Search for the Standard Model Higgs boson at LEP*, Phys. Lett. **B565** (2003) 61–75.

[70] CDF and DØ Collaborations, *Updated combination of CDF and DØ searches for Standard Model Higgs boson production with up to 10.0 fb^{-1} of Data*, FERMILAB-CONF-12-318-E, 2012.

[71] ATLAS Collaboration, *Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector*, ATLAS-CONF-2017-046, 2017.
URL: `https://cds.cern.ch/record/2273853`

[72] ATLAS Collaboration, *Modelling $Z \to \tau\tau$ processes in ATLAS with τ-embedded $Z \to \mu\mu$ data*, JINST **10** (2015) P09018.

[73] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007.

[74] A. Chodos et al., *A New Extended Model of Hadrons*, Phys. Rev. **D9** (1974) 3471–3495.

[75] C. Adloff et al., *Deep inelastic inclusive ep scattering at low x and a determination of $\alpha(s)$*, Eur.Phys.J. **C21** (2001) 33–61.

[76] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics*, Sov. Phys. JETP **46** (1977) 641–653.

[77] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244–289.

[78] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012.

[79] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31–145.

[80] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys.Rev. **D82** (2010) 074018.

[81] *LEP Design Report: Vol.2. The LEP Main Ring*, CERN-LEP-84-01, 1984.
URL: `https://cds.cern.ch/record/102083`

[82] S. Claudet et al., *Cryogenic Heat Load and Refrigeration Capacity Management at the Large Hadron Collider (LHC)*, LHC-PROJECT-Report-1171, 2008.
URL: `https://cds.cern.ch/record/1173062`

[83] ATLAS Collaboration, *ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[84] CMS Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, CERN-LHCC-2006-001, 2006.
URL: `https://cds.cern.ch/record/922757`

[85] LHCb Collaboration, *LHCb technical design report: Reoptimized detector design and performance*, CERN-LHCC-2003-030, 2003.
URL: `https://cds.cern.ch/record/630827`

[86] ALICE Collaboration, *ALICE high-momentum particle identification: Technical Design Report*, CERN-LHCC-98-019, 1998.
URL: `https://cds.cern.ch/record/381431`

[87] LHeC Collaboration, *A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector*, J. Phys. **G39** (2012) 075001.

[88] O. Bruning, H. Burkhardt and S. Myers, *The large hadron collider*, Prog. Part. Nucl. Phys. **67** (2012) 705–734.

[89] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report 1*, CERN-LHCC-99-014, 1999.
URL: `https://cds.cern.ch/record/391176`

[90] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report 2*, CERN-LHCC-99-015, 1999.
URL: `https://cds.cern.ch/record/391177`

[91] ATLAS Collaboration, *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*, ATL-PHYS-PROC-2016-104, 2016.
URL: `https://cds.cern.ch/record/2209070`

[92] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, 2010.
URL: `https://cds.cern.ch/record/1291633`

[93] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J. **C70** (2010) 787–821.

[94] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*, CERN-LHCC-96-042, 1996.
URL: `https://cds.cern.ch/record/331062`

[95] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*, ATLAS-TDR-3, 1996.
URL: `https://cds.cern.ch/record/331062`

[96] H. Bethe, *Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie*, Annalen der Physik **397** (1930) 325–400.

[97] ATLAS Collaboration, *The configuration system of the ATLAS Trigger*, IEEE Trans.Nucl.Sci. **55** (2008) 392–398.

[98] ATLAS Collaboration, *ATLAS Data Acquisition and High Level Trigger system*, JINST **11** (2016) P06008.

[99] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77** (2017) 317.

[100] ATLAS Collaboration, *Data Quality Monitoring Framework for ATLAS Experiment: Performance Achieved with Colliding Beams at LHC*, ATL-DAQ-PROC-2011-007, 2011.
URL: `http://cds.cern.ch/record/1322433`

[101] ATLAS Collaboration, *Data Quality Monitoring Framework for the ATLAS Experiment at the LHC*, ATL-DAQ-CONF-2008-006, 2007.
URL: `http://cds.cern.ch/record/1061900`

[102] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81–86.

[103] ATLAS Collaboration, *Persistent Data Layout and Infrastructure for Efficient Selective Retrieval of Event Data in ATLAS*, arXiv:1109.3119, 2011.
URL: `https://cds.cern.ch/record/1382676`

[104] ATLAS Collaboration, *Implementation of the ATLAS Run 2 event data model*, J. Phys. Conf. Ser. **664** (2015) 072045.

[105] J. Seng, *Fat Unicorn Character Image*, Toonapaloza.net (2009).

[106] ATLAS Collaboration, *ATLAS computing: Technical design report*, ATLAS-TDR-17, 2005.
URL: `https://cds.cern.ch/record/837738`

[107] ATLAS Collaboration, *Expected performance of the ATLAS experiment: detector, trigger and physics*, CERN-OPEN-2008-020, 2009.
URL: `https://cds.cern.ch/record/1125884`

[108] ATLAS Collaboration, *Calorimeter clustering algorithms: Description and performance*, ATL-LARG-PUB-2008-002, 2008.
URL: `https://cds.cern.ch/record/1099735`

[109] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. **C77** (2017) 490.

[110] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. **C76** (2016) 292.

[111] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024, 2016.
URL: `http://cds.cern.ch/record/2157687`

[112] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2016-015, 2016.
URL: `https://cds.cern.ch/record/2203514`

[113] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **0804** (2008) 063.

[114] M. Cacciari and G. P. Salam, *Dispelling the N**3 myth for the* k_t *jet-finder*, Phys.Lett. **B641** (2006) 57–61.

[115] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. **D96** (2017) 072002.

[116] ATLAS Collaboration, *A new tagger for the charge identification of b-jets*, ATL-PHYS-PUB-2015-040, 2015.
URL: `https://cds.cern.ch/record/2048132`

[117] M. L. Perl et al., *Evidence for Anomalous Lepton Production in* e^+e^- *Yusufu and Reichert, Joseph and Hartmann, Nikolai Marcel", Annihilation*, Phys. Rev. Lett. **35** (1975) 1489–1492.

[118] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014.
URL: `http://cds.cern.ch/record/1700870`

[119] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, ATLAS-CONF-2014-032, 2014.
URL: `http://cds.cern.ch/record/1706245`

[120] ATLAS Collaboration, *Electron trigger performance in 2015 ATLAS data*, ATL-COM-DAQ-2015-102, 2015.
URL: `https://cds.cern.ch/record/2034867`

[121] ATLAS Collaboration, *ATLAS Tau Trigger in Run 2*, ATLAS-CONF-2017-061, 2017.
URL: `http://cds.cern.ch/record/2274201`

[122] J. H. Friedman, *Stochastic gradient boosting*, Comput. Stat. Data Anal. **38** (2002) 367–378.

[123] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-027, 2015.
URL: `https://cds.cern.ch/record/2037904`

[124] A. Elagin et al., *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. **A654** (2011) 481–489.

[125] ATLAS Collaboration, *Measurement of the* $H \to \tau^+\tau^-$ *cross-section in* 13 *TeV Collisions with the ATLAS Detector*, ATL-COM-PHYS-2017-446, 2017.
URL: `https://cds.cern.ch/record/2261605`

[126] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. **C76** (2016) 581.

[127] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083, 2013.
URL: `http://cds.cern.ch/record/1570994.`

[128] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 580.

[129] F. Meloni, *Primary vertex reconstruction with the ATLAS detector*, JINST **11** (2016) C12060.

[130] ATLAS Collaboration, *Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of sqrts = 7 TeV*, ATLAS-CONF-2010-038, 2010.
URL: `https://cds.cern.ch/record/1277678`

[131] D. Zanzi, *Search for the Standard Model Higgs Boson in Hadronic $\tau^+\tau^-$ Decays with the ATLAS Detector*, CERN-THESIS-2014-085, Munich, 2014.

[132] ATLAS Collaboration, *ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823.

[133] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[134] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285.

[135] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012.
URL: `https://cds.cern.ch/record/1474107`

[136] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016) 023001.

[137] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145.

[138] N. Davidson, T. Przedzinski and Z. Was, *PHOTOS Interface in C++: Technical and Physics Documentation*, CERN-PH-TH-2010-26, 2010.
URL: `https://cds.cern.ch/record/1304806`

[139] K. Hamilton et al., *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222.

[140] R. V. Harlander, *Virtual corrections to $gg \to H$ to two loops in the heavy top limit*, Phys. Lett. **B492** (2000) 74–80.

[141] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040.

[142] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the Powheg method*, JHEP **11** (2007) 070.

[143] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the Powheg BOX*, JHEP **06** (2010) 043.

[144] E. Bagnaschi et al., *Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM*, JHEP **02** (2012) 088.

[145] K. Hamilton, P. Nason and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05** (2015) 140.

[146] T. Sjöstrand, S. Mrenna and P. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867.

[147] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079.

[148] R. Ball et al., *Parton distributions for the LHC Run II*, JHEP **1504** (2014) 040.

[149] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014.
URL: `https://cds.cern.ch/record/1966419`

[150] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039.

[151] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601.

[152] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038.

[153] S. Höche et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027.

[154] R. Gavin et al., *FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order*, Comput. Phys. Commun. **182** (2011) 2388–2403.

[155] C. Anastasiou et al., *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, Phys. Rev. **D69** (2004) 094008.

[156] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024.

[157] S. Alioli, S.-O. Moch and P. Uwer, *Hadronic top-quark pair-production with one jet and parton showering*, JHEP **01** (2012) 137.

[158] S. Alioli et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111.

[159] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. **C71** (2011) 1547.

[160] P. Artoisenet et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015.

[161] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026.

[162] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018.

[163] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, CHEP-2003-MOLT007, 2003.

[164] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-015, 2015.
URL: `https://cds.cern.ch/record/2037613`

[165] ATLAS Collaborations, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, CERN-EP-2016-117, 2016.
URL: `https://cds.cern.ch/record/2208146`

[166] L. Teodorescu, *Artificial neural networks in high-energy physics*, CERN-08-002, 2008.

[167] B. P. Roe et al., *Boosted decision trees, an alternative to artificial neural networks*, Nucl. Instrum. Meth. **A543** (2005) 577–584.

[168] A. Hoecker et al., *TMVA: Toolkit for Multivariate Data Analysis*, CERN-OPEN-2007-007, 2007.
URL: `https://cds.cern.ch/record/1019880`

[169] F. J. Massey, *The Kolmogorov-Smirnov test for goodness of fit*, J. Amer. Stat. Assoc. **46** (1951) 68–78.

[170] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554.

[171] L. Breiman et al., *Classification and regression trees*, Chapman and Hall/CRC, 1984.

[172] W. A. Rolke, A. M. Lopez and J. Conrad, *Limits and confidence intervals in the presence of nuisance parameters*, Nucl. Instrum. Meth. **A551** (2005) 493–503.

[173] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, ATL-PHYS-PUB-2011-11, 2011.
URL: `https://cds.cern.ch/record/1379837`

[174] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016, 2012.

[175] L. Moneta et al., *The RooStats Project*, arXiv:1009.1003, 2010.
URL: `https://cds.cern.ch/record/1289965`

[176] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput. Phys. Commun. **10** (1975) 343–367.

[177] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Annals Math. Statist. **9** (1938) 60–62.

[178] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Trans. Amer. Math. Soc. **54** (1943) 426–482.

Appendix

APPENDIX A

Event Modelling

This chapter summarises additional details and studies concerning the event modelling. Further information are provided in Chapter (7).

List of Simulated Samples

This section lists simulated samples used in the $\tau_{\text{had}}\tau_{\text{had}}$ analysis, including the corresponding cross sections and filter efficiencies. Further details can be found in [125].

DSID	Dataset Name	XSec [pb]	K-Factor	Eff. [%]
345123	PowhegPy8EG_NNLOPS_nnlo_30_ggH125_tautauh30h20	3.0469	1	18.89
345076	PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_tautauh30h20	0.23721	1	19.829
345211	PowhegPy8EG_NNPDF30_AZNLO_WmH125J_Winc_MINLO_tautau	0.033417	1	100
345212	PowhegPy8EG_NNPDF30_AZNLO_WpH125J_Winc_MINLO_tautau	0.052685	1	100
345217	PowhegPy8EG_NNPDF30_AZNLO_ZH125J_Zinc_MINLO_tautau	0.055438	1	100
343365	aMcAtNloPythia8EvtGen_A14_NNPDF23_NNPDF30ME_ttH125_dilep	0.05343	1	100

Table A.1: List of simulated signal samples and corresponding parameters used in the analysis. The Higgs boson mass is set to $m_H = 125\,\text{GeV}$.

DSID	Dataset Name	XSec [pb]	K-Factor	Eff. [%]
344443	Sherpa_NNPDF30NNLO_Ztautau2jets_Min_N_TChannel_CSSKIN	0.63487	1	100
364210	Sherpa_221_NN30NNLO_Ztt_Mll10_40_MAXHTPTV0_70_BVeto	2417.9	0.9751	96.527
364211	Sherpa_221_NN30NNLO_Ztt_Mll10_40_MAXHTPTV0_70_BFilter	2414.2	0.9751	3.3594
364212	Sherpa_221_NN30NNLO_Ztt_Mll10_40_MAXHTPTV70_280_BVeto	50.37	0.9751	89.015
364213	Sherpa_221_NN30NNLO_Ztt_Mll10_40_MAXHTPTV70_280_BFilter	50.44	0.9751	10.985
364214	Sherpa_221_NN30NNLO_Ztt_Mll10_40_MAXHTPTV280_E_CMS_BVeto	3.2834	0.9751	85.081
364215	Sherpa_221_NN30NNLO_Ztt_Mll10_40_MAXHTPTV280_E_CMS_BFilter	3.2788	0.9751	14.326
364128	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CVetoBVeto	1981.6	0.9751	82.142
364129	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CFilterBVeto	1978.8	0.9751	11.314
364130	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_BFilter	1981.8	0.9751	6.4453
364131	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_CVetoBVeto	110.37	0.9751	68.883
364132	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_CFilterBVeto	110.51	0.9751	18.29
364133	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_BFilter	110.87	0.9751	11.089
364134	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CVetoBVeto	40.781	0.9751	60.821
364135	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CFilterBVeto	40.74	0.9751	22.897
364136	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_BFilter	40.761	0.9751	13.442
364137	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_CVetoBVeto	8.5502	0.9751	56.036
364138	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_CFilterBVeto	8.6707	0.9751	26.245
364139	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV280_500_BFilter	8.6804	0.9751	17.313
364140	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV500_1000	1.8096	0.9751	100
364141	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV1000_E_CMS	0.14834	0.9751	100
344775	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_h30h20	831.33	0.9751	14.707
344779	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV70_140_h30h20	46.352	0.9751	20.358
344782	Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_h30h20	17.064	0.9751	26.126

Table A.2: List of simulated background samples and corresponding parameters describing the $Z/\gamma^{*}\rightarrow\tau\tau$ background component.

DSID	Dataset Name	XSec [pb]	K-Factor	Eff. [%]
364184	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_CVetoBVeto	19152	0.9702	82.495
364185	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_CFilterBVeto	19153	0.9702	12.934
364186	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV0_70_BFilter	19163	0.9702	4.4594
364187	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_CVetoBVeto	947.65	0.9702	67.382
364188	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_CFilterBVeto	946.73	0.9702	22.222
364189	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV70_140_BFilter	943.3	0.9702	10.396
364190	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV140_280_CVetoBVeto	339.36	0.9702	59.622
364191	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV140_280_CFilterBVeto	339.63	0.9702	29.025
364192	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV140_280_BFilter	339.55	0.9702	11.229
364193	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_CVetoBVeto	72.048	0.9702	54.569
364194	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_CFilterBVeto	71.976	0.9702	31.648
364195	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV280_500_BFilter	72.026	0.9702	13.426
364196	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV500_1000	15.046	0.9702	100
364197	Sherpa_221_NNPDF30NNLO_Wtaunu_MAXHTPTV1000_E_CMS	1.2339	0.9702	100

Table A.3: List of simulated background samples and corresponding parameters describing the $W + \text{jets}$ background component.

DSID	Dataset Name	XSec [pb]	K-Factor	Eff. [%]
410000	PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_nonallhad	696.11	1.1949	54.3
410007	PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_allhad	695.99	1.1951	45.62
410009	PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_dil	696.12	1.1949	10.53
410011	PowhegPythiaEvtGen_P2012_singletop_tchan_lept_top	43.739	1.0094	100
410012	PowhegPythiaEvtGen_P2012_singletop_tchan_lept_antitop	25.778	1.0193	100
410013	PowhegPythiaEvtGen_P2012_Wt_inclusive_top	34.009	1.054	100
410014	PowhegPythiaEvtGen_P2012_Wt_inclusive_antitop	33.989	1.054	100
410025	PowhegPythiaEvtGen_P2012_SingleTopSchan_noAllHad_top	2.0517	1.0046	100
410026	PowhegPythiaEvtGen_P2012_SingleTopSchan_noAllHad_antitop	1.2615	1.0215	100

Table A.4: List of simulated background samples describing contributions from top processes.

DSID	Dataset Name	XSec [pb]	K-Factor	Eff. [%]
363355	Sherpa_221_NNPDF30NNLO_ZqqZvv	15.564	1	27.976
363356	Sherpa_221_NNPDF30NNLO_ZqqZll	15.563	1	13.961
363357	Sherpa_221_NNPDF30NNLO_WqqZvv	6.7973	1	100
363358	Sherpa_221_NNPDF30NNLO_WqqZll	3.437	1	100
363359	Sherpa_221_NNPDF30NNLO_WpqqWmlv	24.717	1	100
363360	Sherpa_221_NNPDF30NNLO_WplvWmqq	24.734	1	100
363489	Sherpa_221_NNPDF30NNLO_WlvZqq	11.413	1	100
363490	Sherpa_221_NNPDF30NNLO_llll	1.2557	1	100
363491	Sherpa_221_NNPDF30NNLO_lllv	4.5877	1	100
363492	Sherpa_221_NNPDF30NNLO_llvv	12.465	1	100
363493	Sherpa_221_NNPDF30NNLO_lvvv	3.2274	1	100
363494	Sherpa_221_NNPDF30NNLO_vvvv	0.60154	1	100

Table A.5: List of simulated background samples describing the di-boson background component.

Normalisation Studies

To understand the origins of the differences in the Z/γ^* normalisation factors observed between the analysis channels, several studies were performed. The studies described in the following were done in the CBA Preselection with a tighter requirement of $\Delta R_{\tau\tau} < 2.4$. This requirement does not affect the observed normalisation difference.

- Since the tau trigger group observed issues in the modelling of the trigger efficiency dependent on the leading jet p_T, the effect of a mismodelling in the trigger jet selection was studied by removing the p_T cut of the leading jet completely. Consecutively, it was raised in several steps below and above the current cut value of 70 GeV. As can be seen in Tab. A.6, no noticeable effect on r_Z was observed.

Leading Jet p_T	L1_J25 matching	r_{QCD}	r_Z
-	no	0.7873 ± 0.0278	0.8378 ± 0.0228
	yes	0.8175 ± 0.0334	0.8188 ± 0.0228
50 GeV	no	0.7926 ± 0.0313	0.8474 ± 0.0234
	yes	0.8145 ± 0.0337	0.8302 ± 0.0231
60 GeV	no	0.7952 ± 0.0330	0.8638 ± 0.0231
	yes	0.8010 ± 0.0340	0.8555 ± 0.0228
70 GeV	no	0.7895 ± 0.0359	0.8710 ± 0.0225
	yes	0.7903 ± 0.0363	0.8695 ± 0.0225
80 GeV	no	0.8078 ± 0.0414	0.8707 ± 0.0228
	yes	0.8105 ± 0.0417	0.8710 ± 0.0228
90 GeV	no	0.8676 ± 0.0501	0.8520 ± 0.0239
	yes	0.8739 ± 0.0505	0.8504 ± 0.0240
100 GeV	no	0.8149 ± 0.0561	0.8764 ± 0.0233
	yes	0.8207 ± 0.0565	0.8752 ± 0.0234

Table A.6: Normalisation factors for different offline jet p_T thresholds and jet to trigger-jet matching in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

- The p_T thresholds for the τ_{had} candidates were raised to eliminate effects from underperforming ditau trigger components. No noticeable change was observed as can be seen from Tab. A.7.
- The fit of r_Z was repeated separately for different numbers of reconstructed tracks associated with the τ_{had} candidates. The different normalisation factors for all permutations for leading and subleading τ_{had} can be found in Tab. A.8. While none of the resulting r_Z factors are compatible with 1, the 1-prong/1-prong category differs significantly from the other combinations.
- To exclude effects from the simulation in the specific phase space, a comparison between MADGRAPH and the default SHERPA generated samples was performed. When moving to Madgraph samples r_Z changes by about 10%.

$p_T(\tau_0)$ [GeV]	$p_T(\tau_1)$ [GeV]	r_{QCD}	r_Z
40	30	0.790 ± 0.036	0.870 ± 0.023
	35	0.761 ± 0.044	0.879 ± 0.024
	40	0.637 ± 0.055	0.909 ± 0.026
45	30	0.760 ± 0.041	0.889 ± 0.023
	35	0.730 ± 0.050	0.896 ± 0.024
	40	0.648 ± 0.060	0.906 ± 0.027
	45	0.650 ± 0.080	0.883 ± 0.036
50	30	0.746 ± 0.048	0.894 ± 0.023
	35	0.720 ± 0.056	0.899 ± 0.025
	40	0.649 ± 0.069	0.901 ± 0.028
	45	0.671 ± 0.088	0.887 ± 0.038
	50	0.650 ± 0.124	0.898 ± 0.046
55	30	0.758 ± 0.056	0.889 ± 0.024
	35	0.785 ± 0.070	0.868 ± 0.028
	40	0.726 ± 0.086	0.863 ± 0.032
	45	0.737 ± 0.106	0.873 ± 0.040
	50	0.663 ± 0.131	0.907 ± 0.047
	55	0.818 ± 0.196	0.939 ± 0.058
60	30	0.724 ± 0.060	0.905 ± 0.025
	35	0.749 ± 0.076	0.900 ± 0.029
	40	0.731 ± 0.095	0.878 ± 0.034
	45	0.759 ± 0.121	0.884 ± 0.043
	50	0.663 ± 0.150	0.913 ± 0.050
	55	0.844 ± 0.210	0.930 ± 0.060
	60	0.835 ± 0.259	0.925 ± 0.069

Table A.7: Normalisation factors for different τ_{had} p_T thresholds.

Scenario	No τID-SF		Inclusive τID-SF	
$\tau_0\tau_1$	r_{QCD}	r_Z	r_{QCD}	r_Z
all Prong nOS	0.79 ± 0.04	1.02 ± 0.03	0.79 ± 0.04	0.87 ± 0.02
all Prong SS	1.66 ± 0.08	1.07 ± 0.02	1.66 ± 0.08	0.91 ± 0.02
1p1p	1.71 ± 0.11	1.11 ± 0.03	1.70 ± 0.11	0.96 ± 0.03
1p3p	1.44 ± 0.22	0.95 ± 0.07	1.45 ± 0.22	0.80 ± 0.06
3p1p	1.51 ± 0.21	1.02 ± 0.06	1.56 ± 0.22	0.84 ± 0.05
3p3p	1.71 ± 0.43	1.04 ± 0.14	1.88 ± 0.47	0.83 ± 0.11

Table A.8: Normalisation factors for different Fake background templates in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, separated by number of reconstructed tracks in the τ_{had} pair and with/without identification efficiency scale factors.

- The identification scale factors for τ_{had} were measured in an inclusive $\tau_{\text{lep}}\tau_{\text{had}}$ selection. It is possible, that the events use for the measurement are dominated by τ_{had} candidates with lower transverse momentum than in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. A comparison was made with scale factors binned in τ_{had} p_T. No noticeable effect on r_Z was observed.
- Since the τ_{had} identification scale factors are significantly different between medium and tight working points all permutations of leading and subleading τ_{had} identification thresholds were investigated. As can be seen from Tab. A.9, the small changes in r_Z are consistent with the size difference between the respective scale factors.

ID scenario τ_0/τ_1	Sherpa r_{QCD}	r_Z
tight/tight	0.7722 ± 0.0356	0.8812 ± 0.0220
tight/med	0.6427 ± 0.0205	0.9130 ± 0.0223
med/tight	0.6540 ± 0.0186	0.9204 ± 0.0211
med/med	0.5588 ± 0.0109	0.9458 ± 0.0215

Table A.9: Normalisation factors for different identification threshold combinations for the τ_{had} pair.

- The τ_{had} identification scale factors for the default selection are different from 1, but close to consistent with 1 within their uncertainties:
  - 1.074 ± 0.06 for τ_{had} with one associated track
  - 1.11 ± 0.07 for τ_{had} with three associated tracks

  To understand the effect of the application of these scale factors, a fit was performed without applying them. The resulting r_Z is compatible with 1 within statistical uncertainties.

CBA Preselection Fit AND	Sherpa r_{QCD}	r_Z
$m_{\text{MMC}} > 50$ GeV	0.7919 ± 0.0359	0.8626 ± 0.0216
$m_{\text{MMC}} > 55$ GeV	0.7912 ± 0.0355	0.8641 ± 0.0208
$m_{\text{MMC}} > 60$ GeV	0.7905 ± 0.0351	0.8625 ± 0.0199
$m_{\text{MMC}} > 65$ GeV	0.7895 ± 0.0349	0.8618 ± 0.0190
$m_{\text{MMC}} > 70$ GeV	0.7906 ± 0.0350	0.8605 ± 0.0185
$m_{\text{MMC}} > 75$ GeV	0.7890 ± 0.0352	0.8594 ± 0.0184
$m_{\text{MMC}} > 80$ GeV	0.7957 ± 0.0358	0.8565 ± 0.0190

Table A.10: Normalisation factors for different lower cuts on m_{MMC} in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

- Lower thresholds were set on the m_{MMC} to understand whether the mis-modelling in the low-mass region - partially resolved by the $\Delta\phi_{\tau\tau}$ reweighting - affects the normalisation. No noticeable effect on the normalisation factors are observed, as shown in Tab. A.10.

Detailed CBA Pre-Fit Event Yields

Cut	$Z\tau\tau$	EWK $Z\tau\tau$	$Wjets$	Top	Di-boson + $Z \to \ell\ell$
QCD Fit (DeltaR $<$ 3.0)	10639.7±92.3	75.9±2.5	302.1±48.3	585.9±16.4	196.4±5.5
QCD Fit	9159.0±87.5	56.1±2.1	184.4±26.4	345.3±12.7	148.4±4.6
Preselection	9082.2±87.1	55.5±2.1	163.7±25.3	321.4±12.2	145.7±4.5
CBA VBF	310.9±12.1	17.4±1.2	5.5±2.4	11.1±2.1	5.7±0.9
VBF Low DeltaR	148.8±7.0	9.9±0.9	0.4±0.3	2.2±0.9	2.8±0.5
VBF High DeltaR Tight	102.3±8.3	6.3±0.8	4.9±2.2	4.8±1.2	1.8±0.8
VBF High DeltaR Loose	59.8±5.2	1.2±0.3	0.2±0.7	4.1±1.4	1.0±0.2
CBA Boosted	6922.2±63.4	32.0±1.6	111.6±20.5	242.1±10.5	118.5±4.2
CBA Boosted Tight	3597.1±39.2	20.2±1.2	30.6±5.6	39.5±4.3	67.1±2.5
CBA Boosted Loose	3325.1±49.8	11.8±1.0	81.0±19.7	202.6±9.6	51.5±3.4
r_{QCD}	0.7398 ± 0.0325				
r_{Z}	0.8904 ± 0.0213				

Table A.11: CBA background yields for each considered process.

Cut	gg$H\tau\tau$	VBF$H\tau\tau$	$ttH\tau\tau$	$ZH\tau\tau$	$WH\tau\tau$	$H \to \tau\tau$
QCD Fit (DeltaR $<$ 3.0)	153.8±1.7	54.9±0.4	0.4±0.0	8.3±0.2	11.7±0.3	229.2±1.8
QCD Fit	141.1±1.6	49.6±0.4	0.3±0.0	7.3±0.2	10.4±0.3	208.7±1.7
Preselection	137.0±1.6	48.4±0.4	0.3±0.0	7.1±0.2	10.0±0.3	202.8±1.7
CBA VBF	8.2±0.4	21.5±0.3	0.0±0.0	0.1±0.0	0.1±0.0	30.0±0.5
VBF Low DeltaR	4.2±0.2	12.3±0.2	0.0±0.0	0.1±0.0	0.1±0.0	16.6±0.3
VBF High DeltaR Tight	2.8±0.3	7.5±0.2	0.0±0.0	0.0±0.0	0.0±0.0	10.4±0.3
VBF High DeltaR Loose	1.3±0.1	1.7±0.1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.2
CBA Boosted	105.4±1.4	22.0±0.3	0.2±0.0	6.0±0.2	8.6±0.3	142.3±1.5
CBA Boosted Tight	55.9±1.0	11.8±0.2	0.1±0.0	3.4±0.2	4.9±0.2	76.0±1.1
CBA Boosted Loose	49.5±1.0	10.3±0.2	0.2±0.0	2.6±0.1	3.7±0.2	66.2±1.0
r_{QCD}	0.7398 ± 0.0325					
r_{Z}	0.8904 ± 0.0213					

Table A.12: CBA signal yields for each considered process.

Detailed MVA Pre-Fit Event Yields

Cut	$Z\tau\tau$	EWK $Z\tau\tau$	$Wjets$	Top	Di-boson + $Z \rightarrow \ell\ell$
QCD Fit	12446.4±101.4	76.6±2.5	288.1±58.9	640.3±17.0	207.6±10.5
Preselection	12246.2±100.7	75.7±2.5	282.3±33.4	575.9±16.0	199.9±10.4
MVA Rest	2627.6±67.1	9.3±0.8	84.2±22.4	163.8±8.6	28.3±8.9
MVA VBF	954.5±22.9	27.7±1.6	19.0±4.0	43.3±4.3	14.6±1.4
MVA Boosted	8664.1±71.6	38.6±1.7	179.1±24.4	368.7±12.8	156.9±5.1
r_{QCD}	0.6695 ± 0.0211				
r_Z	0.9359 ± 0.0200				

Table A.13: MVA background yields for each considered process.

Cut	gg$H\tau\tau$	VBF$H\tau\tau$	$ttH\tau\tau$	$ZH\tau\tau$	$WH\tau\tau$	$H \rightarrow \tau\tau$
QCD Fit	169.2±1.8	58.6±0.4	0.5±0.0	8.9±0.2	12.3±0.4	249.5±1.9
Preselection	161.5±1.7	56.3±0.4	0.5±0.0	8.5±0.2	11.7±0.3	238.5±1.8
MVA Rest	27.8±0.8	5.6±0.1	0.1±0.0	1.3±0.1	1.7±0.1	36.4±0.8
MVA VBF	15.9±0.5	28.0±0.3	0.0±0.0	0.3±0.0	0.4±0.1	44.7±0.6
MVA Boosted	117.8±1.5	22.7±0.3	0.4±0.0	6.9±0.2	9.7±0.3	157.4±1.5
r_{QCD}	0.6695 ± 0.0211					
r_Z	0.9359 ± 0.0200					

Table A.14: MVA signal yields for each considered process.

APPENDIX B

BDT Training and Optimisation

Modelling of Input Variables

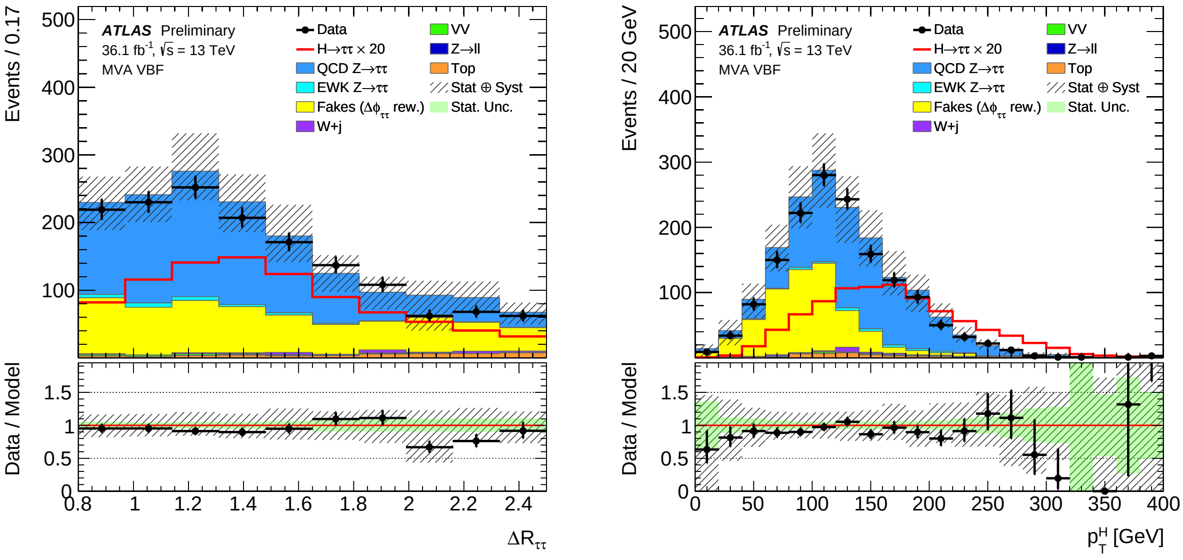


Figure B.1: Modelling of the input variables used in the $\tau_{\rm had}\tau_{\rm had}$ VBF BDT. The uncertainty band includes statistical and systematic uncertainties.

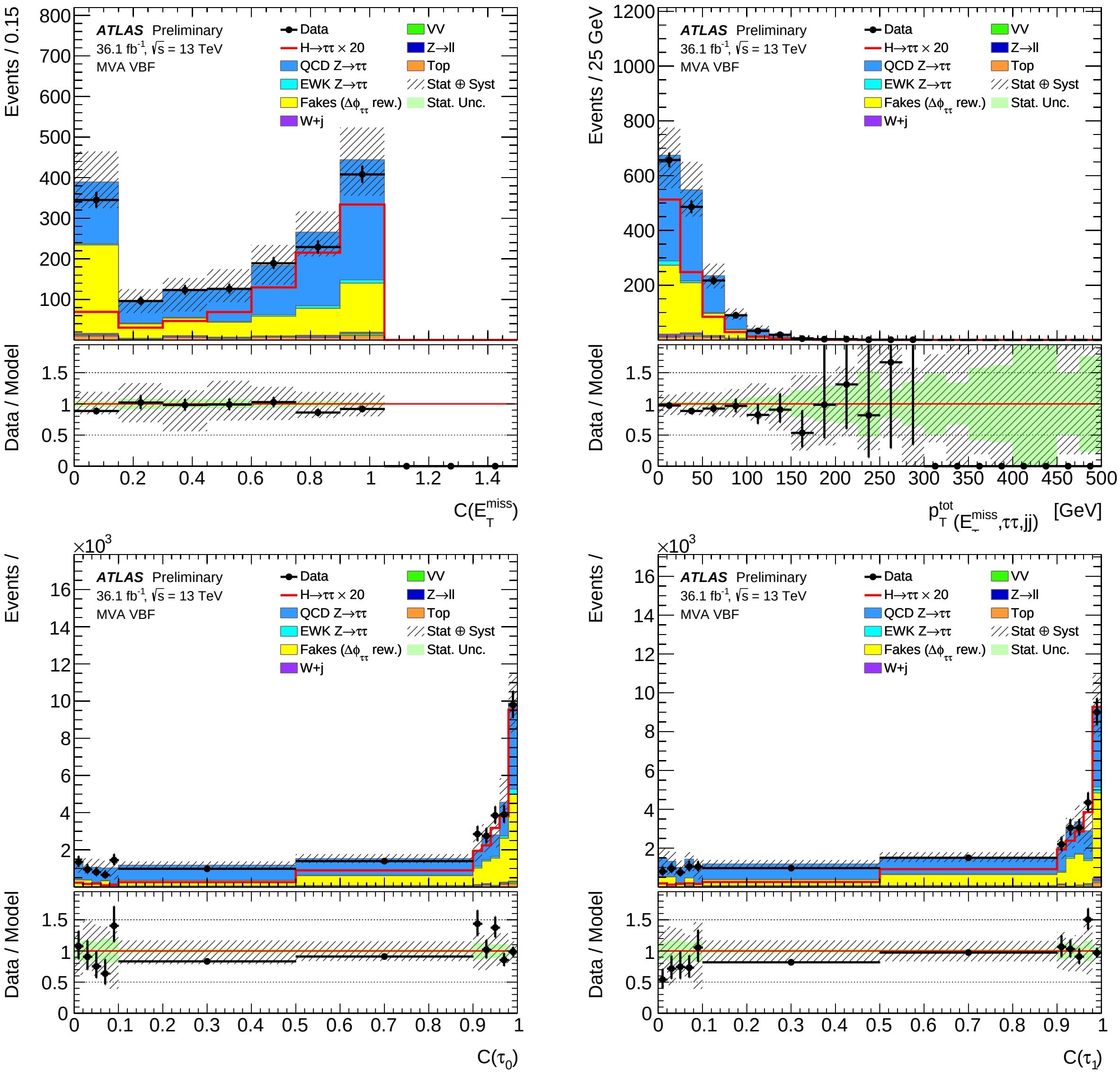


Figure B.2: Modelling of the input variables used in the $\tau_{\rm had}\tau_{\rm had}$ VBF BDT. The uncertainty band includes statistical and systematic uncertainties.

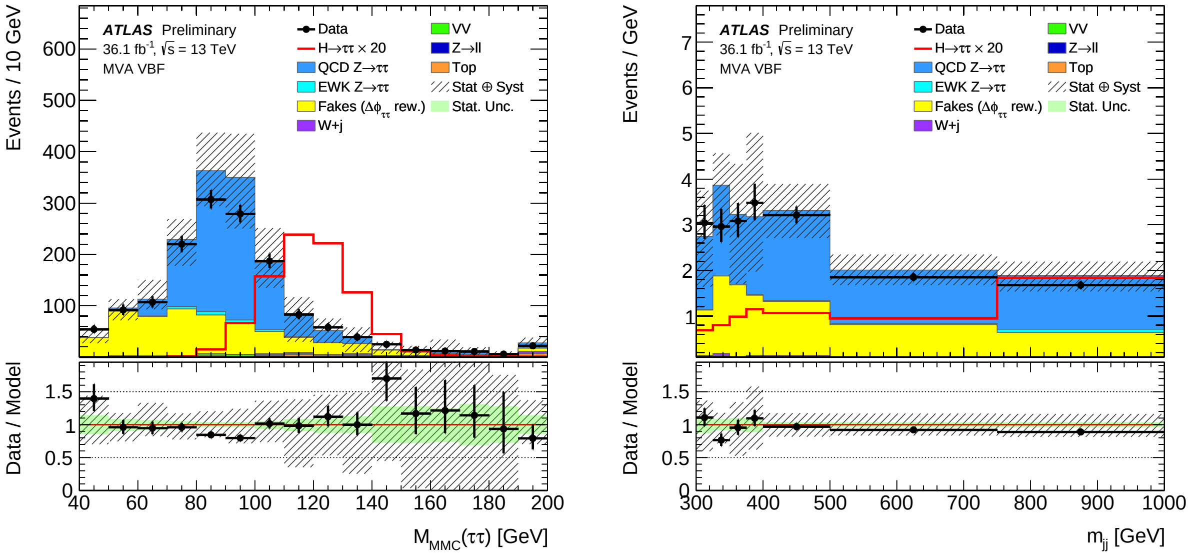


Figure B.3: Modelling of the input variables used in the $\tau_{\text{had}}\tau_{\text{had}}$ VBF BDT. The uncertainty band includes statistical and systematic uncertainties.

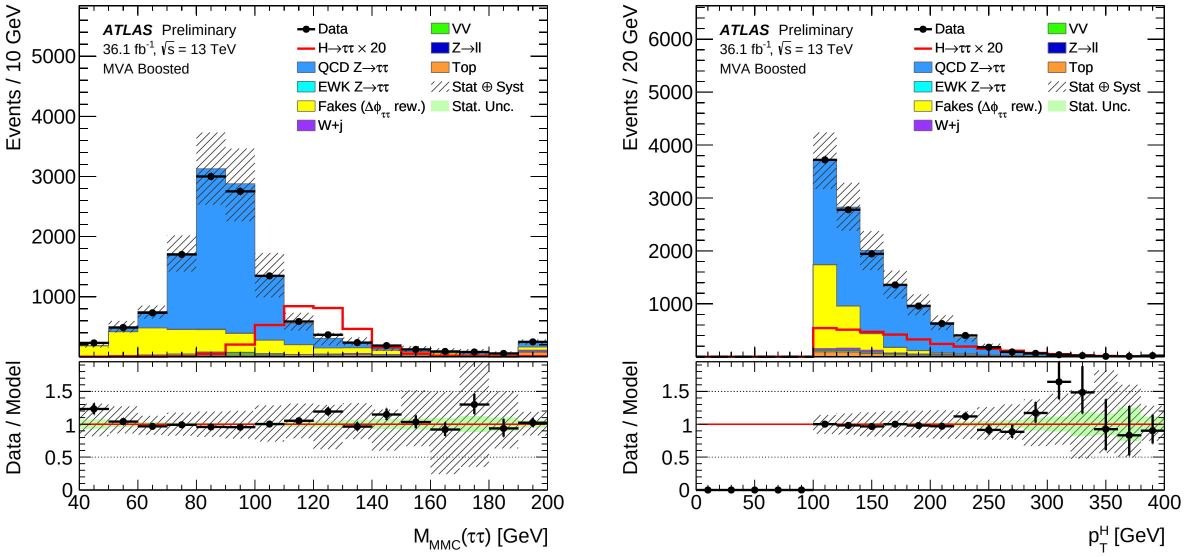


Figure B.4: Modelling of the input variables used in the $\tau_{\text{had}}\tau_{\text{had}}$ Boosted BDT. The uncertainty band includes statistical and systematic uncertainties.

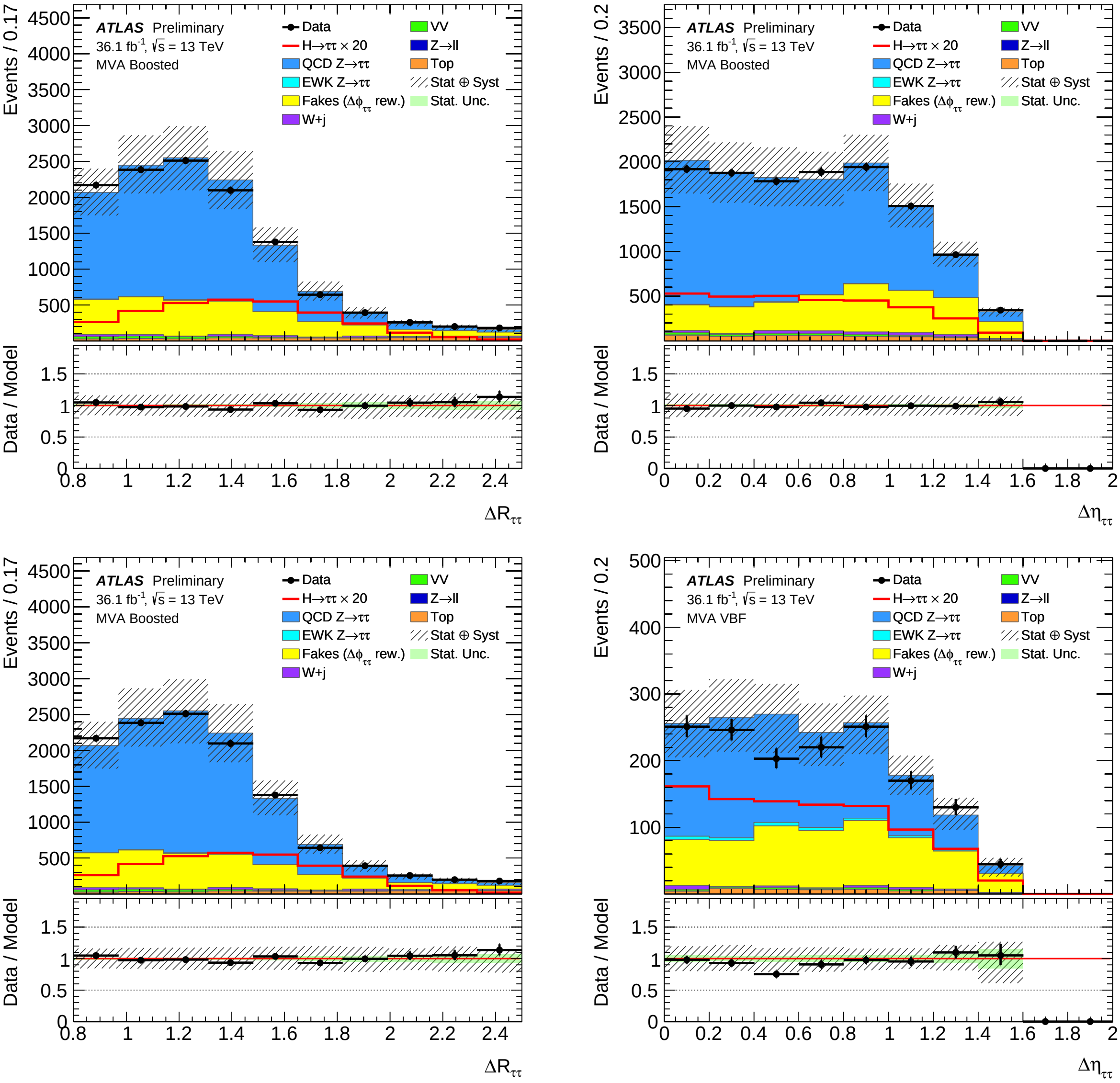


Figure B.5: Modelling of the input variables used in the $\tau_{\rm had}\tau_{\rm had}$ Boosted BDT. The uncertainty band includes statistical and systematic uncertainties.

Full Input Variable Correlation Matrices

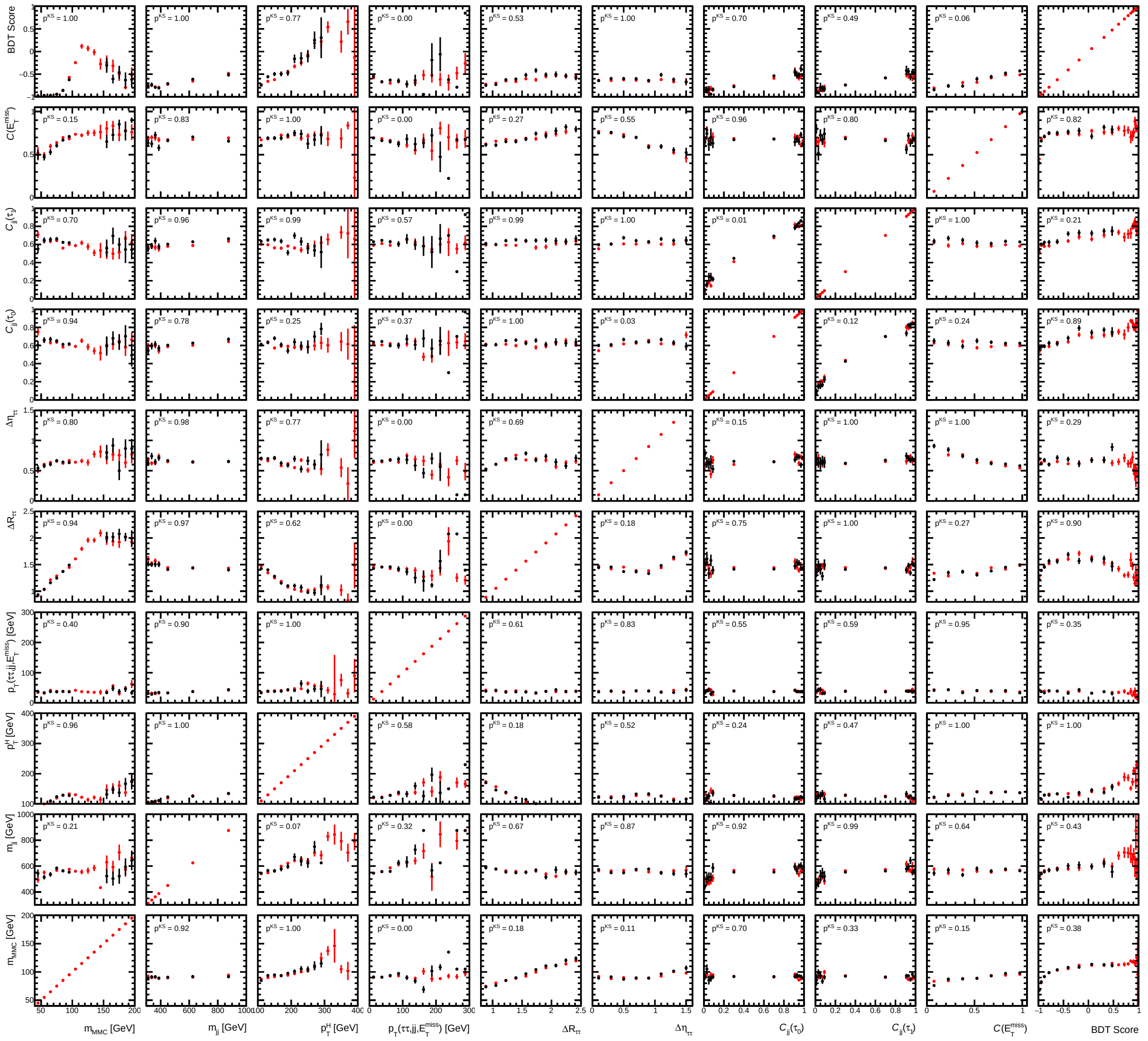


Figure B.6: Correlations between input variables used in the BDT in the MVA VBF region.

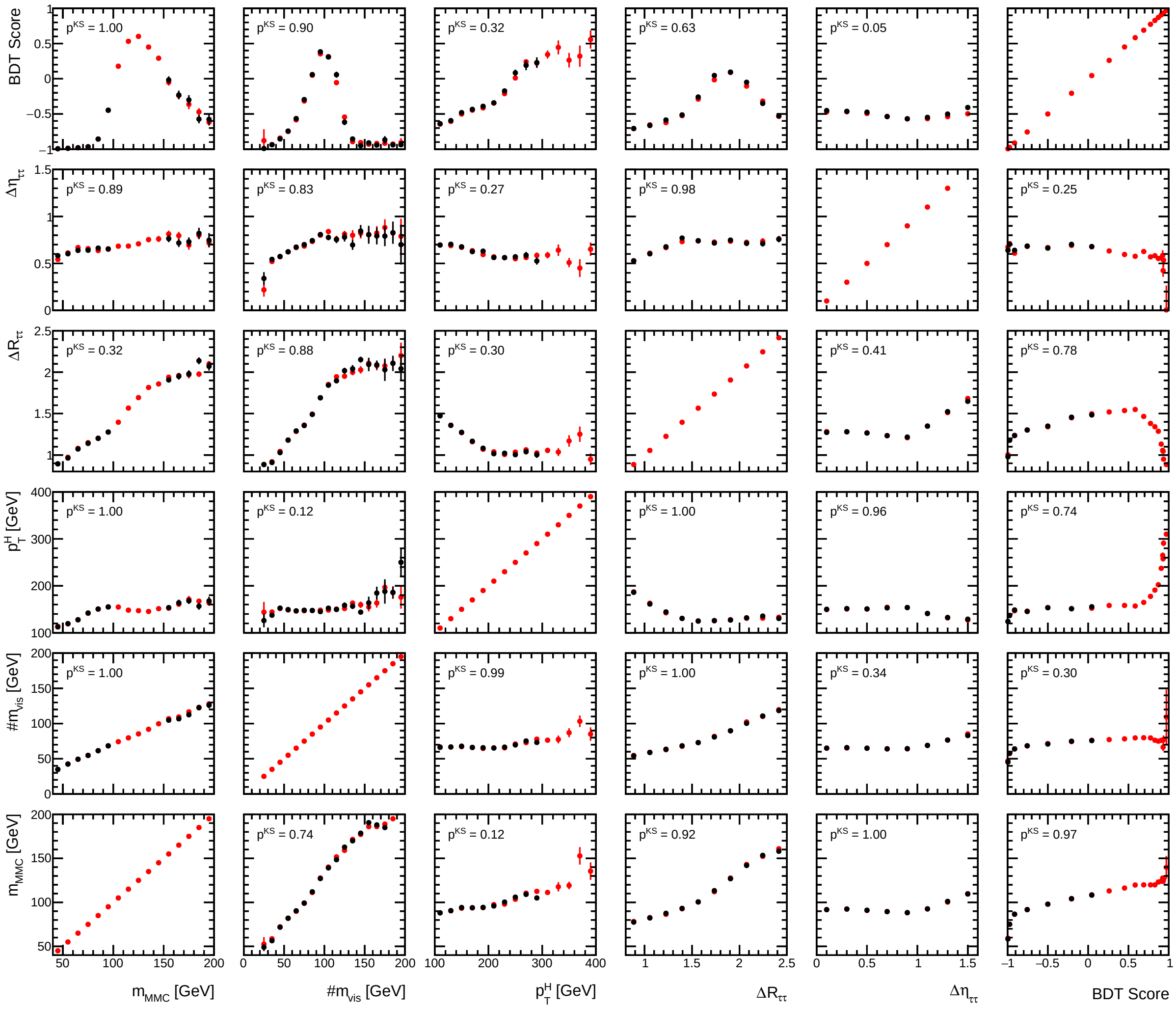


Figure B.7: Correlations between all input variables used in the BDT in the MVA Boosted region.

Training Statistics Improvement

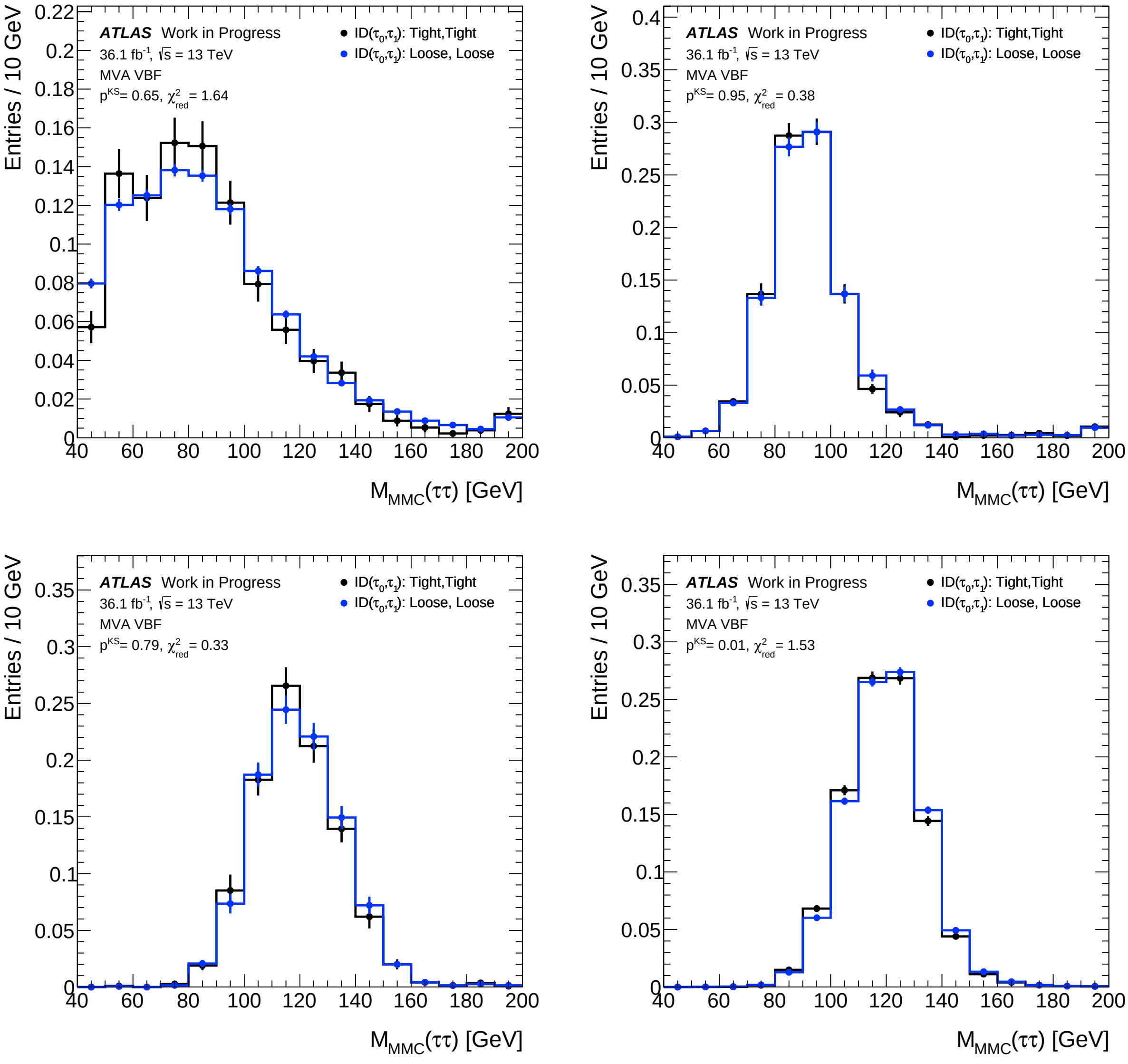


Figure B.8: Comparison of the m_{MMC} distribution between default and loosened BDT training cuts, for Fakes (topleft), $Z/\gamma^* \rightarrow \tau\tau$ (top right), ggH (bottom left) and VBFH (bottom right).

BDT Input Variable List Trimming for Boosted

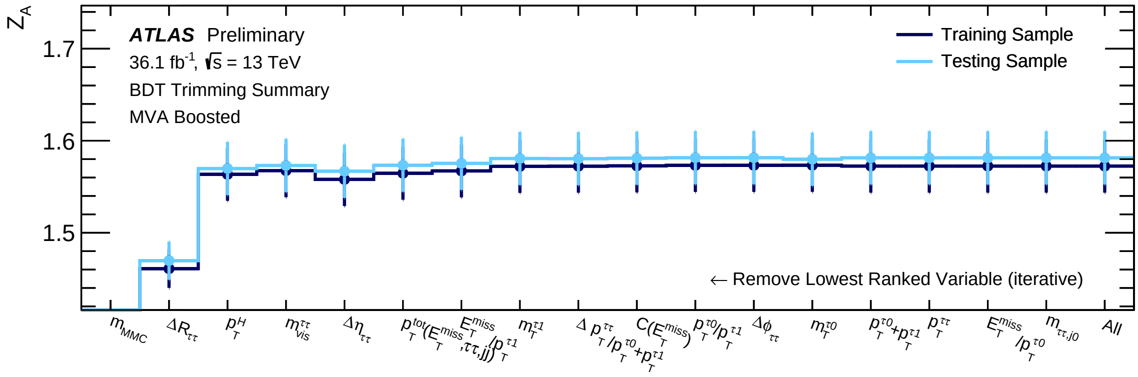


Figure B.9: Trimming scan for BDT in the MVA Boosted region.

BDT Parameter Scans

In order to find non-overtrained, well-performing BDT configurations a scan of the BDT parameter range was performed.
Firstly, a scan over a wide range of parameters is performed for the MVA VBF and Boosted categories separately in order to find a suitable BDT to perform the variable list trimming. With the trimmed variable list at hand another scan was performed, selecting a parameter range which proved to yield well-performing BDTs in the first iteration. The figure-of-merits to judge the BDT performance are:

- **Performance**, that is separation of signal and background:
  - ROC integral
  - $\mathcal{Z}_{\mathrm{A}}$
- **Overtraining**, the agreement between BDT output on training and testing set
  - p^{KS}: Kolmogorov-Smirnov test probability
  - χ^2_{red}: χ^2 over degrees of freedom (reduced χ^2)
- **Shape**, manual tests on shape of BDT distributions
  - $\mathcal{S}_{\prime}$ bin-by-bin difference: Each bin in signal (bkgd) is higher (lower) than the previous one (left to right). If not, add bin difference to counter. Then check for scenario with lowest diff.
  - $\mathcal{S}_{\infty}$ highest bin: Highest bin in at highest (lowest) BDT score for signal (bkgd).
  - $\mathcal{S}_{\in}$ combines $\mathcal{S}_{\prime}, \mathcal{S}_{\infty}$

All figures of merits were evaluated on training and testing set where applicable. Furthermore, all BDTs were trained on the even and odd set.

MVA VBF Category

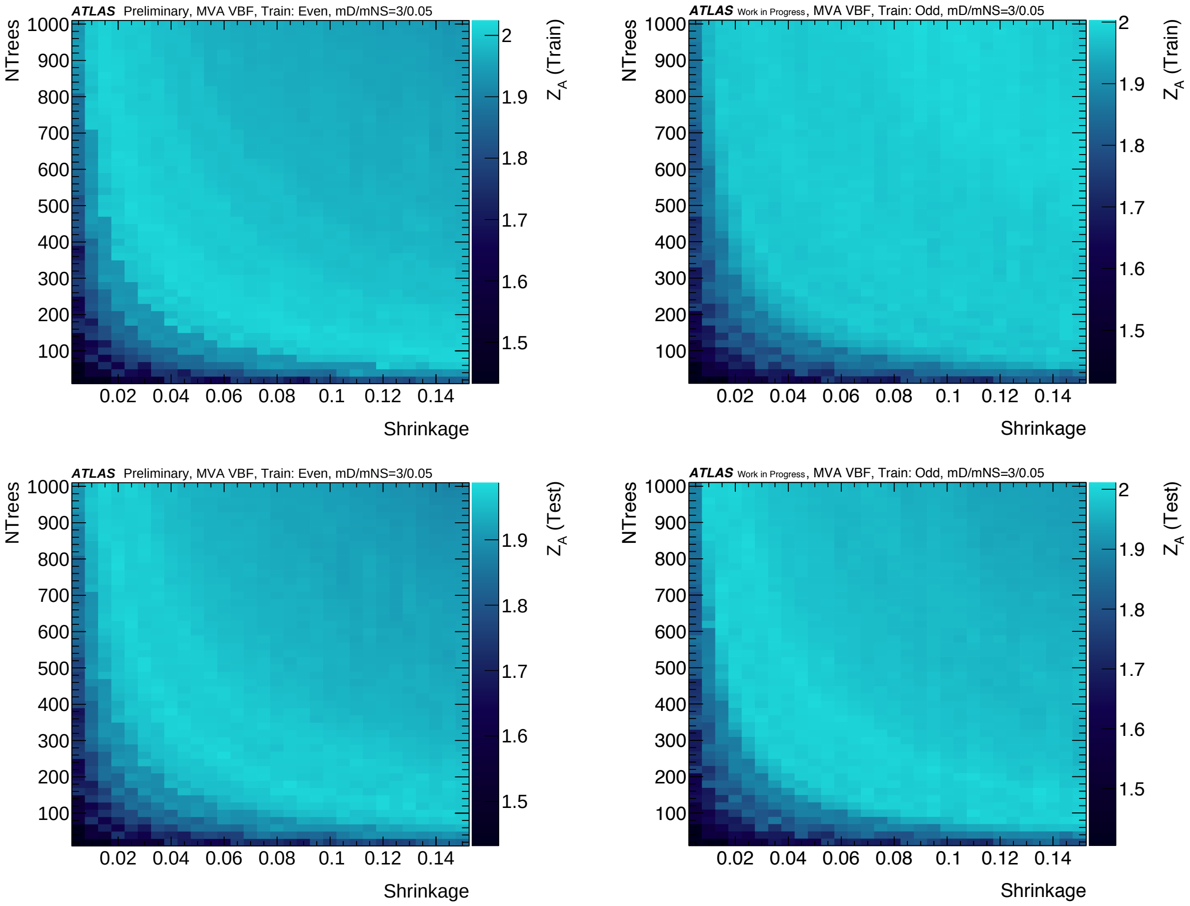


Figure B.10: Detailed parameter scan for optimal BDT Regularisation in MVA VBF. Shown is the Asimov significance $\mathcal{Z}_\mathrm{A}$ in dependence on number of trees `nTrees`. For the figures on the left (right), the BDTs were trained on the even (odd) set.

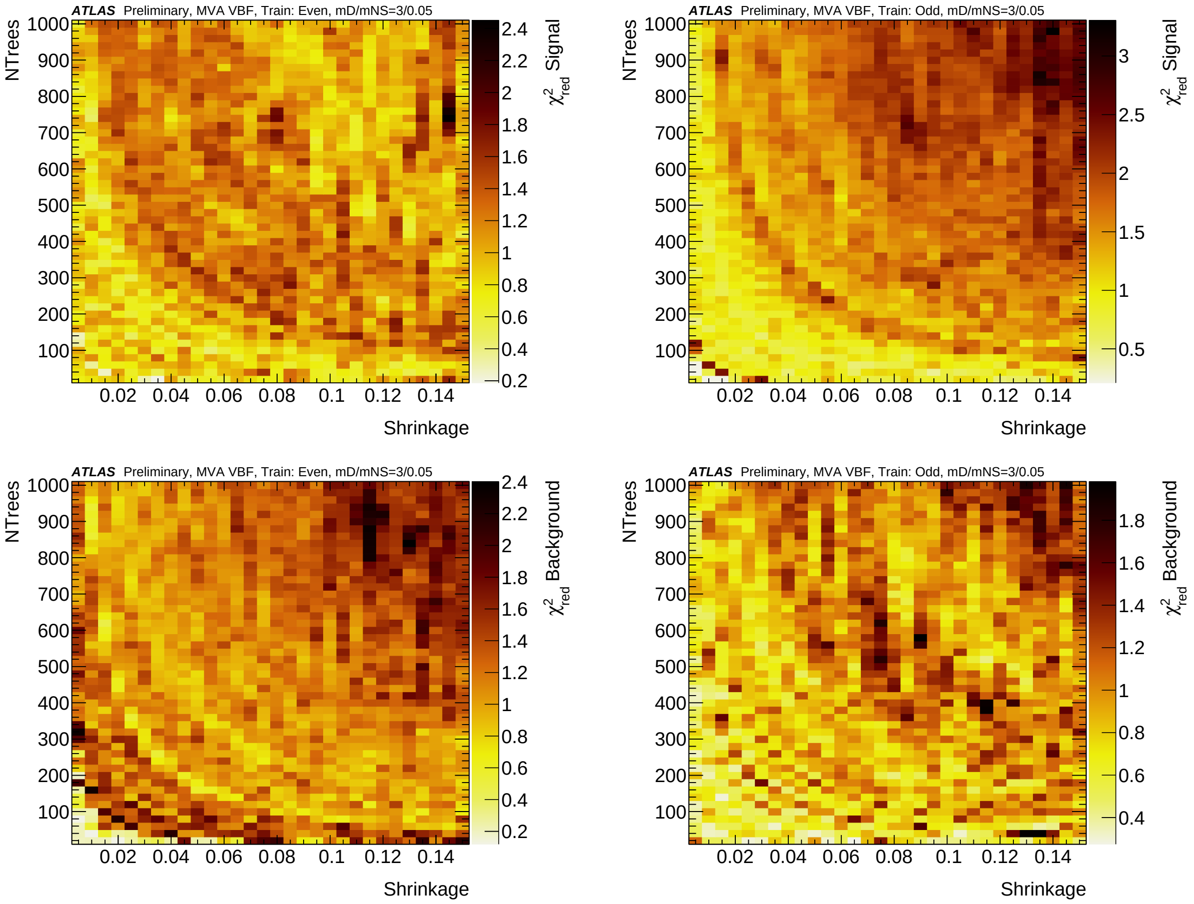


Figure B.11: Detailed parameter scan for optimal BDT Regularisation in MVA VBF. Shown is the χ^2_{red} between signal distribution of training and testing events in dependence on number of trees. For the figures on the left (right), the BDTs were trained on the even (odd) set.

MVA Boosted Category

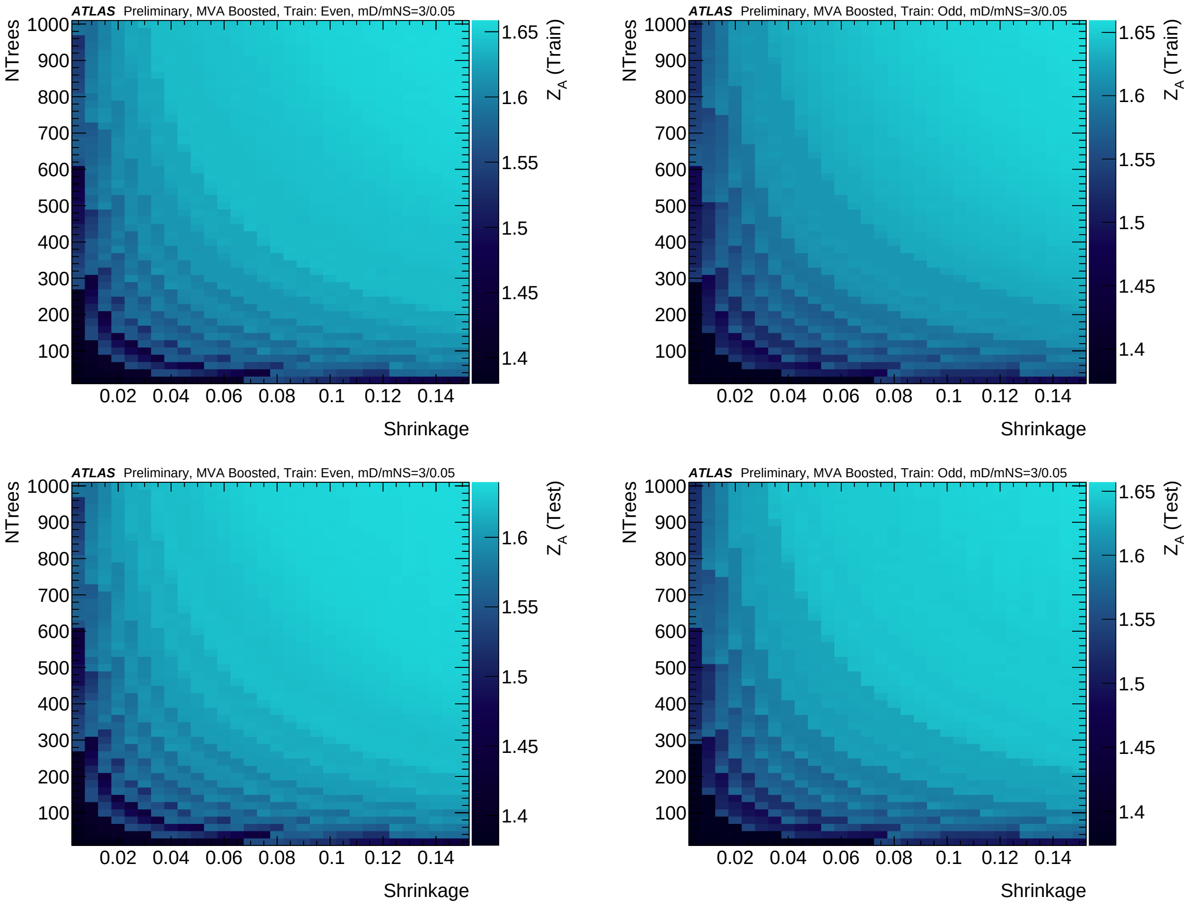


Figure B.12: Detailed parameter scan for optimal BDT Regularisation in MVA Boosted. Shown is the Asimov significance $\mathcal{Z}_{\mathrm{A}}$ in dependence on number of trees `nTrees`. For the figures on the left (right), the BDTs were trained on the even (odd) set.

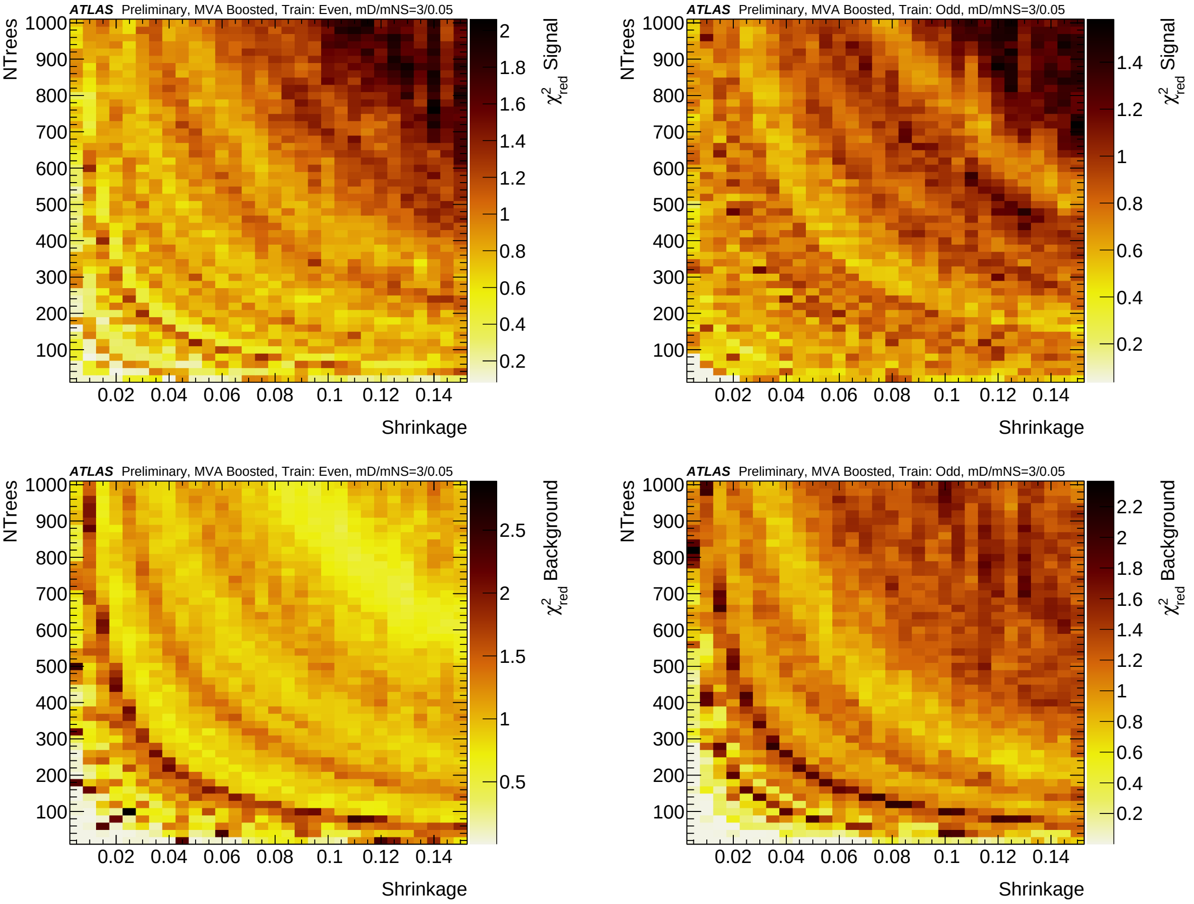


Figure B.13: Detailed parameter scan for optimal BDT Regularisation in MVA Boosted. Shown is the χ^2_{red} between signal distribution of training and testing events in dependence on number of trees. For the figures on the left (right), the BDTs were trained on the even (odd) set.

Scan Results for the BDT in MVA VBF

Criteria	Shrinkage	MaxDepth	MinNodeSize	NTrees	NCuts
p^{KS}	0.015	4.0	5.0	20.0	140.0
χ^2_{red}	0.01	4.0	5.0	80.0	140.0
Z_0	0.01	6.0	5.0	740.0	140.0
ROC	0.045	5.0	5.0	1000.0	140.0
Z_0, p^{KS}	0.12	4.0	5.0	140.0	140.0
$Z_0, \chi^2_{\mathrm{red}}$	0.035	4.0	5.0	560.0	140.0
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}$	0.035	4.0	5.0	560.0	140.0
ROC, p^{KS}	0.04	4.0	5.0	960.0	140.0
ROC, χ^2_{red}	0.035	4.0	5.0	560.0	140.0
ROC, $p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}$	0.035	4.0	5.0	560.0	140.0
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}, \mathcal{S}_0$	0.035	4.0	5.0	560.0	140.0
$Z_0, \chi^2_{\mathrm{red}}, \mathcal{S}_1$	0.035	4.0	5.0	560.0	140.0
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}, \mathcal{S}_2$	0.035	4.0	5.0	560.0	140.0

Table B.1: Parameter setting for scan in MVA VBF on the odd split samples.

Criteria	ROC Train	ROC Test	Z_0 Train	Z_0 Test	χ^2_{red} Sig	χ^2_{red} Bg	p^{KS} Sig	p^{KS} Bg
p^{KS}	93.39	92.79	1.63	1.62	0.02	0.44	1.00	1.00
χ^2_{red}	94.02	93.50	1.70	1.70	0.44	0.60	1.00	0.66
Z_0	95.86	95.36	1.99	2.00	1.12	0.90	0.60	0.45
ROC	97.46	96.14	2.00	1.97	1.86	1.21	0.41	0.53
Z_0, p^{KS}	96.43	95.89	1.99	1.99	1.18	0.91	0.87	0.92
$Z_0, \chi^2_{\mathrm{red}}$	96.52	95.95	1.99	1.99	1.22	0.93	0.79	0.84
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}$	96.52	95.95	1.99	1.99	1.22	0.93	0.79	0.84
ROC, p^{KS}	97.04	96.11	1.98	1.98	2.20	1.18	0.56	0.62
ROC, χ^2_{red}	96.52	95.95	1.99	1.99	1.22	0.93	0.79	0.84
ROC, $p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}$	96.52	95.95	1.99	1.99	1.22	0.93	0.79	0.84
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}, \mathcal{S}_0$	96.52	95.95	1.99	1.99	1.22	0.93	0.79	0.84
$Z_0, \chi^2_{\mathrm{red}}, \mathcal{S}_1$	96.52	95.95	1.99	1.99	1.22	0.93	0.79	0.84
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}, \mathcal{S}_2$	96.52	95.95	1.99	1.99	1.22	0.93	0.79	0.84

Table B.2: Figure of merits for parameter scan in MVA VBF on the odd split samples.

Scan Results for the BDT in MVA Boosted

Criteria	Shrinkage	MaxDepth	MinNodeSize	NTrees	NCuts
p^{KS}	0.005	3.0	5.0	260.0	140.0
χ^2_{red}	0.04	3.0	5.0	20.0	140.0
Z_0	0.135	6.0	5.0	960.0	140.0
ROC	0.14	6.0	5.0	1000.0	140.0
Z_0, p^{KS}	0.1	6.0	5.0	560.0	140.0
$Z_0, \chi^2_{\mathrm{red}}$	0.13	5.0	5.0	980.0	140.0
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}$	0.07	3.0	5.0	980.0	140.0
ROC, p^{KS}	0.1	6.0	5.0	520.0	140.0
ROC, χ^2_{red}	0.13	5.0	5.0	980.0	140.0
ROC, $p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}$	0.07	3.0	5.0	980.0	140.0
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}, \mathcal{S}_0$	0.07	3.0	5.0	980.0	140.0
$Z_0, \chi^2_{\mathrm{red}}, \mathcal{S}_1$	0.145	6.0	5.0	740.0	140.0
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}, \mathcal{S}_2$	0.145	6.0	5.0	740.0	140.0

Table B.3: Parameter setting for scan in MVA Boosted on the odd split samples.

Criteria	ROC Train	ROC Test	Z_0 Train	Z_0 Test	χ^2_{red} Sig	χ^2_{red} Bg	p^{KS} Sig	p^{KS} Bg
p^{KS}	90.07	90.20	1.38	1.38	0.57	0.01	1.00	1.00
χ^2_{red}	89.08	89.27	1.38	1.38	0.38	0.09	1.00	1.00
Z_0	93.18	92.84	1.67	1.66	1.52	1.17	0.83	1.00
ROC	93.22	92.85	1.67	1.66	1.62	1.07	0.71	0.99
Z_0, p^{KS}	92.93	92.74	1.65	1.65	0.92	1.40	1.00	1.00
$Z_0, \chi^2_{\mathrm{red}}$	93.11	92.82	1.67	1.66	1.31	1.01	0.99	1.00
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}$	92.71	92.59	1.65	1.65	0.84	1.15	1.00	0.97
ROC, p^{KS}	92.91	92.73	1.65	1.65	0.97	1.04	1.00	1.00
ROC, χ^2_{red}	93.11	92.82	1.67	1.66	1.31	1.01	0.99	1.00
ROC, $p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}$	92.71	92.59	1.65	1.65	0.84	1.15	1.00	0.97
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}, \mathcal{S}_0$	92.71	92.59	1.65	1.65	0.84	1.15	1.00	0.97
$Z_0, \chi^2_{\mathrm{red}}, \mathcal{S}_1$	93.15	92.84	1.67	1.66	1.42	1.00	0.99	0.99
$Z_0, p^{\mathrm{KS}}, \chi^2_{\mathrm{red}}, \mathcal{S}_2$	93.15	92.84	1.67	1.66	1.42	1.00	0.99	0.99

Table B.4: Figure of merits for parameter scan in MVA Boosted on the odd split samples.

Overfitting Plots for BDTs used in Trimming

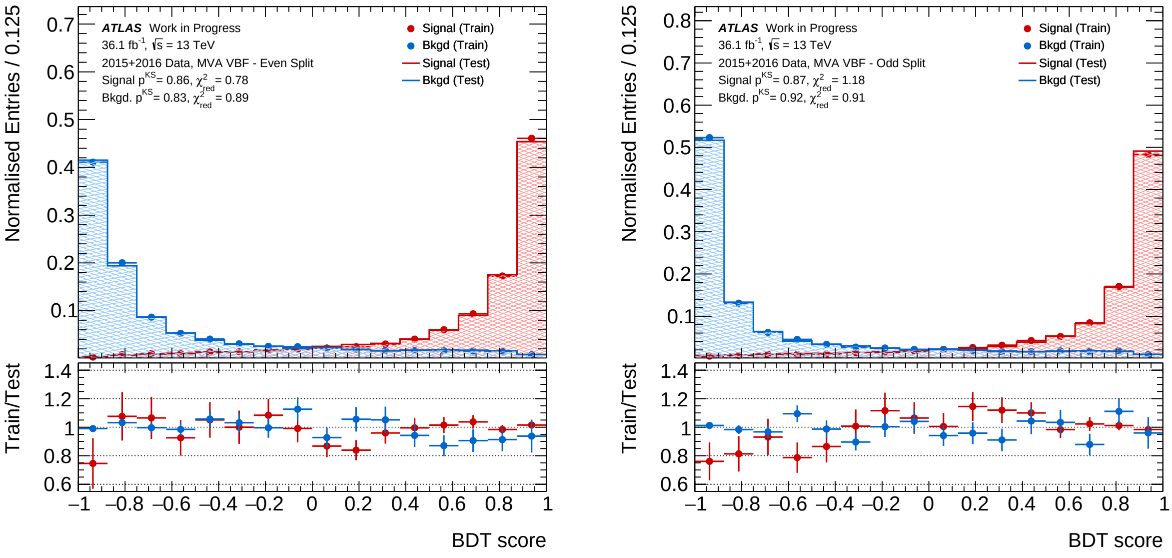


Figure B.14: Overfitting plots for selected BDT in MVA VBF for trimming strategy.

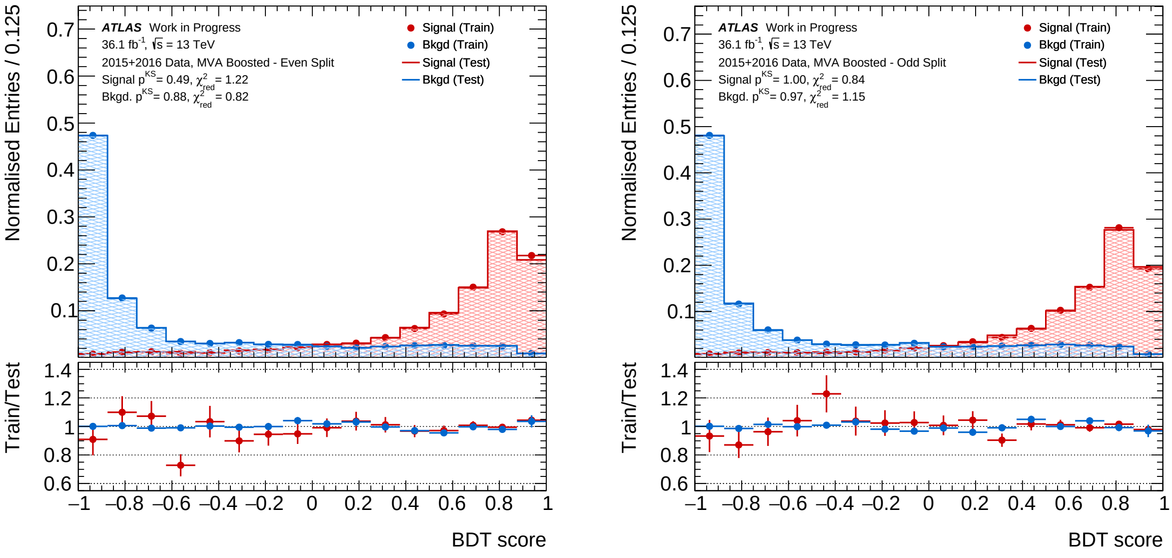


Figure B.15: Overfitting plots for selected BDT in MVA Boosted for trimming strategy.

APPENDIX C

List Of Important Systematic Uncertainties

In the following, sources and naming of the most important systematic uncertainty on the measurement of μ presented in this thesis are listed. Further details can be found in the corresponding publications.

Normalisations and Background Model

norm_boost_Ztt Normalisation factor of the simulated $Z/\gamma^* \rightarrow \tau\tau$ processes in the Boosted signal regions.

norm_vbf_Ztt Normalisation factor of the simulated $Z/\gamma^* \rightarrow \tau\tau$ process in the VBF signal regions.

norm_HH_Fake Normalisation factor of the data-driven Fake background estimate, correlated across the VBF and Boosted signal regions.

hh_fake_contamination Uncertainty on the subtraction of the true τ_{had} contribution from the Fake background template.

hh_dPhiWeightStat Statistical uncertainty on the $\Delta\phi_{\tau\tau}$ based Fake background template correction.

hh_dPhiWeightClosSys Closure uncertainty on the $\Delta\phi_{\tau\tau}$ based Fake background template correction.

Experimental Uncertainties

TAU_EFF_ID_TOTAL Total uncertainty on the τ_{had} identification efficiency measurement.

TAU_EFF_RECO_TOTAL Total uncertainty of the τ_{had} reconstruction methods.

TAU_TES_INSITU Uncertainty of the τ_{had} energy calibration.

jet_jer_np*X* Jet energy resolution nuisance parameters.

jet_jer_crosscalibfwd Jet energy resolution nuisance parameters in the detector forward region.

JES_EtaInter_Model Jet energy scale uncertainty arising from an interpolation to specific detector regions.

JES_PU_Rho Jet energy scale uncertainty for the pileup subtraction method.

JES_EffectiveNP_*X* Jet energy scale uncertainty.

MET_SoftTrk_Scale Calibration scale uncertainty for the $E_{\text{T}}^{\text{miss}}$ soft track term.

MET_SoftTrk_ResoPara Resolution uncertainty on the $E_{\text{T}}^{\text{miss}}$

Uncertainties on Theoretical Predictions

ZttTheory_MUR_MUF Renormalisation and factorisation scale uncertainty.

ZttTheory_PDF Parton density function uncertainty.

ZttTheory_CKK CKKW matching scale uncertainty.

Theo_sig_alphaS QCD scale uncertainty.

Theo_ggH_sig_qcd_*X* QCD scale uncertainty.

APPENDIX D

Nuisance Parameter Correlations

In Chapter (9), nuisance parameter correlation matrices for a fit on Asimov data are presented. In this section the corresponding matrices for the unconditional fit are depicted.

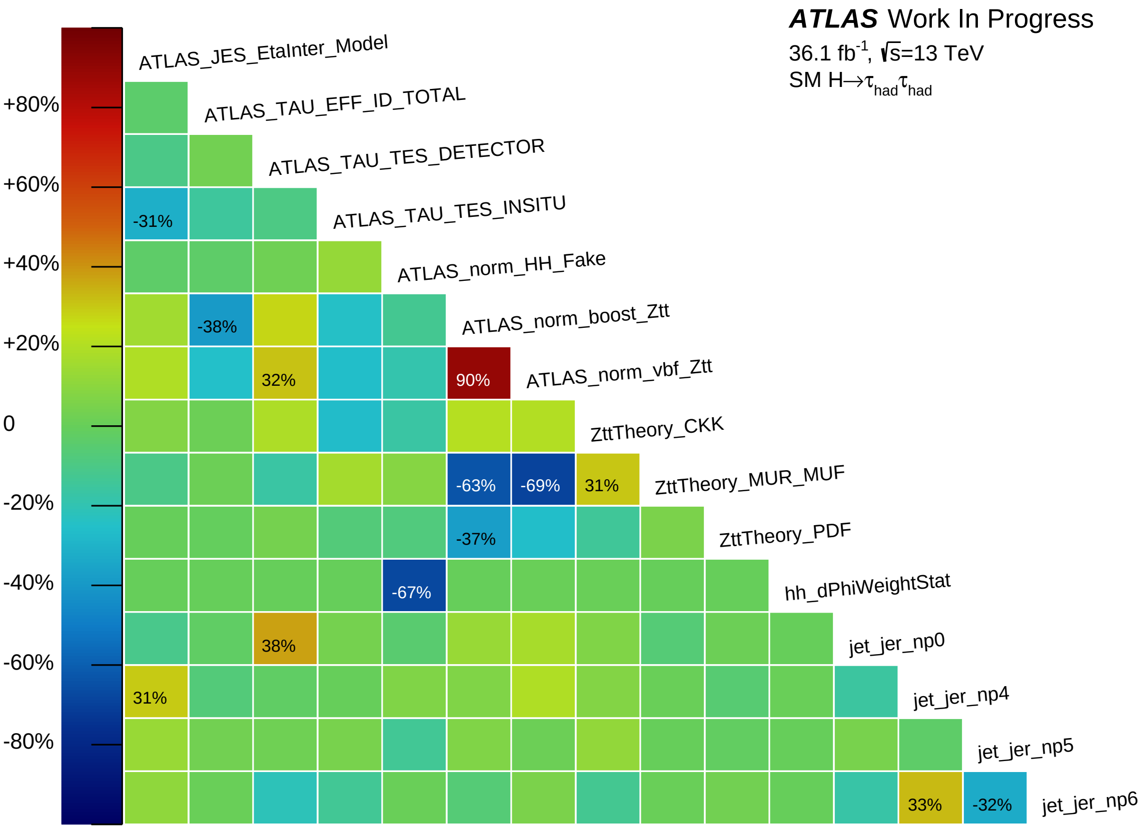


Figure D.1: Correlations between nuisance parameters for the unconditional fit to Data for the cut-based analysis approach.

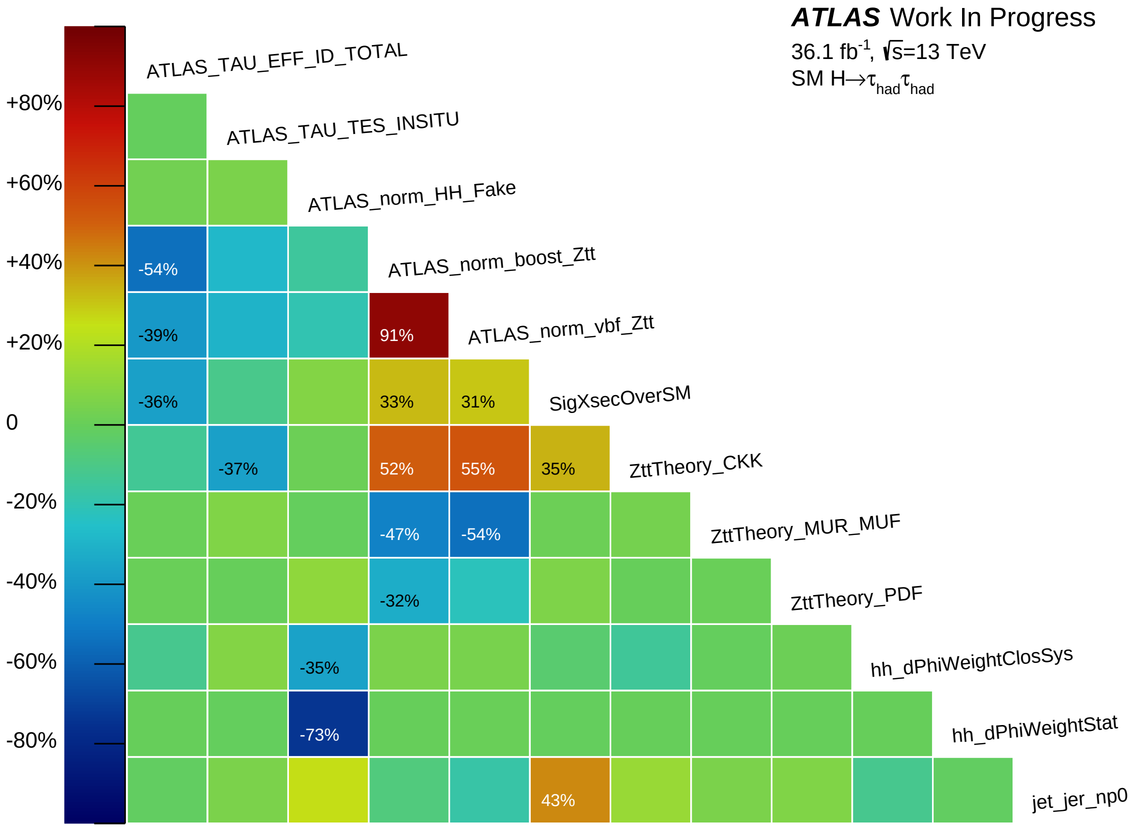


Figure D.2: Correlations between nuisance parameters for the unconditional fit to Data for the multivariate analysis approach.

List of Figures

List of Tables