Measurement of the Higgs boson production in the $H \to \tau\tau \to \tau_{lep}\tau_{had}$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC

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

Antonio De Maria

aus Lamezia Terme (Catanzaro)

Göttingen, 2018

Betreuungsausschuss

Prof. Dr. Arnulf Quadt
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Stan Lai
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Vincenzo Cavasinni
Dipartimento di Fisica, Università di Pisa

Mitglieder der Prüfungskommission:

Referent:	Prof. Dr. Arnulf Quadt II. Physikalisches Institut, Georg-August-Universität Göttingen
Koreferent:	Prof. Vincenzo Cavasinni Dipartimento di Fisica, Università di Pisa

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Stan Lai
II. Physikalisches Institut, Georg-August-Universität Göttingen

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

Prof. Flavio Costantini
Dipartimento di Fisica, Università di Pisa

Dr. Sandra Leone
Istituto Nazionale di Fisica Nucleare, Pisa

Tag der mündlichen Prüfung: 05.07.2018

Referenz: II.Physik-UniGö-Diss-2019/04

Measurement of the Higgs boson production in the $H \to \tau\tau \to \tau_{lep}\tau_{had}$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC

Abstract

This thesis presents a measurement of the Standard Model Higgs boson production in the $H \to \tau\tau$ decay with particular emphasis on the final state in which one tau decays hadronically and the other one decays leptonically, referred to as the $H \to \tau\tau \to \tau_{lep}\tau_{had}$ channel. The analysis uses a dataset corresponding to 36.1 fb^{-1} collected by the ATLAS detector in the 2015 and 2016 Run 2 data taking period at a center of mass energy $\sqrt{s} = 13$ TeV.

The decay of the Higgs boson into a tau lepton pair is currently the only accessible channel to establish the Higgs-Yukawa coupling to leptons. In this context, the $H \to \tau\tau \to \tau_{lep}\tau_{had}$ channel plays an important role due to the high branching ratio and the moderate background.

A cut-based analysis has been performed to enhance the sensitivity to the signal considering the kinematic features of the two dominant Higgs boson production modes: gluon fusion and vector-boson fusion.

The results of a profile likelihood fit to the di-tau mass distribution will be presented considering the $\tau_{lep}\tau_{had}$ standalone fit, as well as the combination of all possible tau pair decay channels. From the combined fit, a signal strength $\mu = 1.08^{+0.175}_{-0.171}(\text{stat.})^{+0.265}_{-0.222}(\text{syst.})$ and the cross section times branching ratio $\sigma_{H\tau\tau}(VBF) = 0.28^{+0.09}_{-0.09}(\text{stat.})^{+0.10}_{-0.09}(\text{syst.})$ pb, $\sigma_{H\tau\tau}(ggH) = 2.97^{+1.03}_{-1.01}(\text{stat.})^{+1.67}_{-1.23}(\text{syst.})$ pb values have been measured. The results are in good agreement with the Standard Model prediction within the uncertainties. The observed (expected) significance, considering only the Run 2 data, is 4.36 (4.13) standard deviations from the background-only hypothesis. The combination of Run 1 and Run 2 data gives a significance of 6.37 (5.43), which is above the 5 standard deviations observation threshold.

Messung der Higgs-Boson-Produktion im $H \to \tau\tau \to \tau_{lep}\tau_{had}$ Zerfallskanal bei $\sqrt{s} = 13$ TeV mit dem ATLAS Detektor

Abstract

Diese Dissertation präsentiert eine Messung der Standard-Modell Higgs-Boson-Produktion im $H \to \tau\tau$ Zerfallskanal mit besonderem Schwerpunkt auf dem Endzustand, in dem ein τ hadronisch und das andere leptonisch zerfällt. Die Analyse verwendet 36,1 fb^{-1} an Daten, die vom ATLAS-Detektor während der Run 2 Datennahme-Perioden in 2015 und 2016 bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV aufgenommen wurden.
Der Zerfall des Higgs-Bosons in ein τ-Lepton-Paar ist gegenwärtig der einzige Kanal, der einen Nachweis der Yukawa-Kopplung des Higgs-Bosons zu Leptonen erlaubt. In diesem Zusammenhang spielt der $H \to \tau\tau \to \tau_{lep}\tau_{had}$ Kanal aufgrund seines großen Verzweigungsverhältnisses und dem gemäßigten Untergrund eine wichtige Rolle.
Eine schnittbasierte Analyse wurde durchgeführt, um die Sensitivität zum Signalprozess zu erhöhen. Diese berücksichtigt die beiden dominanten Higgs-Boson-Produktionsmechanismen: Die Gluon-Fusion und die Vektor-Boson-Fusion.
Das Ergebnis eines Profile-Likelihood-Fits an die Ditau-Massen-Verteilung, sowohl im $\tau_{lep}\tau_{had}$ Sub-Kanal, als auch in Kombination mit allen möglichen τ-Paar-Zerfallskanälen, wird präsentiert. Aus dem kombinierten Fit ergeben sich eine Signalstärke von $\mu = 1,08^{+0,175}_{-0,171}(\text{stat.})^{+0,265}_{-0,222}(\text{syst.})$ und Wirkungsquerschnitt-mal-Verzweigungsverhältnisse von $\sigma_{H\tau\tau}(VBF) = 0,28^{+0,09}_{-0,09}(\text{stat.})^{+0,10}_{-0,09}(\text{syst.})$ pb, $\sigma_{H\tau\tau}(ggH) = 2,97^{+1,03}_{-1,01}(\text{stat.})^{+1,67}_{-1,23}(\text{syst.})$ pb. Die Ergbnisse stehen im Rahmen der Messunsicherheiten in guter Übereinstimmung mit dem Standard-Modell. Die gemessene (erwartete) Signifikanz liegt 4,36 (4,13) Standardabweichungen über der Null-Hypothese, wenn nur die Run 2 Daten berücksichtigt werden. Die Kombination von Run 1 und Run 2 Daten ergibt eine Signifikanz von 6,37 (5,43), was oberhalb der Beobachtungs-Schwelle von fünf Standardabweichungen liegt.

Misura della produzione del bosone di Higgs nel decadimento $H \to \tau\tau \to \tau_{lep}\tau_{had}$ a $\sqrt{s} = 13$ TeV con il detector ATLAS al LHC

Abstract

Questa tesi presenta la misura della produzione del bosone di Higgs all'interno del Modello Standard nel decadimento $H \to \tau\tau$ con particolare enfasi allo stato finale in cui uno dei due τ decade adronicamente e l'altro decade leptonicamente (canale $H \to \tau\tau \to \tau_{lep}\tau_{had}$). L'analisi utilizza un campione di dati corrispondenti a 36.1 fb^{-1} registrati dal rivelatore ATLAS durante il Run 2 nel periodo di presa dati 2015 e 2016 ad un'energia nel centro di massa $\sqrt{s} = 13$ TeV.
Il decadimento del bosone di Higgs in una coppia di leptoni τ è attualmente l'unico modo possibile per misurare l'accoppiamento di Yukawa tra il bosone di Higgs ed i leptoni. In questo contesto, lo stato finale $H \to \tau\tau \to \tau_{lep}\tau_{had}$ ha un ruolo importante dovuto all'alta frazione di decadimento e ad un fondo moderato.
L'analisi è basata su una selezione ottimizzata per aumentare la sensitività del segnale considerando le caratteristiche cinematiche dei principali canali di produzione del bosone di Higgs: la fusione di gluoni e la fusione di bosoni vettori.
I risultati sono stati estratti da un profile likelihood fit sulla massa della coppia di τ e saranno presentati considerando prima solamente lo stato finale $\tau_{lep}\tau_{had}$ e poi considerando tutti i possibili stati finali derivanti dai decadimenti dei due τ. Considerando il fit combinato, sono stati misurati la signal strength $\mu = 1.08^{+0.175}_{-0.171}(\text{stat.})^{+0.265}_{-0.222}(\text{syst.})$ ed il prodotto delle sezioni d'urto per le frazioni di decadimento $\sigma_{H\tau\tau}(VBF) = 0.28^{+0.09}_{-0.09}(\text{stat.})$ $^{+0.10}_{-0.09}(\text{syst.})$ pb, $\sigma_{H\tau\tau}(ggH) = 2.97^{+1.03}_{-1.01}(\text{stat.})^{+1.67}_{-1.23}(\text{syst.})$ pb. I risultati sono in buon accordo con le previsioni del Modello Standard all'interno delle incertezze stimate. La significatività osservata (attesa), considerando solo i dati raccolti nel Run 2, è 4.36 (4.13) deviazioni standard rispetto all'ipotesi di solo fondo. Dalla combinazione dei dati raccolti nel Run 1 e nel Run 2 si ottiene una significatività di 6.37 (5.43), che è superiore alla soglia di 5 deviazioni standard stabilita per l'osservazione di un nuovo processo.

Non è una sfida
non è una rivalsa
non è la finzione di essere meglio
non è la vittoria
l'applauso del mondo di ciò che succede
il senso profondo

Niccolò Fabi - È non è

Contents

1. Introduction **1**

2. The Standard Model of particle physics **3**
2.1. Fundamental particles in the Standard Model . . . 3
2.2. Gauge symmetries . . . 4
2.3. The Higgs boson . . . 9
2.4. Limits of the Standard Model . . . 16

3. The Large Hadron Collider and the ATLAS experiment **21**
3.1. The Large Hadron Collider . . . 21
3.2. The ATLAS detector . . . 24

4. τ reconstruction and identification **31**
4.1. Reconstruction of hadronic τ decays . . . 31
4.2. Identification . . . 32
4.3. $\tau_{had\text{-}vis}$ identification efficiency measurement . . . 32

5. Di-tau mass estimation **53**
5.1. The Collinear Approximation Technique . . . 53
5.2. *Missing Mass Calculator (*MMC*)* . . . 54
5.3. MMC retuning for the $\tau_{lep}\tau_{had}$ final state . . . 56
5.4. Comparison of the Collinear Approximation and MMC mass estimation . . . 59

6. Search for $H \to \tau\tau$ decay in the $\tau_{lep}\tau_{had}$ final state **61**
6.1. Data and Monte Carlo Samples . . . 62
6.2. Object Reconstruction . . . 63
6.3. Preselection Region . . . 65
6.4. Background estimation . . . 67
6.5. Signal Regions definition . . . 86

6.6. Systematic Uncertainties . . . 100
6.7. Fit model description . . . 105
6.8. Binning optimisation strategy . . . 110
6.9. Results for standalone $\tau_{lep}\tau_{had}$ fit . . . 112
6.10. Results for Combined Fit . . . 123

7. Analysis improvements **133**
7.1. Usage of E_T^{miss} Significance to reject $Z \to ll$ background . . . 133
7.2. Addition of a low electron/muon p_T region using TLT trigger . . . 135

8. Conclusion **151**

Appendices **155**

A. OS-SS background estimation correction factors **157**

B. Nuisance parameters list **161**

Bibliography **165**

CHAPTER 1

Introduction

The discovery of the Higgs boson in 2012 [1],[2], is an important milestone in investigating the origin of electroweak symmetry breaking, and the experimental confirmation of the Brout-Englert-Higgs mechanism [3],[4],[5],[6] in the Standard Model of particle physics.
The measurements rely principally on studies of the bosonic decay modes, $H \to \gamma\gamma$, $H \to ZZ^*$ and $H \to WW^*$. In order to establish the mass generation also for fermions, it is important to demonstrate the direct coupling of the Higgs boson to fermions and its proportionality to mass. In this context, the decay of the Higgs boson into a τ pair is currently one of the most important candidate due to the distinct $H \to \tau\tau$ final state topologies and good τ reconstruction and identification in the ATLAS detector.
An excess of events over the expected background from the other Standard Model processes was measured by ATLAS with an observed (expected) significance of 4.5 (3.5) standard deviations during the Run 1 data taking period [7]. Combining ATLAS and CMS analyses, the coupling of the Higgs boson to the fermion sector has been established with the observation of the $H \to \tau\tau$ decay mode with a signal significance of 5.5 standard deviations [8].
Using the collision data available now at $\sqrt{s} = 13$ TeV, the detailed program of the Higgs boson property measurements will be extended to reach a higher precision compared to 7 and 8 TeV analyses, due to the expected increase of data statistics and the expected increase of the Higgs boson production cross section. The $H \to \tau\tau$ channel will continue to play an important role in terms of measurements of the Higgs boson couplings to τ leptons, as well as measurements of the other properties of the Higgs boson, such as its charge-parity (CP) quantum numbers.
The thesis is organised as follows. In the second chapter, a summary of the Standard Model of particle physics is presented with particular focus on the Higgs boson sector. The third chapter contains the description of the experimental setup of the LHC

and the ATLAS experiment. The fourth and the fifth chapters give an overview of the reconstruction and identification of the tau leptons in ATLAS, and the invariant mass estimate of a resonance decaying into a τ pair. These are both fundamental bases for the description of the $H \to \tau\tau$ search, which is presented in the sixth chapter with emphasis on final state in which one τ decays hadronically and the other one leptonically. In the seventh chapter, several analysis improvements are discussed. Finally, the last chapter contains a summary of the analysis results.

CHAPTER 2

The Standard Model of particle physics

The *Standard Model* (SM) is the quantum field theory that, up to now, gives the best description of the fundamental interaction phenomenology. It gives the theoretical formulation of three of the four interactions: electromagnetic, weak and strong interactions. Gravity is negligible at SM energy scale and it is not included in the model. In the context of the SM, the Higgs boson plays an important role; this particle results as consequence of the *spontaneous symmetry breaking*, which, through the *Higgs mechanism*, gives mass to the other fundamental particles.

2.1. Fundamental particles in the Standard Model

The SM is a quantum field theory which describes the universe in terms of interacting particles which can be divided in two categories:

- *fermions*, half integer spin particles which are the fundamental components of the matter and can be subsequently divided into *leptons* and *quarks*;
- *bosons*, integer spin particles which are the force carriers.

Fermions are grouped in three families, called *generations*, which are ordered according their mass. Each lepton family consists of a charged lepton, which can interact electromagnetically or weakly, and a neutrino, which can interact only weakly; a summary of the lepton properties is shown in Table 2.1 .
Quarks are also organised into three generations with six *flavour* types; they have non-integer charge and they are subject to strong, weak and electromagnetic forces. They are not present in nature as free particles, instead they are combined together to form *hadrons* which can be divided in *baryons* (qqq) and *mesons* ($q\bar{q}$). A summary of the

Particle	Symbol	Charge	Spin	Mass (MeV/c^2)
Electron Neutrino	ν_e	0	1/2	$\leq 2 \times 10^{-6}$
Electron	e	-1	1/2	0.511
Muon Neutrino	ν_μ	0	1/2	≤ 0.17
Muon	μ	-1	1/2	105.6
Tau Neutrino	ν_τ	0	1/2	≤ 15.5
Tau	τ	-1	1/2	1776.82

Table 2.1.: Summary of the properties of the three leptonic families in the SM [9].

Particle	Symbol	Charge	Spin	Mass (GeV/c^2)
Up	u	2/3	1/2	0.0023
Down	d	-1/3	1/2	0.0048
Charm	c	2/3	1/2	1.275
Strange	s	-1/3	1/2	0.095
Top	t	2/3	1/2	173.21
Bottom	b	-1/3	1/2	4.18

Table 2.2.: Summary of the properties of the three quark families in the SM [9].

quark properties can be found in Table 2.2 .

As already mentioned, bosons are the mediator of the fundamental interactions: the photon is the electromagnetic force carrier, gluons are the strong force carriers, and finally $W^\pm$ and Z^0 are responsible for weak interactions. A summary of the boson properties is shown in Table 2.3 .

Particle	Symbol	Charge	Spin	Mass (GeV/c^2)
$W^\pm$	$W^\pm$	± 1	1	80.39
Z	Z	0	1	91.18
Photon	γ	0	1	0
Gluons	g_α, $\alpha = 1, ..., 8$	0	1	0

Table 2.3.: Summary of the properties of the gauge bosons in the SM [9].

2.2. Gauge symmetries

The mathematical formulation of the SM is based on gauge invariance of the quantum relativistic theory [10]: this theory was born around 1940 to describe *Quantum*

Electrodynamics (QED). Also weak interactions (Glashow, Salam and Weinberg theory [11],[12],[13]) and strong interactions (Yang-Mills theory [14]) are well described by gauge invariance.
When a physical law has a symmetry, it means that it remains unchanged after a transformation despite variations of the system. A symmetry can be defined as *global* when the transformation is not dependent on space-time coordinates, otherwise it can be referred to as *local* symmetry. A physical system which is described by a Lagrangian which is invariant under local symmetries is called *gauge invariant*. The gauge symmetry group of the SM is given by the product:

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$$

where $SU(3)_C$ is the symmetry group of strong interactions associated with *colour* (red, blue, green) quantum number conservation. This theory is known as *Quantum Chromodynamics* (QCD).
The $SU(2)_L \otimes U(1)_Y$ is the symmetry group of the electroweak sector (EWK), in which electromagnetic and weak interactions are grouped in a unique theory. The L indicates that the $SU(2)$ group is related to the weak *left-handed* isospin doublet, while Y refers to the weak *right-handed* hypercharge singlets.
The SM is a local theory and, given the contributions of two symmetry groups, it is possible to decompose the full SM Lagrangian as sum of two components:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EWK}$$

2.2.1. QED

Before describing the SM Lagrangian, it is necessary to introduce QED gauge invariance [15]. The starting point is the Lagrangian of a massive fermion represented by the field ψ:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{2.1}$$

where γ^μ are the Dirac's matrices and $\bar{\psi} = \psi^\dagger \gamma^0$. The Lagrangian in Eqn. 2.1 is invariant under the transformation of the group $U(1)$:

$$\psi \xrightarrow{U(1)} \psi' = e^{iQ\theta(x)}\psi \tag{2.2}$$

where Q is the electric charge operator and it is the generator of the group $U(1)$; $\theta(x)$ is a coordinate dependent phase.
Gauge invariance is assured by the substitution of partial derivative with the *covariant derivative*:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iQA_\mu \tag{2.3}$$

and imposing that the potential vector A_μ follows the transformation:

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \theta(x) \tag{2.4}$$

Substituting the previous relations in Eqn. 2.1, the Lagrangian becomes:

$$\mathcal{L} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - Q(\bar{\psi}\gamma^{\mu}\psi)A_{\mu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.5}$$

where the gauge invariance is recovered by the interaction term between the fields ψ and A_{μ}. Moreover, a new term containing the tensor $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ is added. The gauge invariance forbids a mass term for the field A_{μ}, so the photon is massless as confirmed by many experimental results.

2.2.2. QCD

The procedure described for QED can be extended also to QCD, where the $U(1)$ group can be substituted by the $SU(3)$ group. This group describes the transformations of the colour fields of the quarks. The free Lagrangian is given by:

$$\mathcal{L}_0 = \bar{q}_j(i\gamma^{\mu}\partial_{\mu} - m)q_j \qquad j = 1, 2, 3 \tag{2.6}$$

where q_1, q_2 and q_3 are the colour fields. The first step to get the gauge invariance is to require that $\mathcal{L}_0$ in 2.6 is invariant under local transformation of quark fields:

$$q(x) \xrightarrow{SU(3)} q^{'}(x) = e^{i\alpha(x)_a T_a} q(x) \tag{2.7}$$

where $\alpha(x)_a$ is a coordinate-dependent phase and T_a ($a = 1, ..., 8$) are the $SU(3)$ group generators. Conventionally $T_a = \lambda_a/2$, where λ_a are the Gell-Mann matrices. The $SU(3)$ group is a non abelian group because the generators T_a do not commute:

$$[T_a, T_b] = i f_{abc} T_c \tag{2.8}$$

where f_{abc} are the structure constants of the group.
As in QED, in QCD the partial derivative should also have to be substituted by the covariant derivative:

$$\partial_{\mu} \rightarrow D_{\mu} = \partial_{\mu} + igT_a G^a_{\mu} \tag{2.9}$$

where G^a_{μ} are the eight gauge fields associated with the gluons which follow the transformation:

$$G^a_{\mu} \rightarrow G^a_{\mu} - \frac{1}{g}\partial_{\mu}\alpha_a - f_{abc}\alpha_b G^c_{\mu} \tag{2.10}$$

where g is the coupling constant of the strong interactions. The gauge invariant Lagrangian for the QCD is then given by:

$$\mathcal{L} = \bar{q}(i\gamma^{\mu}\partial_{\mu} - m)q - g(\bar{q}\gamma^{\mu}T_a q)G^a_{\mu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{2.11}$$

The first term in Eqn. 2.11 represents the free term, the second term represents the interaction between quarks and gluons, and the last one is the kinematic term between the two tensors $G^a_{\mu\nu}$ which are defined as:

$$G^a_{\mu\nu} = \partial_{\mu}G^a_{\nu} - \partial_{\nu}G^a_{\mu} - g f_{abc} G^b_{\mu} G^c_{\nu} \tag{2.12}$$

Substituting Eqn. 2.12 in Eqn. 2.11, it can be observed that this term is not purely kinematic but it leads to gluon self-interaction. Finally, gauge invariance requires that gluons are also massless.

2.2.3. The electroweak theory

The electroweak theory of Weinberg and Salam was born to unify the QED and the weak interactions, that is with the aim to describe two different forces as manifestation of the same interaction. Its group symmetry is $SU(2)_L \otimes U(1)_Y$, and its generators are the weak isospin operator $\vec{T} = \vec{\sigma}/2$ (where $\vec{\sigma}$ are the Pauli matrices) and the hypercharge operator Y. It is a *chiral* theory, that is the *left-* and *right-handed* fermion components transform in different way under local gauge transformations:

$$\begin{aligned} \psi_L \rightarrow \psi_L^{'} = e^{i\vec{\alpha}(x)\vec{T}+i\beta(x)Y}\psi_L \\ \psi_R \rightarrow \psi_R^{'} = e^{i\beta(x)Y}\psi_R \end{aligned} \tag{2.13}$$

where $\alpha(x)$ and $\beta(x)$ are the phases of the local gauge transformations, ψ_L is the isospin doublet which represents the *left-handed* fermions and ψ_R is the isospin singlet which represents the *right-handed* fermions:

$$\psi_{l_L} = \begin{pmatrix} \widehat{\nu}_l \\ \widehat{l} \end{pmatrix} \qquad \psi_{q_L} = \begin{pmatrix} \widehat{u} \\ \widehat{d} \end{pmatrix} \qquad \psi_{l_R} = \widehat{l}_R \qquad \psi_{q_R} = \widehat{u}_R, \widehat{d}_R \tag{2.14}$$

where $\widehat{\nu}_l = \nu_e, \nu_\mu, \nu_\tau$, $\widehat{l} = e, \mu, \tau$, $\widehat{u} = u, c, t$ and $\widehat{d} = d, s, b$. The hypercharge Y is related to the electric charge Q and to the third component of the isospin through the relation:

$$Q = T_3 + \frac{Y}{2} \tag{2.15}$$

Table 2.4 summarises the quantum numbers for the electroweak group. The gauge

Particle	Charge Q	Isospin T_3	Hypercharge Y
$(\widehat{\nu_l})_L$	0	1/2	-1
$(\widehat{l})_L$	-1	-1/2	-1
$(\widehat{l})_R$	-1	0	-2
$(\widehat{u})_L$	2/3	1/2	1/3
$(\widehat{d})_L$	-1/3	-1/2	1/3
$(\widehat{u})_R$	2/3	0	4/3
$(\widehat{d})_R$	-1/3	0	-2/3

Table 2.4.: Quantum numbers of the electroweak group [9].

invariance relations for this theory bring to the following electroweak Lagrangian:

$$\begin{aligned}\mathcal{L}_{EWK} = \quad & \bar{\psi}_L\gamma^\mu\big[i\partial_\mu - g\frac{1}{2}\vec{\tau}\vec{W}_\mu + \frac{g^{'}B_\mu}{2}\big]\psi_L + \bar{\psi}_R\gamma^\mu[i\partial_\mu + g^{'}B_\mu]\psi_R \\ & - \frac{1}{4}\vec{W}_{\mu\nu}\vec{W}^{\mu\nu} - \frac{1}{4}\vec{B}_{\mu\nu}\vec{B}^{\mu\nu}\end{aligned} \tag{2.16}$$

where the $\vec{W}_\mu$ are the fields of the $SU(2)_L$ group, B_μ is the field of the $U(1)_Y$ group, and g and $g^{'}$ are the couplings of the corresponding interactions. In analogy with QED, the partial derivative can be substituted with the covariant derivative

$$D_\mu = \partial_\mu + ig\vec{T}W_\mu(x) + ig^{'}\frac{Y}{2}B_\mu(x) \tag{2.17}$$

Moreover, the fields transform according to the following relations:

$$\begin{aligned}W_\mu(x) \rightarrow W^{'}_\mu(x) = W_\mu(x) - \tfrac{1}{g}\partial_\mu\alpha(x) - \alpha(x)\times W_\mu(x) \\ B_\mu(x) \rightarrow B^{'}_\mu(x) = B_\mu(x) - \tfrac{1}{g^{'}}\partial\mu\beta(x)\end{aligned} \tag{2.18}$$

Experimental results have shown the existence of two charged and massive vector bosons, $W^\pm_\mu$, which correspond to combinations of the first two components of $\vec{W}_\mu$:

$$W^\pm_\mu = \frac{W^1_\mu \mp iW^2_\mu}{\sqrt{2}} \tag{2.19}$$

Also two neutral bosons (the Z and the photon) have been observed; they are given by the combinations of W^3_μ and B_μ, which are both neutral and massless. The combination is possible through the introduction of the *Weinberg angle* (θ_w):

$$\begin{aligned}A_\mu = \cos\theta_w B_\mu + \sin\theta_w W^3_\mu \\ Z_\mu = \cos\theta_w W^3_\mu + \sin\theta_w B_\mu\end{aligned} \tag{2.20}$$

The coupling constants g and $g^{'}$ are related to the electromagnetic coupling constant e through the Weinberg angle:

$$e = g\sin\theta_w = g^{'}\cos\theta_w \tag{2.21}$$

The last two terms in Eqn. 2.16 are the kinematic terms of the vector bosons, in which the following tensorial fields are used:

$$\begin{aligned}W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - gW_\mu \times W_\nu \\ B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu\end{aligned} \tag{2.22}$$

Also, in this case, the gauge invariance forbids the presence of massive terms for bosons and fermions, such as $m^2W_\mu W^\nu$ for bosons and $m^2f\bar{f}$ for fermions; this is in contrast to experimental observations.

2.3. The Higgs boson

To explain the presence of massive fermions and bosons in the SM, it is necessary to postulate the existence of a scalar boson, called the *Higgs* boson. The particles get a mass by interacting with the Higgs field through the so-called *Higgs mechanism* [16],[17],[18] and preserving the gauge invariance.

2.3.1. The Higgs Mechanism

In order to describe how the vector bosons get a mass in the SM, it is simpler to start from spontaneous global gauge symmetry breaking. We consider a complex scalar field $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ and the Lagrangian:

$$\mathcal{L} = (\partial_\mu \phi)^*(\partial^\mu) - V = (\partial_\mu \phi)^*(\partial^\mu) - \mu^2 \phi^* \phi - \lambda(\phi^* \phi)^2 \tag{2.23}$$

where $\lambda > 0$. If $\mu^2 > 0$, the fundamental state is unequivocably defined for $\phi = 0$. Considering the case $\mu^2 < 0$, the system has a round shape with radius v of potential minima (Fig. 2.1):

$$\phi_1^2 + \phi_2^2 = v^2 = -\frac{\mu^2}{\lambda} \tag{2.24}$$

Without loss of generality, it is possible to translate the field at the minimum of the

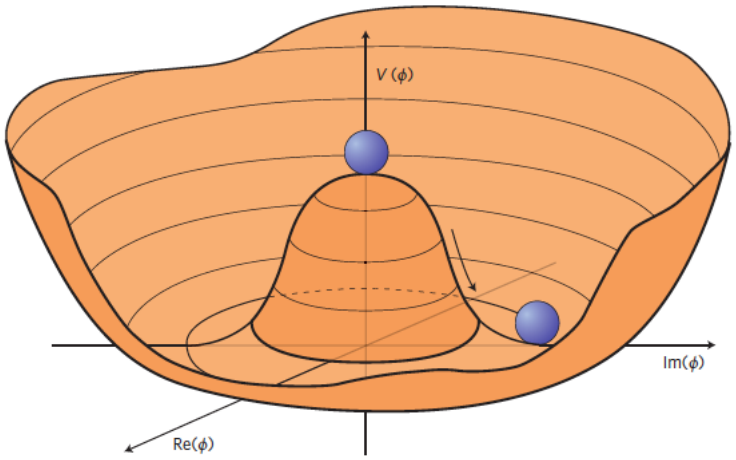


Figure 2.1.: The Higgs potential for $\mu^2 < 0$.

potential and to fix the minimum at $\phi_1 = v$ and $\phi_2 = 0$. The Lagrangian can then be expressed around the minimum in terms of the fields η and ξ, so the field ϕ becomes:

$$\phi(x) = \sqrt{\frac{1}{2}}[v + \eta(x) + i\xi(x)] \tag{2.25}$$

and the Lagrangian in Eqn. 2.23 can be re-written as:

$$\mathcal{L}^{'} = \frac{1}{2}(\partial_\mu \xi)^2 + \frac{1}{2}(\partial^\mu \eta)^2 + \mu^2\eta^2 + \text{higher order terms in } \eta \text{ and } \xi \tag{2.26}$$

In this way, through the symmetry breaking, the scalar field has acquired mass $m_\mu = \sqrt{-2\mu^2}$. Moreover, there is a kinematic term for the field ξ which represents a massless scalar particle, known as the *Goldstone boson*. So the symmetry breaking gives mass to the gauge bosons and generates massless particles given by the choice of the expansion around the potential minimum.

Gauge boson masses

In order to give mass to the $W^\pm$ and Z^0 bosons and to leave the photon massless, it is necessary to extend the spontaneous symmetry breaking mechanism to a local theory and to the $SU(2)_L \otimes U(1)_Y$ group. So it is necessary to add to the electroweak Lagrangian in Eqn. 2.16 an invariant term under $SU(2)_L \otimes U(1)_Y$ using four scalar fields, ϕ_i, to get:

$$\mathcal{L}_{EWK} = \left|\left(i\partial_\mu - g\vec{T}\vec{W}_\mu - g^{'}\frac{Y}{2}B_\mu\right)\phi\right|^2 - V(\phi) \tag{2.27}$$

To preserve the gauge invariance, the fields ϕ_i should belong to an isospin multiplet with hypercharge $Y = 1$:

$$\phi = \begin{pmatrix}\phi^+ \\ \phi^0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2 \\ \phi_3 + i\phi_4\end{pmatrix} \tag{2.28}$$

Using the Lagrangian in Eqn. 2.23, considering $\mu^2 < 0$ and $\lambda > 0$ and the complex field just defined, the potential has a minimum for:

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2} \tag{2.29}$$

It is possible to choose the minimum of the potential such that $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$:

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0 \\ v\end{pmatrix} \tag{2.30}$$

This ensures the electroweak spontaneous symmetry breaking, and consequently that vector bosons acquire masses, and the invariance of $U(1)_{EM}$, which implies a massless photon.
The perturbative expansion around the minimum introduces four scalar fields $\theta_1(x)$, $\theta_2(x)$, $\theta_3(x)$ and $h(x)$. Thus the field ϕ becomes:

$$\phi = e^{i\vec{\sigma}\vec{\theta}(x)/v}\begin{pmatrix}0 \\ \frac{v+h(x)}{\sqrt{2}}\end{pmatrix} \simeq \frac{1}{\sqrt{2}}\begin{pmatrix}\theta_2(x) + i\theta_1(x) \\ v + h(x) - i\theta_3(x)\end{pmatrix} \tag{2.31}$$

The three fields θ_1, θ_2 and θ_3 are the massless Goldstone bosons which are generated by the electroweak spontaneous symmetry breaking. Moreover, the Lagrangian is still

locally invariant under $SU(2)$; thus it is possible to *eliminate* the Goldstone bosons through the appropriate gauge choice. The ϕ field will become:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.32}$$

in which $h(x)$ is the neutral scalar field of the Higgs boson. The scalar degrees of freedom corresponding to the Goldstone bosons will become the longitudinal polarisation of the the massive vector bosons. The masses of the gauge bosons are then given by:

$$M_W = \frac{1}{2}vg \qquad M_Z = \frac{1}{2}\sqrt{g^2 + g'^2} \tag{2.33}$$

while the photon is massless.
It is also possible to determine the value of v using the experimental values of the Fermi coupling constant G_F determined by the muon decay. This is related to the $W^{\pm}$ mass by:

$$\frac{G_F}{\sqrt{2}} = \frac{g}{8M_W^2} \tag{2.34}$$

from which it is possible to get:

$$v^2 = \frac{1}{\sqrt{2}G_F} \simeq (246 \text{ GeV})^2 \tag{2.35}$$

Finally the potential $V(\phi)$ contains the following terms:

$$V(\phi) = -\lambda v^2 h(x) - \lambda v h^3(x) - \frac{1}{4}h^4(x) \tag{2.36}$$

where the second and the third term represent the Higgs boson interaction with gauge fields and the self-interaction. The first term gives the mass of the Higgs boson:

$$M_H = \sqrt{2v^2\lambda} \tag{2.37}$$

which is not predicted by the theory since the value of λ is not known.

Fermion masses

Since it is not possible to insert mass terms for fermions in the Lagrangian in Eqn. 2.16 due to gauge invariance, we can use the same Higgs doublet in Eqn. 2.31 to generate quark and lepton masses. It is necessary to add an invariant term under $SU(2)_L \otimes U(1)_Y$ to the Lagrangian in Eqn. 2.16. Using Eqn. 2.14, the Lagrangian can be written as:

$$\mathcal{L}_{leptons} = -G_l\left[\bar{\psi}_{l_L}\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}\psi_{l_R} + h.c.\right] \tag{2.38}$$

where G_l is the coupling constant between the lepton l and the field ϕ. Introducing the Higgs field and breaking the symmetry, the Lagrangian will contain the lepton mass term and the interaction term between the lepton and the Higgs field:

$$\mathcal{L}_{leptons} = \frac{G_l}{\sqrt{2}}v(\bar{l}_L l_R - \bar{l}_R l_L) - \frac{G_l}{\sqrt{2}}(\bar{l}_L l_R - \bar{l}_R l_L)h \tag{2.39}$$

from which it follows that:

$$m_l = \frac{G_l v}{\sqrt{2}} \tag{2.40}$$

Quark masses are generated in a similar way, however, the only difference is that for quarks the right-hand components are also massive. Writing the Higgs double in the following way:

$$\phi_C = -i\sigma_2\phi^* = \begin{pmatrix} -\bar{\phi}^0 \\ \phi^- \end{pmatrix} \tag{2.41}$$

the additional Lagrangian for quarks is [1]:

$$\mathcal{L}_{quark} = -G_d^{ij}\bar{\psi}'_{iL}\begin{pmatrix}\phi^+ \\ \phi^0\end{pmatrix}d_{jR} - G_u^{ij}\bar{\psi}'_{iL}\begin{pmatrix}-\bar{\phi}^0 \\ \phi^-\end{pmatrix}u_{jR} + h.c. \tag{2.42}$$

where:

$$\psi'_{i_L} = \begin{pmatrix} u \\ d' \end{pmatrix}_L \tag{2.43}$$

Finally, it is possible to write the Lagrangian in the diagonal form:

$$\mathcal{L}_{quarks} = -m_d^i \bar{d}_i d_i\left(1 + \frac{h}{v}\right) - m_u^i \bar{u}_i u_i\left(1 + \frac{h}{v}\right) \tag{2.44}$$

where the symmetry is broken through choosing a particular vacuum state.
It is important to note that fermion masses are free parameters of the theory and their values are not predicted, so it is important to measure them experimentally.

2.3.2. Higgs boson production at LHC

The main Higgs boson production mechanisms at the *Large Hadron Collider* (LHC), described in Chapter 3, are gluon fusion, vector boson fusion, associated production with a gauge boson, and associated production with a $t\bar{t}$ pair. Feynman diagrams for each production mode are shown in Fig. 2.2. The cross sections for the production of a SM Higgs boson as a function of the center of mass energy ($\sqrt{s}$) for proton-proton collisions are shown in Fig. 2.3 [19].
In the following part of this section, each production mode will be discussed.

[1] the index for the quarks is referring to the *Cabibbo mixing* of the flavour eigenstate. In particular:

$$d_i^{'} = \sum_{n=1}^{N} V_{in} d_n$$

where N is the number of quarks ($N = 3$), V_{in} is the element in in the CKM matrix (V_{CKM}) and d_n are the d, s and b quarks, respectively ($n = 1, 2, 3$)

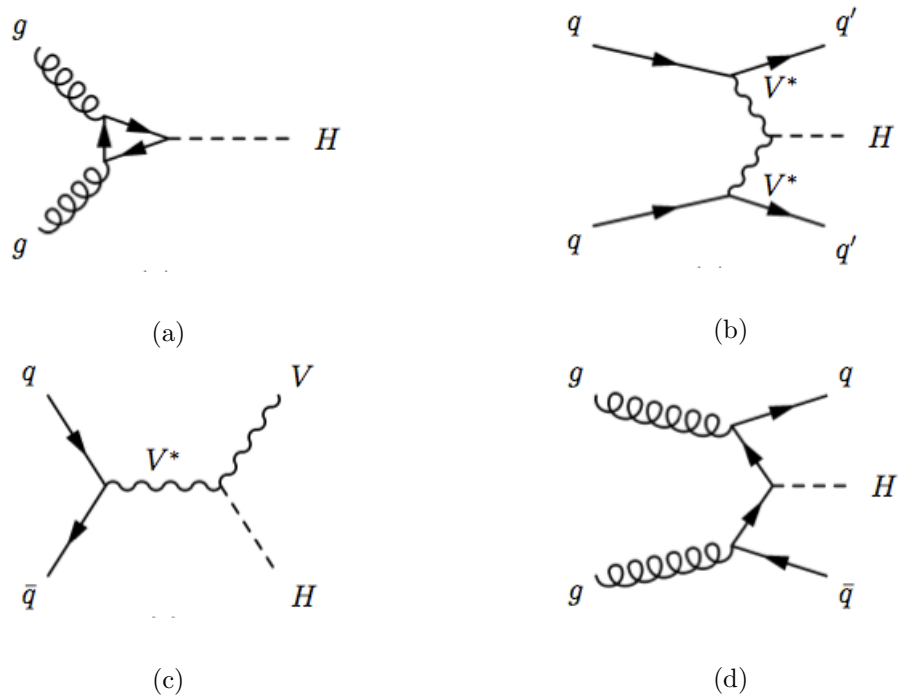


Figure 2.2.: Higgs boson production modes : *a)* gluon fusion, *b)* vector boson fusion, *c)* associated production with a gauge boson, *d)* associated production with a $t\bar{t}$ pair.

Gluon fusion production mechanism

At high-energy hadron colliders, the Higgs boson production mechanism with the largest cross section is the gluon fusion process, $gg \to H + X$, mediated by the exchange of a virtual top quark [20]. Contributions from lighter quarks propagating in the loop are suppressed proportional to m_q^2. The leading Order (LO) and next to leading order (NLO) QCD corrections amount to about 80 % of the total cross section at next to next to next to leading order (N3LO). The next to next to leading order (NNLO) corrections further enhance the cross section approximately by 30 % of the LO+NLO result. Electroweak radiative corrections have been computed at NLO and increase the LO cross section by about 5 % for $m_H = 125$ GeV [21]. At the LHC with a center of mass energy of 13 TeV, the most up-to-date value for the production cross section of a 125 GeV Higgs boson is:

$$\sigma_{ggF}^{N3LO} = 48.6 \text{ pb}^{+2.2 \text{ pb}}_{-3.3 \text{ pb}}(\text{theory}) \pm 1.6 \text{ pb}(\text{PDF} + \alpha_S)$$

Vector boson fusion production mechanism

The SM Higgs production mode with the second-largest cross section at the LHC is the vector-boson fusion (VBF). Higgs production via VBF, $qq \to qqH$, proceeds by the scattering of two (anti-)quarks, mediated by the t- or u-channel exchange of a W or Z boson, with the Higgs boson radiated off the weak boson propagator. The scattered quarks will

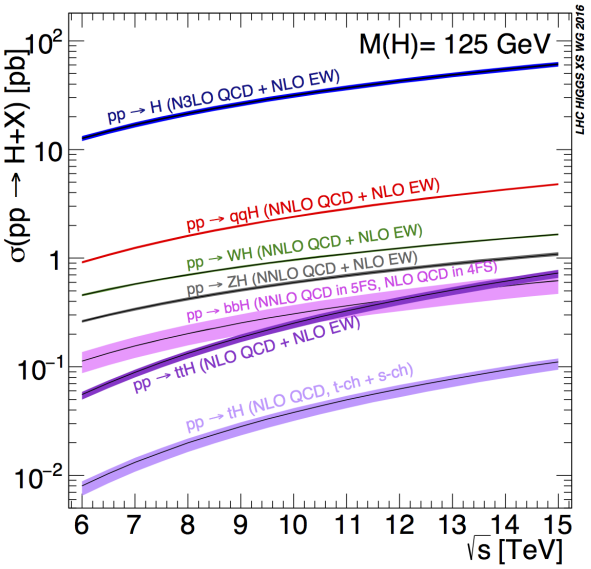


Figure 2.3.: SM Higgs boson production cross sections as a function of the center of mass energy, $\sqrt{s}$, for pp collisions.

produce two back-to-back hard jets in the forward and backward regions of the detector. Because of the colour-singlet nature of the weak-gauge boson exchange, gluon radiation from the central-rapidity regions is strongly suppressed. These characteristic features of VBF processes can be used to distinguish them from the QCD backgrounds, including gluon fusion induced Higgs +2 jet production, and from s-channel WH or ZH production with a hadronically decaying weak gauge boson [22]. The production cross section of a 125 GeV Higgs boson at LHC is:

$$\sigma_{VBF} = 3.748 \text{ pb}^{+0.026 \text{ pb}}_{-0.026 \text{ pb}}(\text{theory}) \pm 0.120 \text{ pb}(\text{PDF} + \alpha_S)$$

WH and ZH associated production mechanism

The next most relevant Higgs boson production mechanisms after gluon fusion and VBF are associated production with W and Z gauge bosons. The cross section for the associated production processes, $pp \to VH + X$, with $V = W^{\pm}, Z$, receive contributions at NLO given by NLO QCD corrections to the Drell-Yan cross section and from NLO EW corrections [23],[24],[25]. At NNLO, the Drell-Yan-like corrections to WH production also give the bulk of the corrections to ZH production [26]. For ZH production, there are also gluon-gluon induced contributions that do not involve a virtual Z gauge boson but are such that the Z boson and H boson couple to gluons via top-quark loops [27]. In addition, WH and ZH production receive non Drell-Yan-like corrections at the NNLO level, where the Higgs is radiated off top-quark loops [28]. The production cross section

of a 125 GeV Higgs boson at the LHC is:

$$\sigma_{WH} = 1.380\ \text{pb}^{+0.01\ \text{pb}}_{-0.02\ \text{pb}}(\text{theory}) \pm 0.03\ \text{pb}(\text{PDF} + \alpha_S)$$

$$\sigma_{ZH} = 0.8696\ \text{pb}^{+0.033\ \text{pb}}_{-0.03\ \text{pb}}(\text{theory}) \pm 0.019\ \text{pb}(\text{PDF} + \alpha_S)$$

Higgs production in association with $t\bar{t}$ pair

The top quark, being the heaviest elementary fermion known in the SM, is predicted to have the largest coupling to the Higgs boson. The associated production of top quarks with the Higgs boson, either in pairs ($t\bar{t}H$) or singly (tH), provides direct experimental access to the top-Higgs coupling. The $t\bar{t}H$ (tH) production mode, while proceeding at a rate of about 100 (1000) times smaller than gluon fusion, bears a highly distinctive experimental signature, which includes leptons and/or jets from the decay of the two (single) top quarks. The production cross section of a 125 GeV Higgs boson at the LHC is:

$$\sigma_{ttH} = 0.508\ \text{pb}^{+0.029\ \text{pb}}_{-0.047\ \text{pb}}(\text{theory}) \pm 0.044\ \text{pb}(\text{PDF} + \alpha_S)$$

2.3.3. Higgs boson decay

A Higgs boson mass of about 125 GeV allows the Higgs boson couplings to many SM particles to be explored. Fig. 2.4 shows the branching ratios for the Higgs boson main decays [29]. The dominant decay modes are $H \to b\bar{b}$ and $H \to WW^*$, followed by $H \to gg$, $H \to \tau^+\tau^-$, $H \to c\bar{c}$ and $H \to ZZ^*$. With much smaller rates, the following Higgs boson decay as a function of the branching ratio are $H \to \gamma\gamma$, $H \to Z\gamma$ and $H \to \mu^+\mu^-$.

Bosonic decays

For a given m_H, the sensitivity of a search channel depends on the production cross section of the Higgs boson, its decay branching fraction, reconstructed mass resolution, selection efficiency and the level of background in the final state. The most important contributions to the search for the Higgs boson in the bosonic final states are the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ channels in which all the final state particles can be very precisely measured and the reconstructed m_H resolution is comparatively excellent (typically 1-2 %). The $H \to WW^* \to 2l2\nu$ channel has relatively large branching fraction, but the m_H resolution is poor (approximately 20 %) due to the presence of neutrinos in the final state.

2.3.4. Fermionic decays

At lowest perturbative order, the partial decay width of the Higgs boson in a pair of fermions is given by :

$$\Gamma(H \to f\bar{f}) = \frac{CG_F m_f^2 M_H}{4\pi\sqrt{2}}\left(1 - \frac{4m_f^2}{m_H^2}\right)^{\frac{3}{2}} \tag{2.45}$$

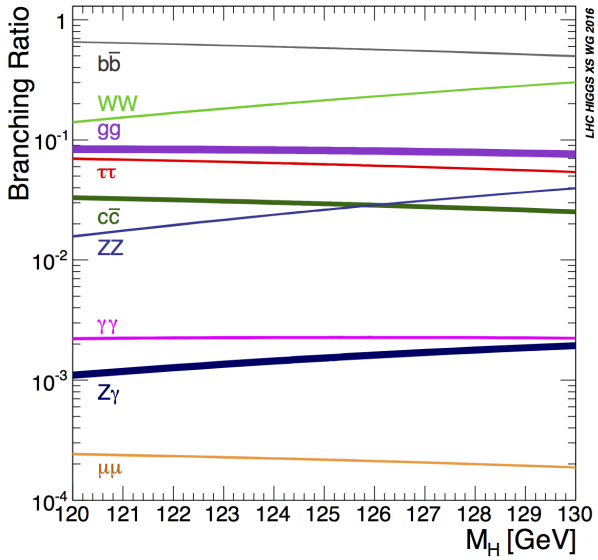


Figure 2.4.: The branching ratios for the main decays of the SM Higgs boson as a function of the mass of the boson. The mass range is chosen to be near $m_H = 125$ GeV.

where C is the colour factor (where $C = 1$ for leptons and $C = 3$ for quarks), G_F is the Fermi coupling constant. In the case of hadronic decays, it is necessary to consider QCD corrections due to loops with an exchange of a gluon, or to gluon emission in the final state. In the limit in which $m_H >> 2m_f$, the decay width at NLO is given by:

$$\Gamma_{NLO}(H \to q\bar{q}) = \frac{3G_F}{4\sqrt{2\pi}} M_H m_q^2 \left[1 + \frac{4}{3}\frac{\alpha_s}{\pi}\left(\frac{9}{4} + \frac{3}{2} log \frac{m_q^2}{M_H^2}\right)\right] \tag{2.46}$$

At hadron colliders, the most promising channels for probing the coupling of the Higgs field to the quarks and leptons are $H \to b\bar{b}$ and $H \to \tau^+\tau^-$, respectively. For a Higgs boson with $m_H \simeq 125$ GeV, the branching fraction to $b\bar{b}$ is about 57 % and to $\tau^+\tau^-$ is about 6 %. Nevertheless, the presence of very large backgrounds makes the isolation of a Higgs boson signal in these channels very challenging.

2.4. Limits of the Standard Model

Despite the fact that the SM has shown to be a robust theory of the elementary particles, there are some points which show the limits of its theoretical structure and some hints that new physics may be required to describe consistently the known phenomena of particle physics.
Some issues come from experimental observations that are not predicted by the SM, some others rely on theoretical conceptions about the nature of a fundamental particles theory.

In this section some of these issues will be described together with also some possible solutions as extension of the SM.

2.4.1. Higgs/Hierarchy problem

As it was shown in Section 2.3, it is possible to generate fermion and boson masses through the introduction of the elementary Higgs field. For model consistency, the Higgs boson mass cannot be so different from W boson mass, and the experimental discovery of the Higgs boson at the LHC indicates that $m_H \simeq 125$ GeV. However, at tree-level the Higgs boson mass receives quadratically divergent radiative corrections from the interactions with the gauge bosons, fermions and self interaction, as shown in Fig. 2.5.

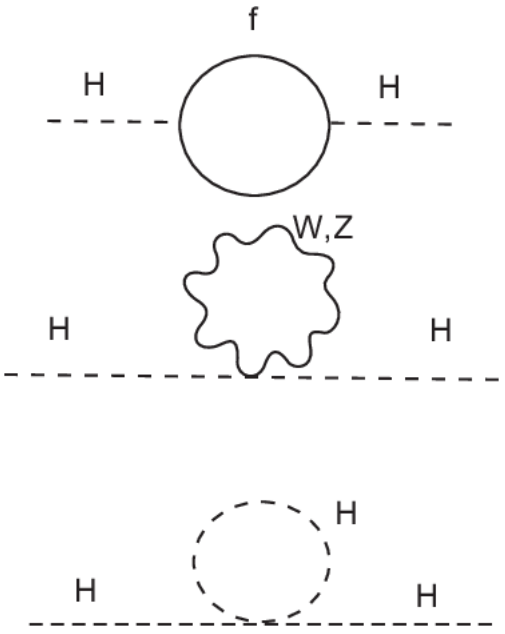


Figure 2.5.: Radiative corrections to the Higgs boson mass.

The Higgs boson mass can be written as:

$$m_H^2 = (m_H^2)_{\text{bare}} + O(\lambda, g^2, h^2)\Lambda^2 \tag{2.47}$$

where Λ is the next higher scale of the theory, after electroweak scale. If there was no higher scale, Λ could be interpreted as an ultraviolet cut-off, assuming that m_H is a measured parameter and that $(m_H)_{\text{bare}}$ is not a physics observable. However, if we consider Λ at the Planck scale ($M_P \simeq 10^{19}$ GeV) so that $\Lambda \simeq 10^{14}$ GeV, there should be a fine tuned cancellation of several orders of magnitude between the bare value of the mass and its correction. This seems highly 'not natural' and it represents a strong argument for the presence of new physics at higher energy scales. Models like

Technicolour, composite Higgs models or Supersymmetry, which will be not described in this thesis, try to address this problem.

2.4.2. Neutrino masses

In the SM, neutrinos are described as massless left-handed particles, with no right-handed component, assuming that right-handed neutrinos do not interact in the SM so they can be omitted. Experimental observations of the flavour oscillation indicate that neutrinos have a small mass. The approximation that $m_\nu << m_f$, where f is any charged fermion in the SM, is good for most applications, but mass terms for neutrinos should be consistently introduced in the SM Lagrangian. This can be done in two ways: introducing a Dirac mass (as it is done for the other fermions) or a Majorana mass.
In the first case, the existence of a right-handed neutrino is assumed, so it is possible to build a four-dimensional Dirac field using both left and right-handed neutrinos; the right-handed neutrino is not yet observed because it does not interact with SM particles, so it is called a *sterile* neutrino. In this scenario, the masses of the neutrinos are generated throughout the Higgs mechanism and it implies that the Yukawa coupling constants should be at least 12 orders of magnitude weaker than the top quark one.
In the case of the Majorana model, neutrinos are considered as their own antiparticles and right-handed neutrinos are still introduced as a charge conjugate of the left-handed neutrinos. In this case, it is possible to introduce neutrino mass term in the Lagrangian without exploiting the Higgs mechanism. However, since Majorana masses violate lepton number conservation, it should be assumed that the introduced Majorana mass is large enough that lepton number violation effects are compatible with observations.

2.4.3. Matter/anti-matter asymmetry

In our Universe, there is a large predominance of matter over anti-matter; from a cosmological point of view it is difficult to describe the evolution of the Universe from a balanced situation, predicted at the Big Bang, to the very asymmetric condition which we observe today. The SM provides a source of charge conjugate-parity (CP) violation that can take into account part of this asymmetry: this is coming from the presence of a complex phase in the CKM matrix which describes the mixing of different quark-flavour within the electroweak interactions. However, the CP violation provided by the CKM matrix of the SM is not enough to justify the observed matter anti-matter asymmetry, therefore there should be other mechanisms which can explain it.

2.4.4. Dark matter and dark energy

The SM can describe the behaviour of ordinary matter, but from cosmological measurements and gravitational effects it was observed that ordinary matter corresponds roughly to 4 % of the content of the Universe. The remaining part is hypothesised to be composed by dark matter (22 %) which has no electromagnetic interaction and can be detected only through gravitational force, and dark energy (74 %) which is associated to

the vacuum space. The dark energy is distributed throughout the Universe and it leads to a repulsive force which tends to accelerate the expansion of the Universe.
The SM cannot offer any explanation for this different kind of matter, and despite some models such as Supersymmetry having dark matter candidates, none of these has been observed in the experiments.

CHAPTER 3

The Large Hadron Collider and the ATLAS experiment

In this chapter, firstly the LHC and the ATLAS experiment will be described. The final part of the chapter is dedicated to the data quality monitoring, a crucial part of the data-taking process in which I was involved for the Pixel detector during Run 2 data taking period.

3.1. The Large Hadron Collider

The LHC is currently the largest and most powerful particle accelerator [30]. It first started in September 2008, and it is the latest addition to CERN's accelerator complex, situated near Geneva, Switzerland. The LHC is a two-ring superconducting hadron accelerator and collider, designed for protons and lead ions, which has been installed in the existing 26.7 km tunnel constructed for the LEP machine. The tunnel geometry includes eight crossing points followed by long straight sections for radio frequency cavities. Only four of the eight crossing points are used for collisions, and host the four main LHC experiments: ATLAS, CMS, LHCb and ALICE.
The machine is designed to collide proton beams with a center-of-mass energy of $\sqrt{s} = 14$ TeV and a luminosity of 10^{34} cm^{-2} s^{-1}. The nominal number of protons per bunch is 15×10^{11}, and the time between successive bunch collisions is designed to be 25 ns.

3.1.1. Accelerator structure

To accelerate counter-rotating proton beams, opposite magnetic dipole fields are required in both rings. The LHC has been designed as a collider with separate magnetic fields and vacuum chambers in the main arcs with common section only at the intersection regions, where the experiments are situated. The LHC relies on superconducting magnets which are cooled to a temperature of 1.9 K using superfluid helium, and operates at fields of

8 T. Superconducting technology is essential in order to obtain such magnetic fields and to avoid big resistive losses. Space limitations in the tunnel led to a two-in-one or twin-bore design for almost all of the LHC superconducting magnets, which accommodates the two beam channels in a common cryostat. The proton bunches accelerated by the LHC are supplied by an injector chain which includes several machines from the CERN accelerator complex, as shown in Fig. 3.1. The protons are first accelerated by the linear accelerator *Linac2* up to 50 MeV, then they reach the energy of 1.4 GeV in the *Proton Synchrotron Booster* (*PSB*). The next part of the chain is the *Proton Synchrotron* (*PS*) which brings the protons to an energy of 25 GeV, and finally the *Super Proton Synchrotron* (*SPS*) to achieve the 450 GeV necessary for the beam injection to the LHC ring.

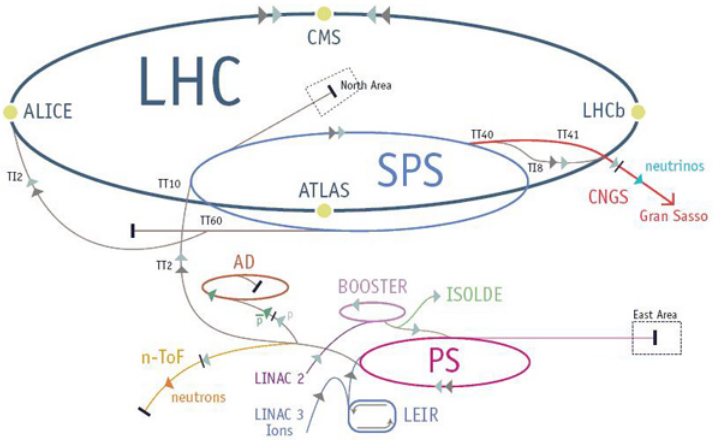


Figure 3.1.: CERN accelerator complex, including the LHC injection chain: Linac2, PSB, PS, SPS (copyright CERN).

Main features of the hadron colliders

In the following section, the main interesting quantities of circular colliders, such as LHC, are described.
Luminosity: in a particle-particle collider the processes under study usually involve a two-body initial-state which leads to an unknown final state: $|1,2>\rightarrow|f>$. The quantity that characterizes such a process is the cross section $\sigma_{|1,2>\rightarrow|f>}$, which is related to the probability that the process happens between the initial states $|1>$ and $|2>$, leading to the final state $|f>$. To write this quantity in terms of physical observables, in a particle accelerator the average number of interactions per bunch crossing is:

$$< \Delta N^{crossing}_{|1,2>\rightarrow|f>} >= \sigma_{|1,2>\rightarrow|f>} \frac{N_1 N_2}{\Sigma} \tag{3.1}$$

where Σ is the transverse section of the bunches, and N_1 and N_2 are the number of particles in each bunch. Thus the rate of interaction with a bunch crossing frequency f, is given by:

$$\frac{dN_{|1,2>\to|f>}}{dt} = \sigma_{1,2>\to|f>}\mathcal{L} \tag{3.2}$$

where $\mathcal{L}$ is the instantaneous luminosity:

$$\mathcal{L} = \frac{fN_1N_2}{\Sigma} \tag{3.3}$$

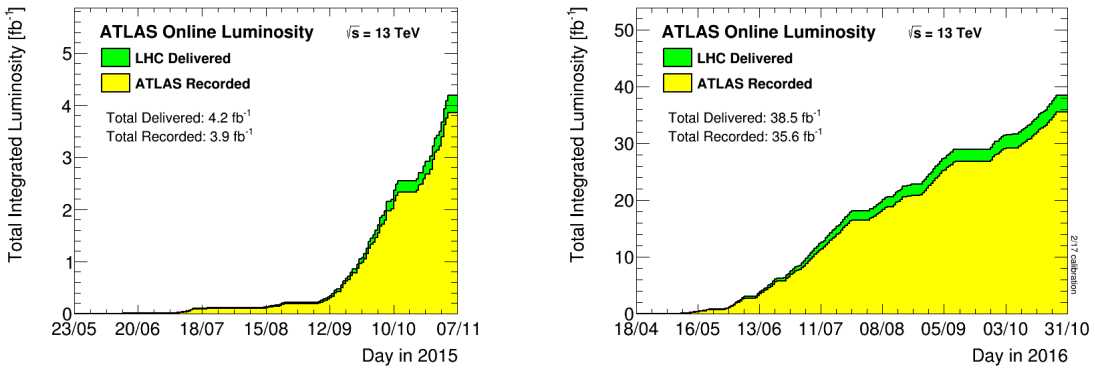
which is measured in units of $\text{cm}^{-2}\text{s}^{-1}$. Taking into account the possibility of different transversal dimensions for the bunches, a more precise definition of the instantaneous luminosity is given by:

$$\mathcal{L} = \frac{fN_1N_2}{4\pi\sigma_x\sigma_y} \tag{3.4}$$

where σ_x and σ_y are the beam transversal dimensions, supposed to have a Gaussian distribution.
The number of total collisions in a given period of time is given by the integral over time of the interaction rate, i.e it is given by the product of the cross section and the integrated luminosity L:

$$N_{events} = \sigma_{|1,2>\to|f>}L \tag{3.5}$$

In Fig. 3.2, the integrated luminosities recorded by the ATLAS experiment during the 2015 and the 2016 data taking periods are shown.



(a) (b)

Figure 3.2.: Integrated luminosity recorded by ATLAS in the 2015 (a) and the 2016 (b) data taking periods [31].

Pile-up: a parameter related to the instantaneous luminosity is the mean number of inelastic interactions per bunch crossing, called *pile-up* events. Pile-up events are mainly soft interactions which become background to the hard interaction targeted by the analysis. To relate the pile-up with the instantaneous luminosity, it can be shown that:

$$\mathcal{L} = \frac{\text{rate}_{\text{inelastic}}}{\sigma_{\text{inelastic}}} = \frac{\mu n_b f}{\sigma_{\text{inelastic}}} \tag{3.6}$$

where μ is the number of inelastic interactions per bunch crossing, n_b is the number of colliding bunches and f is the bunch crossing frequency. The number of pile-up interactions per bunch crossing μ, which is shown in Fig. 3.3 for the 2015 and 2016 data taking period, is proportional to $\mathcal{L}/f$ and increases with the peak luminosity.

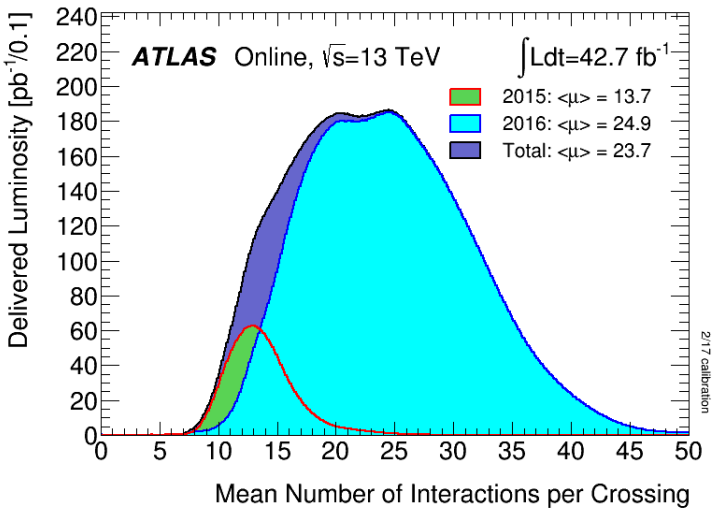


Figure 3.3.: Mean number of pile-up interactions per bunch crossing for 2015+2016 data taking [31].

3.2. The ATLAS detector

The ATLAS (*A Toroidal LHC ApparatuS*) detector is shown in Fig. 3.4. It is over 25 meters high, 44 meters long, and it weights approximately 7000 tones [32]. It is built around the LHC beam pipe and it is centered on one of the LHC collision points. ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (*IP*) in the centre of the detector and the z-axis along the beam direction. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r,ϕ) are used in the transverse (x,y) plane, ϕ being the azimuthal angle around the beam direction. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln(\tan(\theta/2))$. The distance ΔR in the $\eta - \phi$ space is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.
The detector is composed of a series of concentric sub-systems, each sensitive to different types of particles produced in the collisions.
The *Inner Detector* (ID) [33] is closest to the interaction point and it measures trajec-

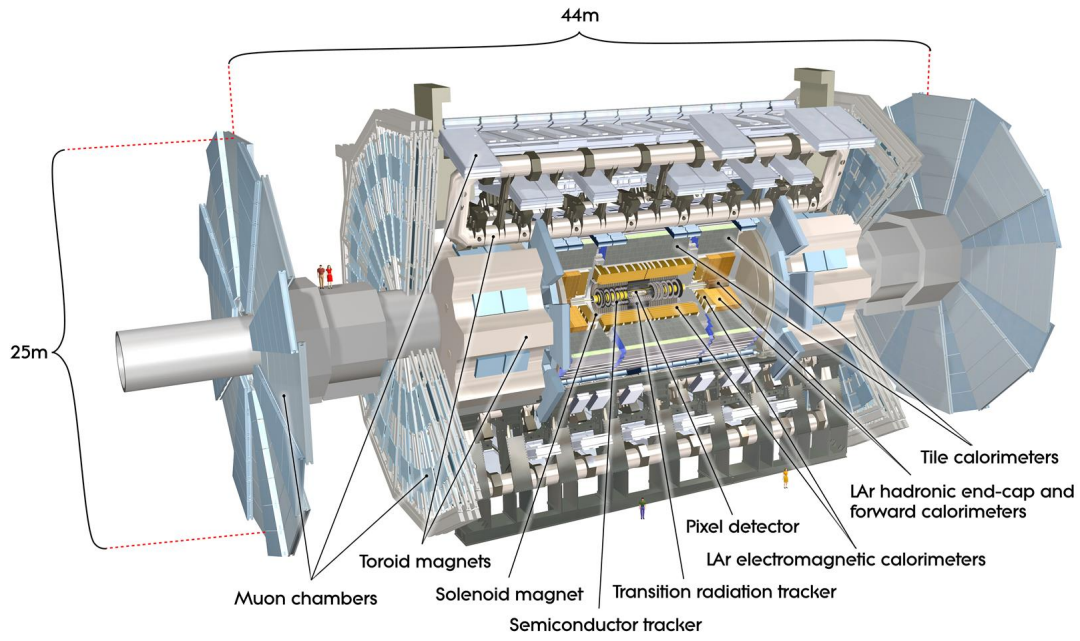


Figure 3.4.: The ATLAS detector.

tories of charged particles as they traverse the detector. To cope with the high particle densities produced by the LHC, the ID has been designed to make high-precision measurements with fine detector granularity. It operates in a 2 Tesla magnetic field provided by a solenoid magnet [34], which allows the ID to serve as a spectrometer in which the curved trajectories of charged particles can be reconstructed. This sub-system consists of central barrel layers, centered on the interaction point, and end-cap wheels or disks at either end of the barrel. It is composed of the Pixel Detector [35], the Semiconductor Tracker (SCT) [36] and the Transition radiation tracker (TRT) [37], as shown in Fig. 3.5. The pixel Detector and SCT sub-detectors provide uniform coverage in ϕ up to $|\eta| \leq 2.5$, while the TRT provides coverage up to $|\eta| = 2.0$. These detectors are built using two technologies: silicon sensors (Pixel Detector and SCT) and straw drift tubes (TRT). When charged particles cross the silicon sensors, they generate electron-hole pairs which can be collected with an applied electric field. This charge is recorded locally in the sensor, identifying the position of the particle. A similar process occurs in the straw drift tubes. Charged particles traversing the drift tubes ionize gas contained within the straw; the electrons produced by ionisation are drifted, using an applied electric field, to the wire at the center of the straw where they are recorded. In addition to being a tracking detector, the TRT also provides particle identification through the detection of transition radiation. Charged particles emit transition radiation photons when traversing the TRT; the probability of emitting a photon is a function of the Lorentz

factor γ. Thus at fixed momentum, electrons will emit more transition radiation photons than charged hadrons and this provides separation power used for particle identification.

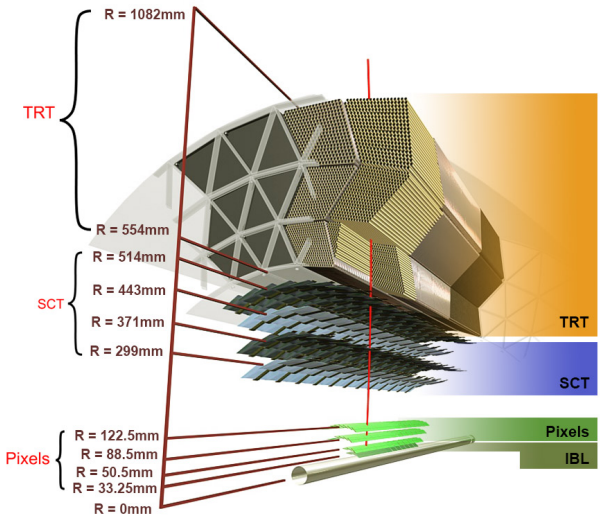


Figure 3.5.: The ATLAS Inner Detector in the barrel region.

Surrounding the ID is the calorimeter system [38]. This is composed of the liquid argon electromagnetic calorimeters, the tile calorimeters, the liquid argon hadronic end-cap calorimeters and the forward calorimeters, as shown in Fig. 3.6. The calorimeters are designed to measure the energy of electrons, photons and hadrons. The ATLAS calorimeters are built as *sampling* calorimeters, where incident particles produce showers of energy in the calorimeter. Only a fraction of the energy produced by the particle is measured by active detector sensors and afterwards the energy of the full shower can be derived from the observed energy.
The energies of electrons and photons are measured by the liquid-argon (LAr) electromagnetic (EM) barrel and end-cap calorimeters. These detectors provide high granularity measurements, which are very useful for particle identification in the range $|\eta| < 2.5$. The Tile calorimeters and the LAr hadronic end-cap calorimeter are designed to measure the energy of hadrons. The scintillator-tile calorimeter is separated into a barrel and two extended barrel cylinders. In the end-caps, $1.5 \leq \eta \leq 3.2$, LAr technology is used for the hadronic calorimeters. The LAr forward calorimeters provide both electromagnetic and hadronic energy measurements, and extend the pseudorapidity to $|\eta| = 4.9$.

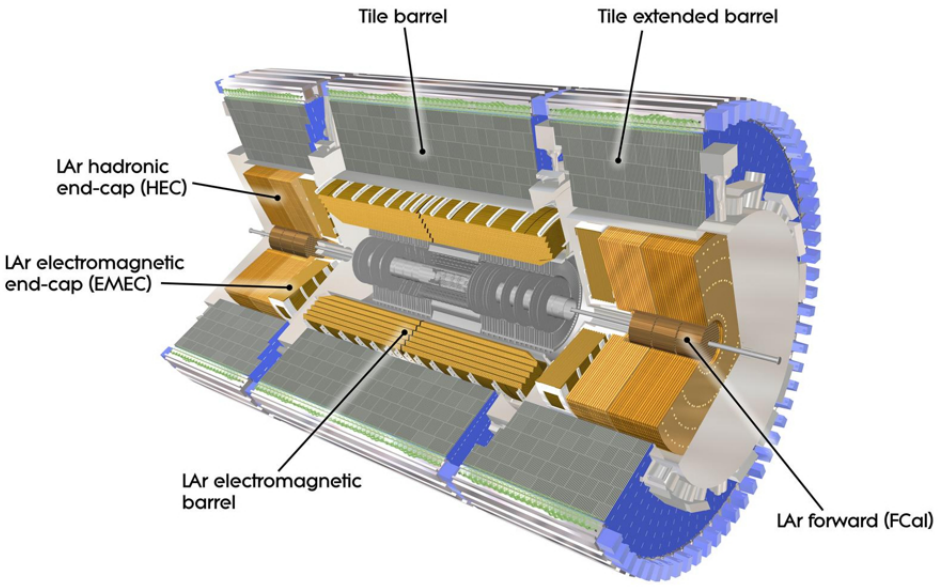


Figure 3.6.: The Calorimeter system.

The Muon Spectrometer (MS), shown in Fig. 3.7, surrounds the calorimeters [39]. It is designed to measure the trajectories of muons leaving the calorimeter. Over the range $|\eta| < 1.4$, magnetic bending is provided by the large barrel toroid. For $1.6 \leq |\eta| \leq 2.7$, muon tracks are bent by small end-cap magnets inserted in both ends of the barrel toroid, while in the region $1.4 \leq |\eta| \leq 1.6$ the bending is provided by a combination of the barrel and end-cap fields.
In the barrel region, the position of the muons are measured in chambers in three cylindrical layers around the beam axis. In the transition and end-cap regions, the chambers are arranged in three planes perpendicular to the beam. For most of the η-range, the muon positions are measured by Monitored Drift Tubes (MDTs), while in the range $2 \leq |\eta| \leq 2.7$, Cathode Strip Chambers (CSCs) are used. The MS includes chambers used also for the trigger in the region $|\eta| \leq 2.4$, where Resistive Plate Chambers (RPC) are used in the barrel while Thin Gap Chambers (tGC) are used in the end-cap regions.

3.2.1. DAQ and Data quality monitoring

The main challenge for the ATLAS experiment is the high event rate. To be able to store and analyse a large amount of data, the *trigger system* has been developed to perform an online event selection [40]. The non-interesting events are rejected in three stages, called Level 1 (L1), Level 2 (L2) triggers and Event Filter (EF). The L2 and EF are usually referred to as High Level Trigger (HLT). The design of the data acquisition

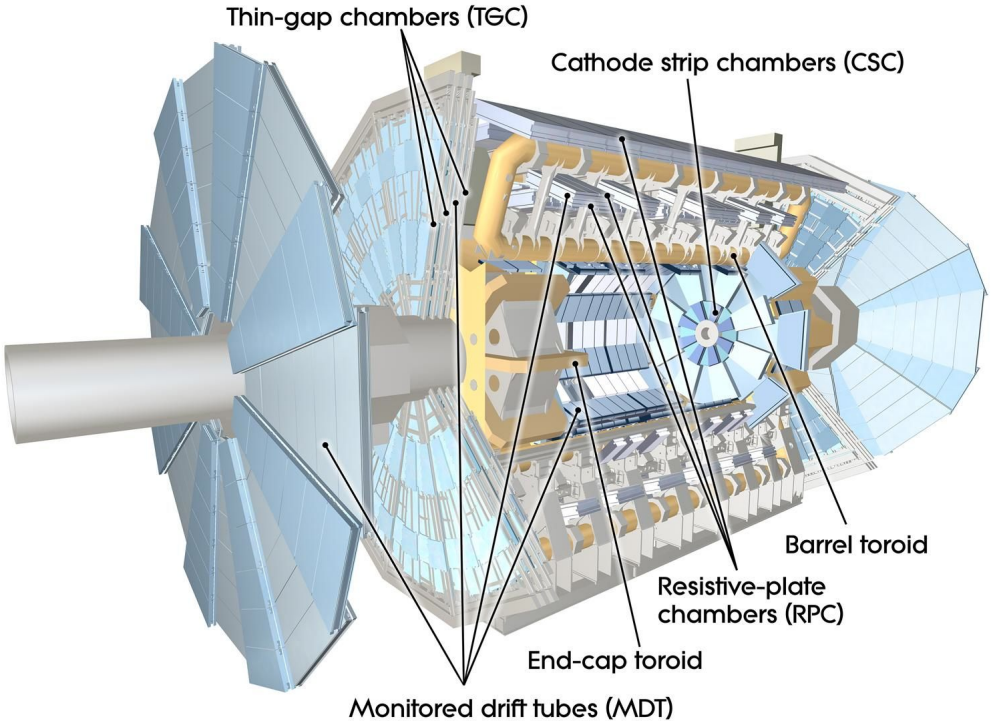


Figure 3.7.: The Muon Spectrometer.

system is such that it allows continuous monitoring of the data from when an event is triggered until it is stored for physics analysis.
The Level 1 (L1) trigger is hardware-based. It is installed on the calorimeters and the MS to reduce material occurrence from cables and read-out electronics. High transverse momentum (p_T) muons are triggered by the RPCs in the barrel and the TGCs in the end caps. High p_T electrons and photons, jets and hadronically decaying τ leptons and missing and total transverse energy are identified in the calorimeters (EM, hadronic and forward) using reduced-granularity information. Trigger decisions are based on simple combinations of objects which are required to coincide or to be vetoed. The detector signals are stored in the front-end pipelines waiting for the L1 trigger decision with a latency of less than 2.5 μs. The output event rate from L1 is 75 kHz for the designed luminosity and bunch spacing of 25 ns. Regions of Interest (RoIs) are defined at this point to be further investigated by the HLT.
The Level 2 (L2) and Event Filter (EF) triggers are both software based and they use sophisticated algorithms for event rejection. The L2 trigger looks into the regions of interest defined at L1 to verify whether a triggered object is valid. Afterwards, for all objects, except for the jets, extra information is retrieved from the Inner Detector. At this point, L2 RoIs are defined. In the end, the L2 trigger further reduces the rate to 3

kHz with average processing time of $\simeq$ 40 ms/event.
In the EF the full event is reconstructed, and offline algorithms are used to reject events. At this final stage, the event rate is reduced to $\simeq$ 200 Hz with an average processing time of $\simeq$ 4 seconds/event. An event is reconstructed at the Event Builder (EB) from the information stored in the readout buffers (ROBs) using the full granularity of the detector. After the EF decision, the event is stored for offline analysis. The data are written to streams depending on the type of trigger that was fired. There are also several calibration streams for specific studies; an important stream which contains $\simeq$ 10 % of the data is the *express* stream which is used for online calibration and Data Quality (DQ) checks.
Data quality monitoring is part of the Monitoring Infrastructure of the ATLAS experiment. It is performed throughout the data acquisition and then during the offline processing of fully reconstructed data. Due to the complexity of the ATLAS experiment, a framework for automatic data quality assessments of incoming data and a visualisation tool for easy identification of problems are both essential [41]. A highly scalable distributed data quality monitoring framework (DQMF) has been developed and is being used to monitor the quality of the data as well as operational conditions of hardware and software elements of the detector, trigger, and data acquisition systems. Online, the DQMF checks thousands of histograms permitting to avoid recording faulty data and alerting the shift crew in case of problems. The shift crew interacts with DQMF via the data quality monitoring display (DQMD), which serves as a platform to alert of problems and investigate them. Offline, the DQMF is used to perform the data quality monitoring on the express stream. Then, after the bulk reprocessing at Tier-0, a full scale validation is performed and the final data quality assessment is done. The information being monitored vary from operational condition, such as errors from readout electronics, to distributions of triggered objects, and are published into a central Online Histogramming Service (OHS). A workflow of the DQMF interaction with the other ATLAS services is shown in Fig. 3.8. Single data quality tests are defined by DQParameters. Each DQParameter specifies what input histogram(s) to use, what algorithm and parameters (DQAlgorithm) to apply and the threshold to classify the result (DQResult) as good or bad. All the DQParameters are grouped in different DQRegions, which in turn can be combined in more general DQRegions, thus forming a hierarchy known as the DQTree. The logic to combine the DQResults of the sub-parameters and sub-regions as well as the information specific to each DQParameter are defined in a configuration database. Within this structure, each sub-detector of the ATLAS experiment is described as a top DQRegion that hosts several tiers of the DQRegions and DQParameters to check the performance of its hardware and recorded data. For each top DQRegion, an application runs the data quality checks and outputs the colour-coded DQResults relaying the quality of the data to a central Information Service. A summary scheme of the DQMF structure is shown in Fig. 3.9.
During the timeframe of the PhD period, I was involved in the optimisation and maintenance of the DQMF for the Pixel Detector; different tasks were performed, from algorithm optimisation to make more robust data quality assessment to the selection of the

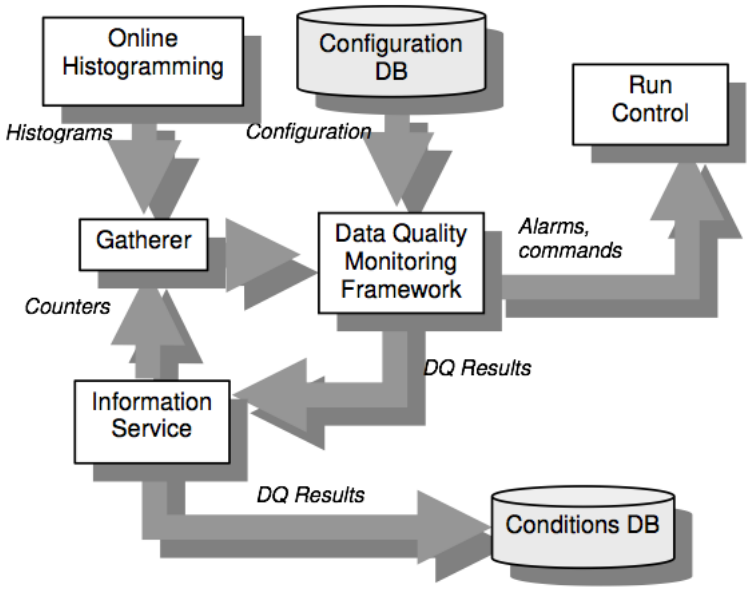


Figure 3.8.: The DQMF interaction with the online services.

relevant monitored information given the different LHC running condition during Run 2 data taking.

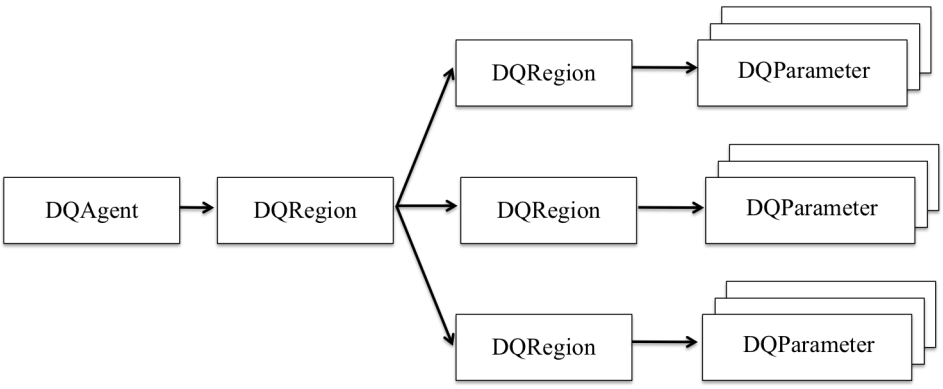


Figure 3.9.: Scheme of the DQMF structure.

CHAPTER 4

τ reconstruction and identification

In the following chapter, the algorithms used to reconstruct and identify hadronic τ decays will be described [42], [43]; at the end of the chapter, the τ efficiency identification measurement will be reported. Results are included in the latest set of scale factors provided by the ATLAS Tau Performance working group (TauWG) to correct the simulated events used in analyses selecting taus in the final state; this is the first attempt to perform such a type of measurement binned in τ p_T, so most of the described procedures to get the results have been developed specifically for this analysis.

4.1. Reconstruction of hadronic τ decays

With a mass of $m_\tau = 1.777$ GeV, the τ is the only lepton heavy enough to decay into hadrons (h), and it does so in about two third of the cases, typically into either one or three charged pions or a kaon and up to two neutral pions (π_0) and one neutrino (ν_τ). The branching fractions for the main τ decay modes are given in Table 4.1. The hadronic τ decay candidates ($\tau_{had\text{-}vis}$) are built from jets reconstructed using the anti-k_t algorithm with a radius parameter value of 0.4 [44],[45]. Three-dimensional clusters of calorimeter cells calibrated using a local hadron calibration serve as inputs to the jet algorithm [46], [47]. Only inputs with $p_T > 10$ GeV and $|\eta| < 2.5$ are considered. The calculation of the $\tau_{had\text{-}vis}$ four-momentum uses clusters within the *core-region* ($\Delta R < 0.2$ from the initial jet-axis). It includes a τ-specific calibration derived from simulated samples, which accounts for out-of-cone energy, underlying event, hadron composition of τ decays and contribution from pileup; the calibration is later confirmed and corrected with an insitu measurement performed using $Z \to \tau\tau$ events. Tracks reconstructed in the inner detector are matched to the $\tau_{had\text{-}vis}$ candidate if they are in the core region and satisfy the following criteria:

- $p_T > 1$ GeV;

Decay Mode	$\mathcal{B}$ [%]
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$	17.8
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$	17.4
$\tau^- \to h^- \nu_\tau$	11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$	4.8
Others	3.2
All had. modes	64.8

Table 4.1.: Branching fractions of the main τ decay modes.

- at least 2 associated hits in the pixel layers of the inner detector;
- at least 7 hits in total in the pixel and silicon microstrip layers;
- requirements are imposed on the distance of closest approach of the tracks to the τ vertex in the transverse plane, $|d_0| < 1.0$ mm, and longitudinally, $|z_0 \times \sin\theta| < 1.5$ mm.

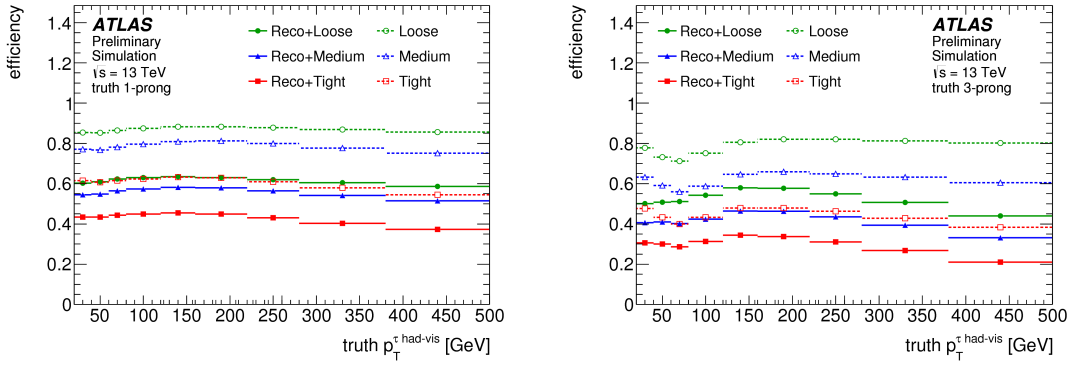
The $\tau_{had\text{-}vis}$ charge is then reconstructed from the sum of the charges of core tracks.

4.2. Identification

The $\tau_{had\text{-}vis}$ identification (ID) algorithm is designed to reject backgrounds from quark- and gluon-initiated jets. The identification uses *Boosted Decision Tree* (BDT) based methods. The BDT for $\tau_{had\text{-}vis}$ candidates with one or three associated tracks are trained separately using simulated $Z/\gamma^* \to \tau\tau$ events for signal and dijet events (selected from data) for background. Three working points labelled *Loose*, *Medium* and *Tight* are provided, corresponding to different τ identification efficiency values, with the efficiency designed to be independent of p_T. The target efficiencies are 0.85, 0.75 and 0.60 for the generated 1-track loose, medium and tight working points, and 0.75, 0.60 and 0.45 for the corresponding 3-track target efficiencies. The identification efficiencies are optimized to be flat versus the τ p_T and the pileup, as shown in Fig. 4.1-4.2. Thus the jet rejection depends on the τ p_T and varies from 15-20 (100-400) for loose identification criteria for the 1 (3) track τ candidates with 20 GeV $< p_T <$ 100 GeV.

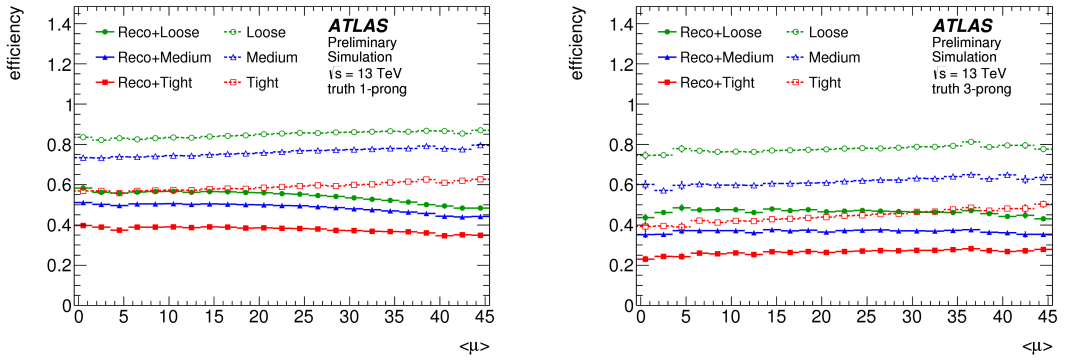
4.3. $\tau_{had\text{-}vis}$ identification efficiency measurement

In this paragraph, the main steps of the analysis for the $\tau_{had\text{-}vis}$ ID efficiency measurement are described. Results are provided as scale factors which need to be applied to Monte Carlo samples to get a better description of the data.



(a) (b)

Figure 4.1.: Efficiency for τ identification (open symbols) and combined reconstruction and identification efficiency (full symbols) as a function of the τ p_T, for 1-track (a) and 3-track (b) τ candidates.



(a) (b)

Figure 4.2.: Efficiency for τ identification (open symbols) and combined reconstruction and identification efficiency (full symbols) as a function of the average number of interactions per event, for 1-track (a) and 3-track (b) τ candidates.

4.3.1. Analysis selection and control region definitions

The analysis is following a tag and probe approach consisting of selecting events triggered by the presence of a muon (tag), as a sign of leptonically decaying τ ($\tau \to \mu\nu_\tau\nu_\mu$), and containing a hadronically decaying τ lepton candidate (probe) in the final state. In this section, *signal* refers to a $\tau_{had\text{-}vis}$ candidate geometrically matched with a generated $\tau_{had\text{-}vis}$ or the $Z \to \tau_\mu\tau_{had}$ event containing such τ candidates. To select $Z \to \tau_\mu\tau_{had}$ events, a single-muon trigger with a threshold of 26 GeV is used, so the corresponding offline reconstructed muon candidate must have $p_T > 27.3$ GeV and be geometrically matched to the online muon. Moreover, the muon is required to pass an isolation requirement and a multi-jet control region (QCD CR) is built inverting this requirement. Events are required to have no reconstructed electrons and at least one $\tau_{had\text{-}vis}$ candidate with 1 or 3 tracks (referred as 1 and 3 prong). If there are multiple $\tau_{had\text{-}vis}$ candidates, only the leading p_T candidate is considered. In addition, a very loose requirement on the $\tau_{had\text{-}vis}$ identification BDT score > 0.3 is made to ensure that the quark/gluon jet composition between signal region and QCD CR are comparable. The muon and $\tau_{had\text{-}vis}$ candidates are required to have opposite-sign electric charges (OS); additional cuts are placed in order to enhance the purity of the signal regions: muon p_T is required to be less than 40 GeV and the visible mass of the $\tau_{had\text{-}vis}$-muon pair ($M_{(\tau,\mu)}$) is required to be in the range 45 GeV $\leq M_{(\tau,\mu)} \leq$ 120 GeV.
In order to reduce Top background contributions, events with b-tagged jets are vetoed and a corresponding Top control region (Top CR) is built by inverting this requirement. A series of selection requirements is used to suppress W+jets events:

- the transverse mass of the muon and the missing energy (E_T^{miss}) system, $m_T = \sqrt{2p_T^\mu E_T^{miss}(1 - \cos\Delta\phi(\mu, E_T^{miss}))}$, is required to be less than 50 GeV, where p_T^μ is the transverse momentum of the muon and $\Delta\phi(\mu, E_T^{miss})$ is the $\Delta\phi$ separation between the muon and the missing transverse momentum;

- the quantity $\Sigma\cos\Delta\phi = \cos\Delta\phi(\mu, E_T^{miss}) + \cos\Delta\phi(\tau_{had\text{-}vis}, E_T^{miss})$ is required to be greater than -0.15, where $\Delta\phi(\tau_{had\text{-}vis}, E_T^{miss})$ is the $\Delta\phi$ separation between the $\tau_{had\text{-}vis}$ and the missing transverse momentum.

A W+jets control region (W CR) is built inverting the requirements mentioned above, $m_T > 60$ GeV and $\Sigma\cos\Delta\phi < 0$, and purified in W+jets events requiring $E_T^{miss} > 30$ GeV. The selection criteria are summarised in Table 4.2.
After the main analysis categories have been defined, the signal/control regions are then split into 4 slices according to the τ p_T interval: 20 GeV $< p_T <$ 25 GeV, 25 GeV $< p_T <$ 30 GeV , 30 GeV $< p_T <$ 40 GeV, $p_T >$ 40 GeV. Additionally, each slice is then divided into 4 slices according the τ working points defined above: *not loose*, *loose not medium*, *medium not tight* and *tight*. Considering both τ p_T and τ ID splits, the global signal region is split into 16 categories both for 1 and 3 prong taus; background estimation, which is described in the Section 4.3.2, is performed for each category separately since background composition can be dependent both on τ p_T and τ ID.

variable	SR region	W+jets CR	QCD CR	Top CR
n bjets	0	0	0	> 0
m_T	< 50 GeV	> 60 GeV	< 50 GeV	> 40 GeV
E_T^{miss}	-	> 30 GeV	-	-
$\Sigma \cos \Delta\phi$	> -0.15	< 0	> -0.15	> -0.15
μ p_T	< 40 GeV	-	< 40 GeV	-
μ^{iso}	yes	yes	no	yes
$m(\tau_{vis}, \mu)$	[45,120] GeV	[45,120] GeV	[45,120] GeV	-

Table 4.2.: Signal/Control regions selection criteria.

4.3.2. OS-SS background estimation method

The dominant background contributions come from the misidentification of jets as $\tau_{had\text{-}vis}$ candidates in multi-jet and W+jets events which are estimated via data-driven methods using dedicated control regions. The multi-jet background is fully taken from a control region defined by inverting the opposite-sign requirement in the signal region; a normalisation factor (r_{QCD}) is used to correct the difference between the same charge and opposite charge $\mu - \tau_{had\text{-}vis}$ events. The same sign component of other backgrounds, estimated from the simulation, is subtracted to avoid double counting. The r_{QCD} normalisation factor is derived in QCD CR as the ratio of the opposite charge yield to the same charge yield. The overall data in the signal region are then described by:

$$N_{OS}^{data} = r_{QCD} \times N_{SS}^{data} + N_{MC}^{Z\to\tau\tau} + N_{MC}^{Z\to ll} + N_{MC}^{W+jets} + N_{MC}^{top} \tag{4.1}$$

with

$$N_{MC}^{X} = N_{MC,OS}^{X} - r_{QCD} \times N_{MC,SS}^{X} \tag{4.2}$$

where N_{SS}^{data} is the number of data events from the same sign control region, and $N_{MC,OS}^{X}$ ($N_{MC,SS}^{X}$) represents the number of simulated events for a given background, X, with the opposite (same) sign charge. To correct possible discrepancies between data and simulation in the W+jets and Top background, a normalisation factor $k_{OS}^{W(Top)}$ ($k_{SS}^{W(Top)}$), is defined as the ratio of data to simulated W+jets (Top) events in the W+jets (Top) CR with opposite (same) sign charge requirement. The number of W+jets (Top) events is then given by:

$$N_{MC}^{W+jets(Top)} = k_{OS}^{W(Top)} \times N_{MC,OS}^{W+jets(Top)} - r_{QCD} \times k_{SS}^{W(Top)} \times N_{MC,SS}^{W+jets(Top)} \tag{4.3}$$

The values of r_{QCD} and $k_{OS}^{W(Top)}$ ($k_{SS}^{W(Top)}$) for the different analysis regions are shown in Tables A.1 and A.2 for 1/3 prong taus, respectively. Modelling distributions for Top/QCD and W CRs in some τ p_T/τ ID slices are shown in Fig. 4.4-4.5, respectively. Overall, good modelling has been observed in the various CRs. Given the low number

of Monte Carlo (MC) events for the W+jets sample which could lead to spiky distributions and unphysical biases, as it can be observed in Fig. 4.5, the W+jets background contribution in the signal region is estimated using:

- the normalisation from W (OS-SS) simulated events, where SS are the events in which the muon and $\tau_{had\text{-}vis}$ candidates are required to have same-sign electric charges;
- the shape of W (OS-SS) contribution for the considered variable from W CR subtracting non W+jets background contributions from data.

The shape extraction could lead to a bias given by kinematic differences between signal region and W CR; systematic uncertainties have been extracted to account for this effect (Section 4.3.3).

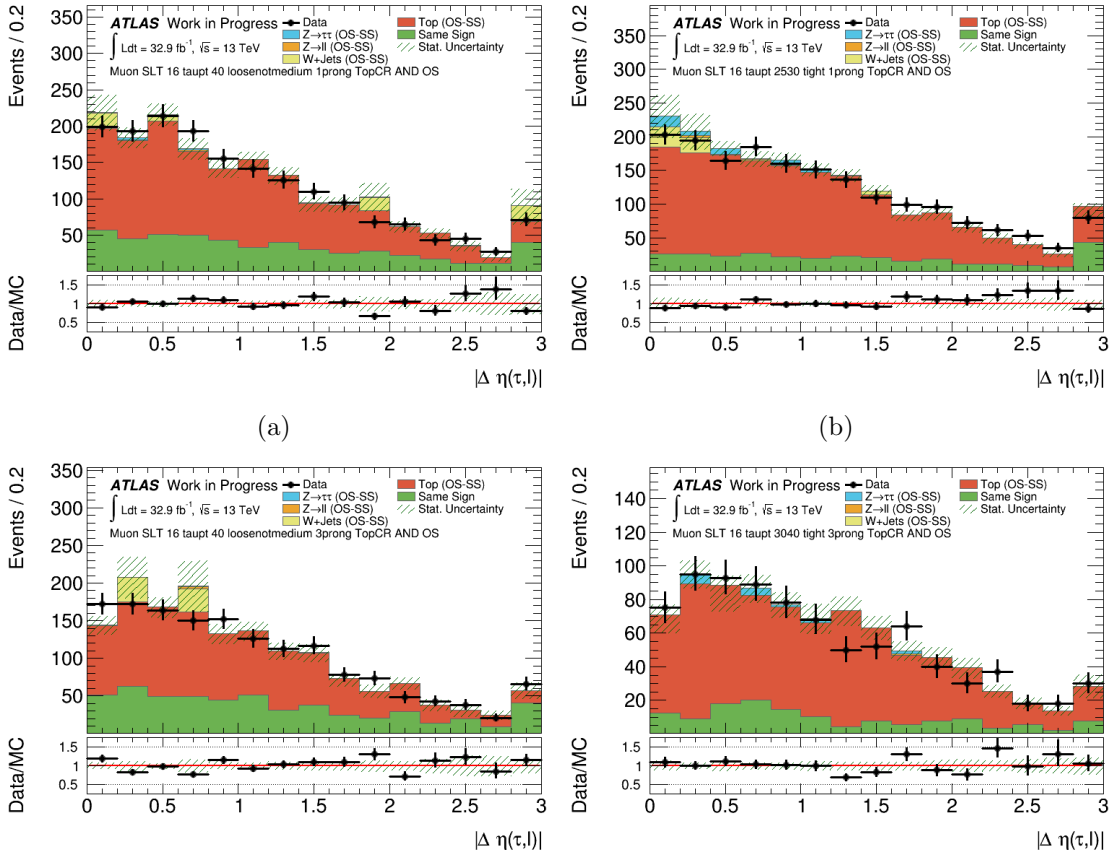


Figure 4.3.: Examples of modelling in Top CR for: *a)* $p_T > 40$ GeV, 1 prong, loose not medium ID category, *b)* $25 < p_T < 30$ GeV, 1 prong, tight ID category, *c)* $p_T > 40$ GeV, 3 prong, loose not medium ID category, *d)* $30 < p_T < 40$ GeV, 3 prong, tight ID category.

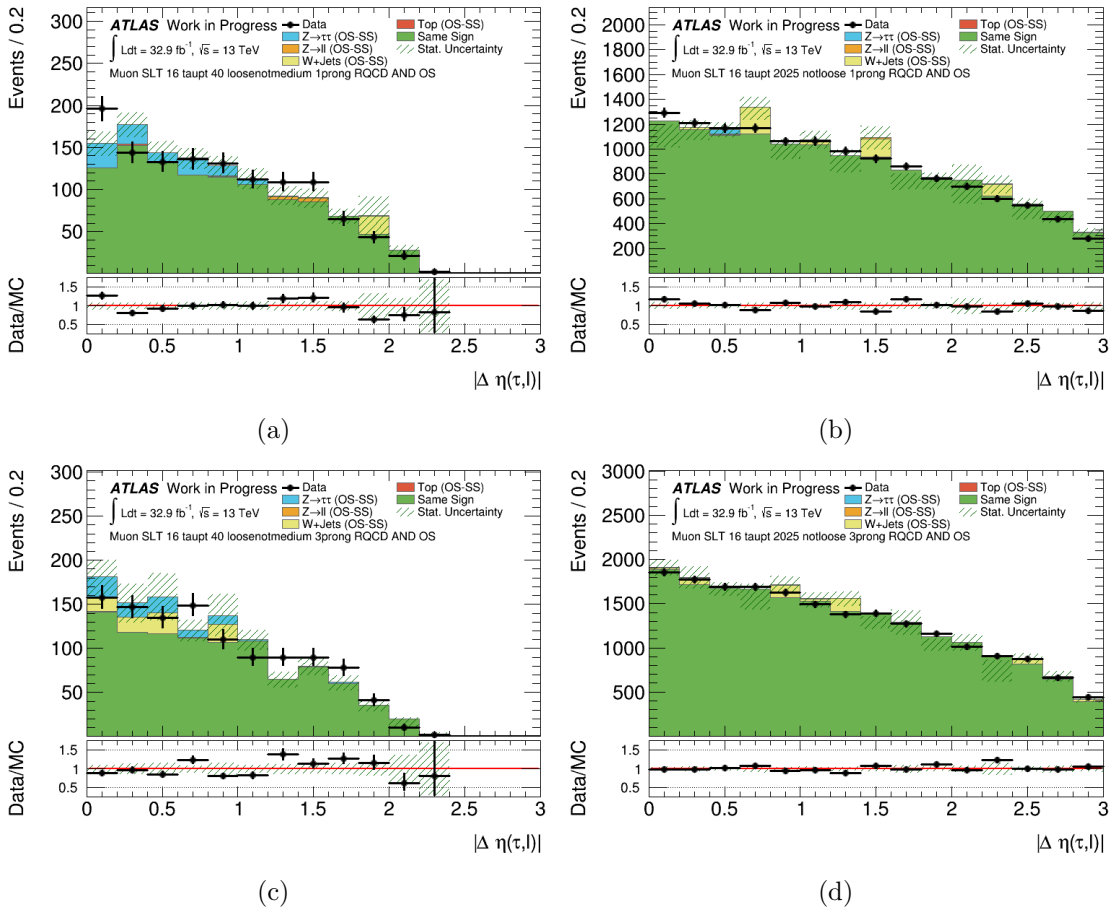


Figure 4.4.: Examples of modelling in QCD CR for: *a)* $p_T > 40$ GeV, 1 prong, loose not medium category ID, *b)* $20 < p_T < 25$ GeV, 1 prong, not loose category ID, *c)* $p_T > 40$ GeV, 3 prong, loose not medium ID category, *d)* $20 < p_T < 25$ GeV, 3 prong, not loose ID category. As it can be observed in the figures, the W+jets background estimation through MC suffers of large statistical uncertainties, justifying its estimation through data-driven methods as reported in the text.

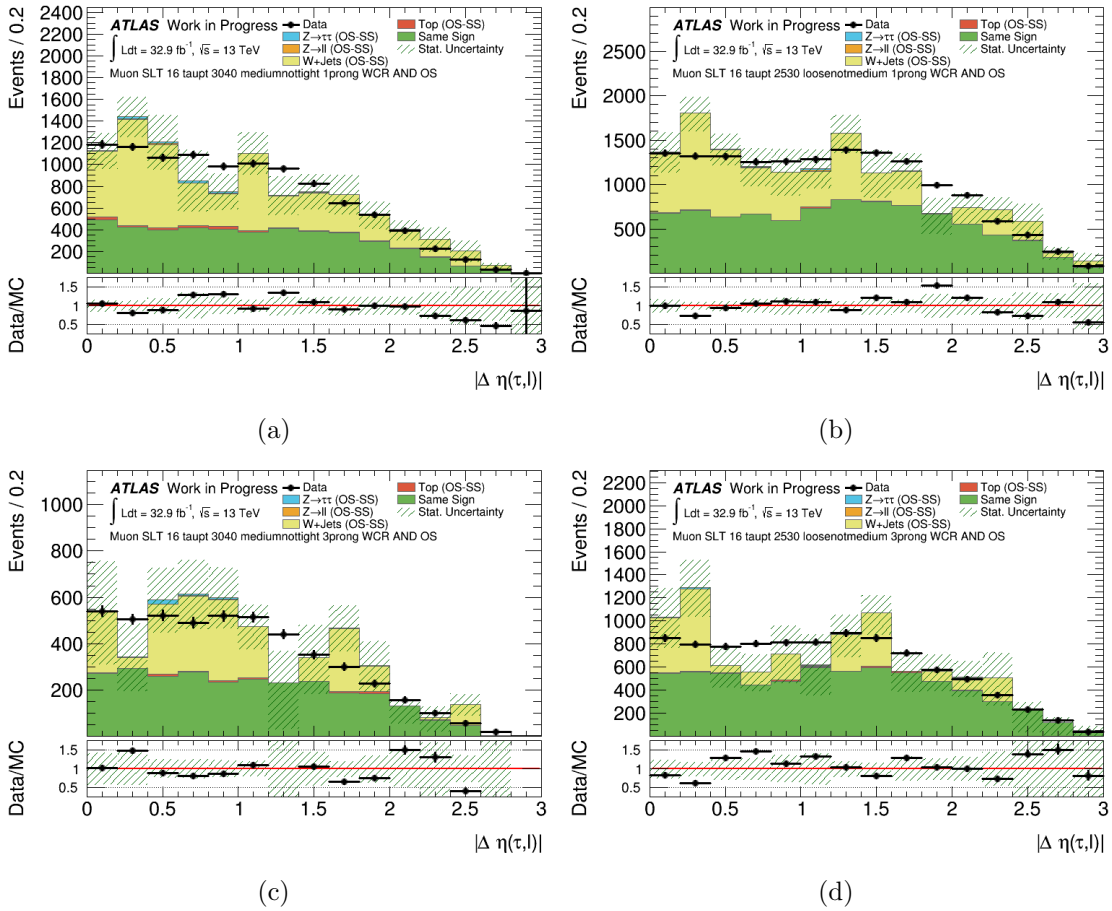


Figure 4.5.: Examples of modelling in W CR for: *a)* 30 GeV $< p_T <$ 40 GeV, 1 prong, medium not tight ID category, *b)* 25 $< p_T <$ 30 GeV, 1 prong, loose not medium category, *c)* 30 GeV $< p_T <$ 40 GeV, 3 prong, medium not tight ID category, *d)* 25 $< p_T <$ 30 GeV, 3 prong, loose not medium ID category.

4.3.3. QCD(W)-Shape and $Z \to \tau\tau$ MC generator systematic uncertainty estimation

Systematic uncertainties related to the multi-jet and W background estimation as well as comparison of different signal MC generators have been derived:

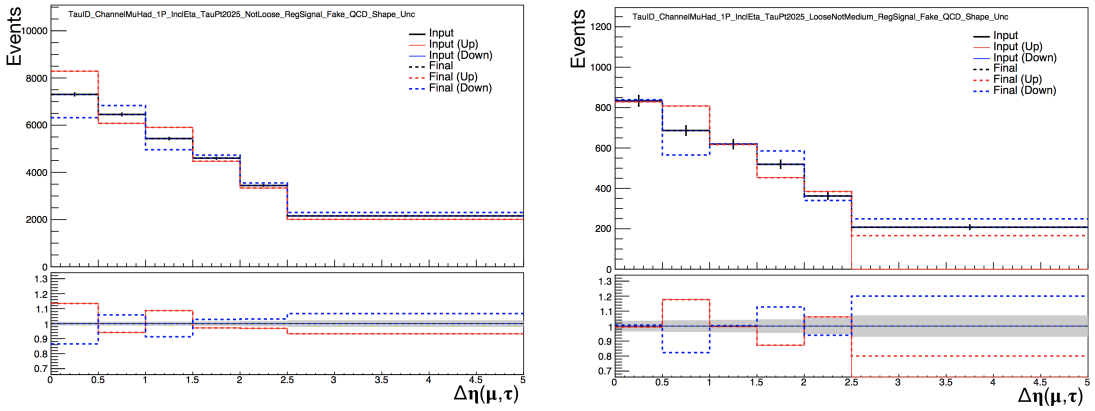
- systematic uncertainty on QCD (called also SS or Fake) background shape has been evaluated comparing (Data - MC background) in OS and SS QCR CRs. Differences have been assigned as systematic uncertainty. Envelopes for this uncertainty for some analysis categories are shown in Fig. 4.6;

- systematic uncertainty on W+jets shape extraction from W CR to signal region have been evaluated comparing W (OS-SS) MC shapes in the two regions. Differ-

ences have been assigned as systematic uncertainty. Envelopes for this uncertainty for some analysis categories are shown in Fig. 4.7;

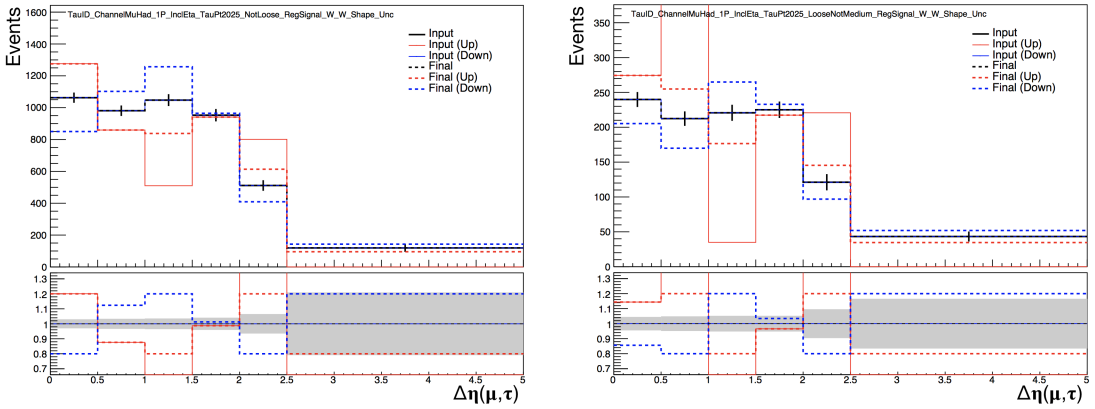
- $Z \to \tau\tau$ MC generator systematic uncertainty has been evaluated comparing shape distributions for $Z \to \tau\tau$ in each analysis category using *Sherpa* and *Powheg* MC generators. Differences have been assigned as systematic uncertainty. Envelopes for this uncertainty for some analysis categories are shown in Fig. 4.8.

These uncertainties will be later used in the fit to extract the final results.



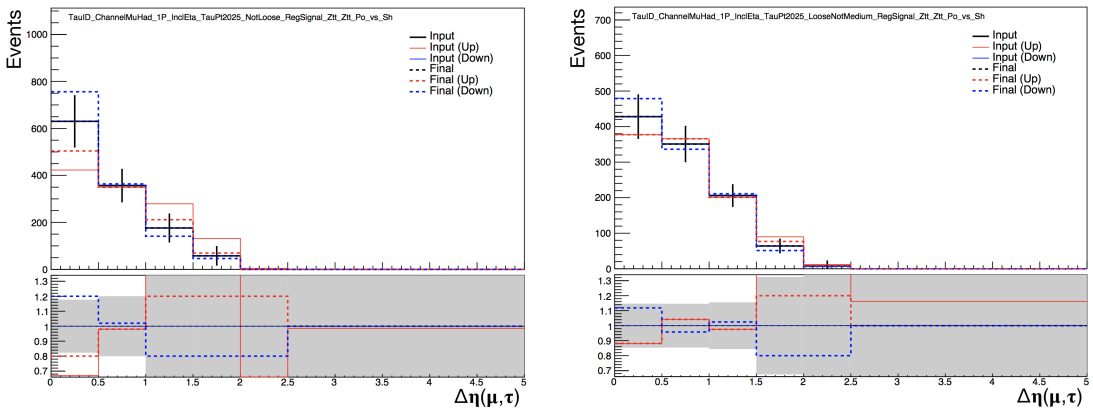
(a) (b)

Figure 4.6.: Envelope plots for $\Delta\eta(\mu,\tau)$ QCD Shape uncertainty for 20 GeV $<$ p_T $<$ 25 GeV, 1 prong : *a)* not loose, *b)* loose not medium ID categories.



(a) (b)

Figure 4.7.: Envelope plots for $\Delta\eta(\mu,\tau)$ W Shape uncertainty for 20 GeV $<$ p_T $<$ 25 GeV, 1 prong : *a)* not loose, *b)* loose not medium ID categories.



(a) (b)

Figure 4.8.: Envelope plots for $\Delta\eta(\mu,\tau)$ $Z \to \tau\tau$ MC generator uncertainty for 20 GeV $< p_T <$ 25 GeV, 1 prong : *a)* not loose, *b)* loose not medium ID categories.

4.3.4. Scale Factor extraction

To extract the scale factors, a fit is performed using the *Maximum Likelihood Estimation* technique to get the $Z \to \tau\tau$ normalisation factor in each analysis category. The variable chosen to perform the fit is the $\Delta\eta$ separation between the muon and the $\tau_{had\text{-}vis}$, $\Delta\eta(\mu,\tau)$; this has been chosen because of:

- good signal/background separation power;
- low correlation with τ-ID related variables;
- low kinematic bias for W+jets background estimation.

Prefit distributions of this variable in all analysis signal region categories are shown in Fig. 4.9-4.12. In general, good modelling has been observed in all the cases. The fit has also been performed using the τ track multiplicity variable, which is defined as the number of core ($\Delta R < 0.2$) and outer ($0.2 < \Delta R < 0.6$) tracks associated to the $\tau_{had\text{-}vis}$. The tracks are weighted by a distance to the τ candidate which is calculated in a similar way to the distance measure in an anti-kt jet algorithm. Results using τ track multiplicity are compatible with the results obtained using $\Delta\eta(\mu,\tau)$; however, this variable has less signal/background separation power with respect to $\Delta\eta(\mu,\tau)$ which leads to larger fit uncertainties. For this reason, results using track multiplicity have only been used as cross check of the fit procedure.

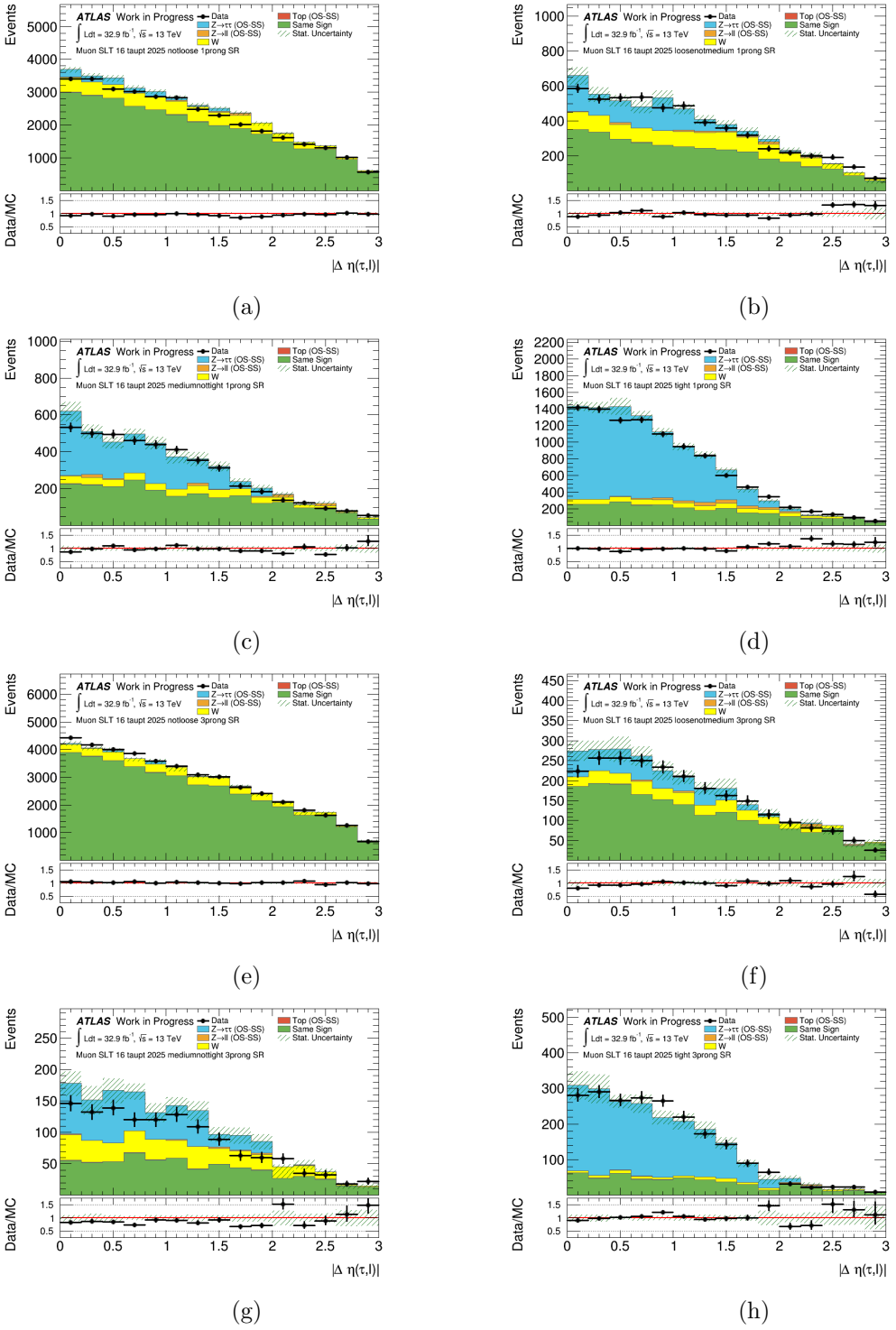


Figure 4.9.: $\Delta\eta(\mu,\tau)$ prefit distribution for 20 GeV $<$ p_T $<$ 25 GeV, 1 prong : *a)* not loose, *b)* loose not medium, *c)* medium not tight, *d)* tight ID categories; 3 prong : *e)* not loose, *f)* loose not medium, *g)* medium not tight, *h)* tight ID categories.

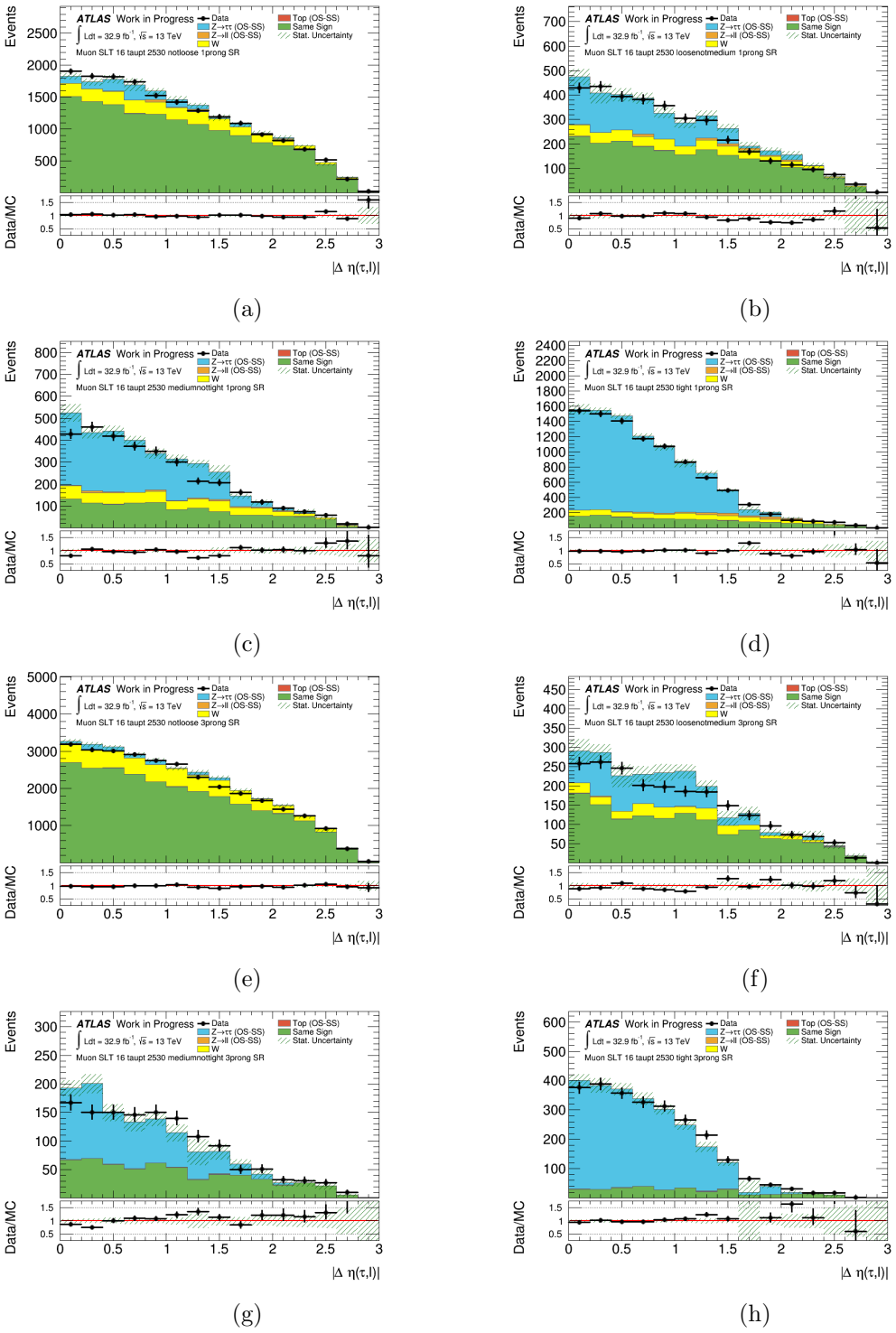


Figure 4.10.: $\Delta\eta(\mu,\tau)$ prefit distribution for 25 GeV $< p_T <$ 30 GeV, 1 prong : *a)* not loose, *b)* loose not medium, *c)* medium not tight, *d)* tight ID categories; 3 prong : *e)* not loose, *f)* loose not medium, *g)* medium not tight, *h)* tight ID categories.

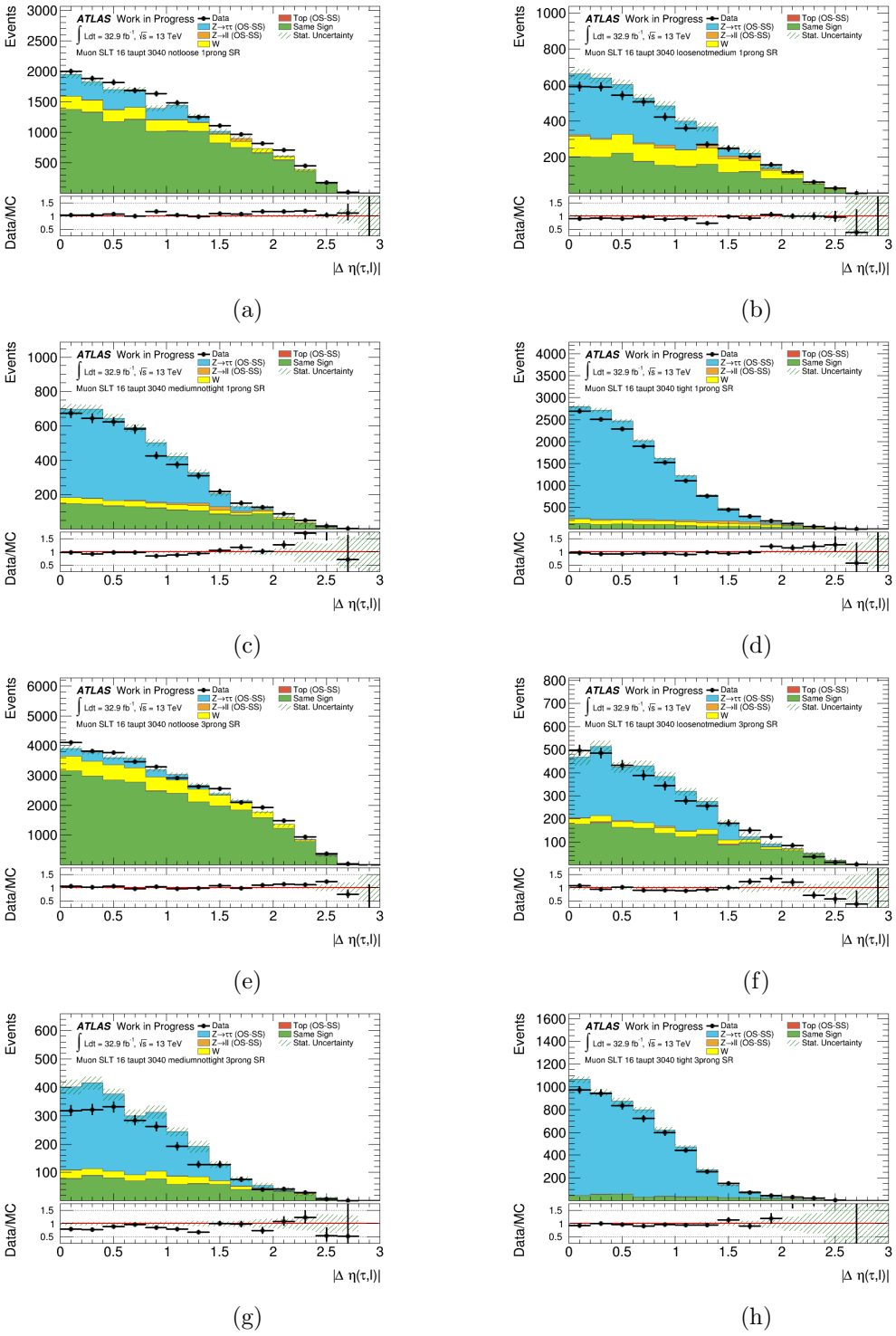


Figure 4.11.: $\Delta\eta(\mu,\tau)$ prefit distribution for 30 GeV $<$ p_T $<$ 40 GeV, 1 prong : *a)* not loose, *b)* loose not medium, *c)* medium not tight, *d)* tight ID categories; 3 prong : *e)* not loose, *f)* loose not medium, *g)* medium not tight, *h)* tight ID categories.

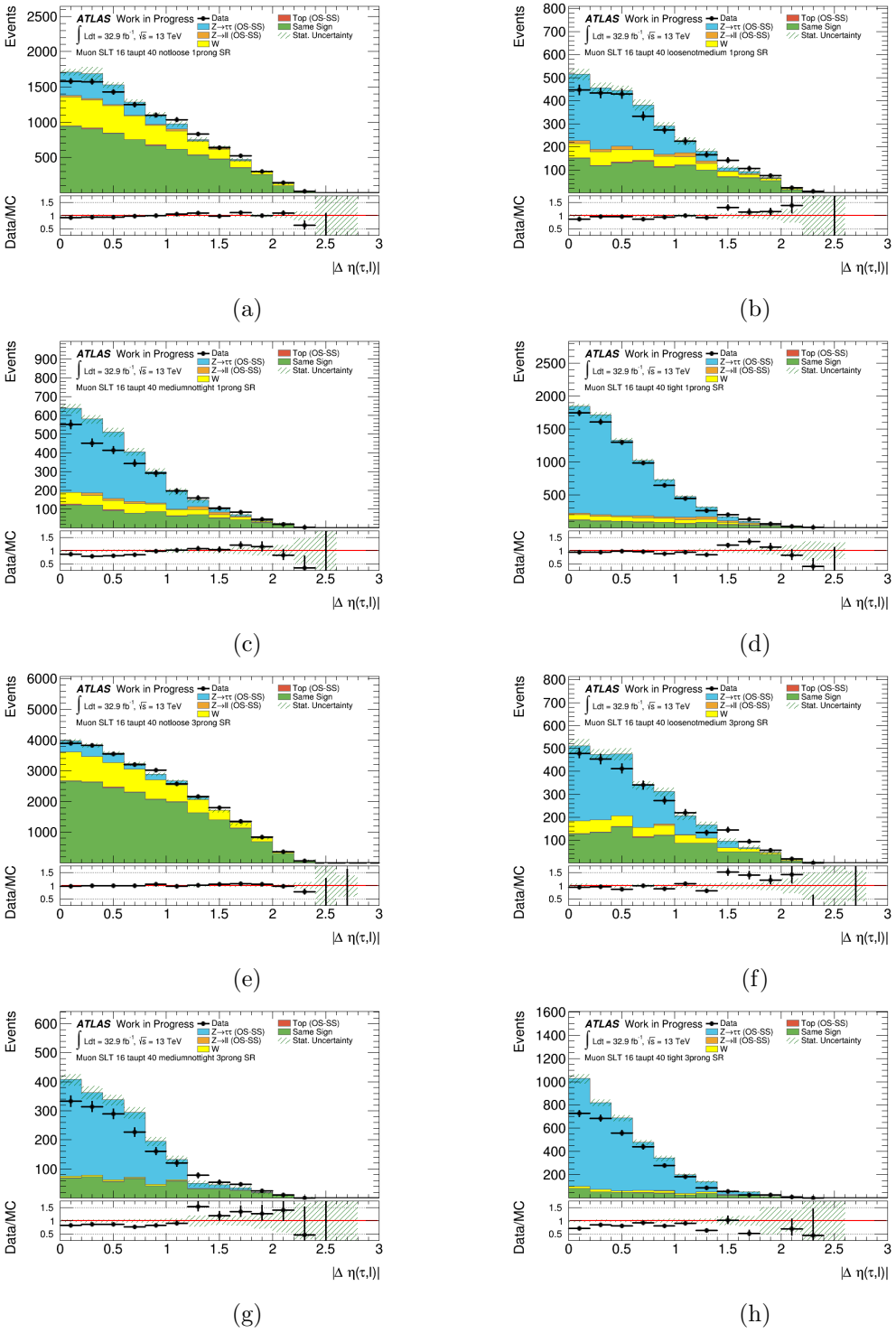


Figure 4.12.: $\Delta\eta(\mu,\tau)$ prefit distribution for $p_T > 40$ GeV, 1 prong : *a)* not loose, *b)* loose not medium, *c)* medium not tight, *d)* tight ID categories; 3 prong : *e)* not loose, *f)* loose not medium, *g)* medium not tight, *h)* tight ID categories.

Results

Postfit distributions for $\Delta\eta(\mu,\tau)$ in all analysis signal region categories are shown in Fig. 4.13-4.16. After the fit, each signal region and the extraction of the $Z \to \tau\tau$ normalisation, scale factors are computed comparing the signal selection efficiency pre/postfit for the different τ ID level: Loose, Medium, Tight. The selection efficiency is defined as the ratio between the number of events which pass an ID level with respect to the total number of selected events:

$$\epsilon = \frac{N(Z \to \tau\tau)_{pass-IDreg.}}{N(Z \to \tau\tau)_{all-reg.}} \tag{4.4}$$

Table 4.3 shows the pre/postfit efficiencies for different τ ID level in different τ p_T regions.
Scale factors are then extracted using the formula:

$$SF = \frac{\epsilon_{postfit}}{\epsilon_{prefit}} = \frac{\left(\frac{N(Z\to\tau\tau)_{pass-IDreg.}}{N(Z\to\tau\tau)_{all-reg.}}\right)_{postfit}}{\left(\frac{N(Z\to\tau\tau)_{pass-IDreg.}}{N(Z\to\tau\tau)_{all-reg.}}\right)_{prefit}} \tag{4.5}$$

Table 4.4 shows the final scale factor values for different τ ID level in different τ p_T regions. To get the breakdown of the error in statistical and systematic components, all the fit uncertainties have been divided in two sets: one set of uncertainties due to systematic effects (*systematic set*) and another set which includes uncertainties due to the event sample sizes in the control and signal regions (*statistical set*). Then the following procedure has been applied:

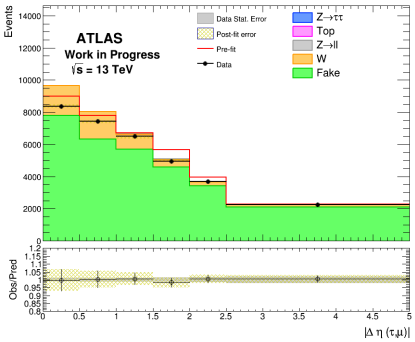
- a fit has been performed considering both the statistical and the systematics set of uncertainties (*main fit*) to extract both the statistical and systematic component of the error;
- a second fit has been performed fixing the uncertainties in the systematic set to their postfit values derived from the main fit; this will lead to the extraction of only the statistical component of the error;
- the systematic component of the error has been derived from the difference in quadrature of the errors extracted from the fits performed in the previous steps.

As further cross-check, scale factor values have been recalculated changing the fit setup to use different background and signal correlation schemes, like:

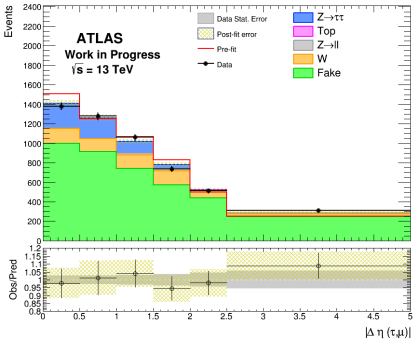
- fitting simultaneously more than one category with the same ID level but different p_T range;
- decorrelation of the Fake and W+jets normalisation factors.

All the different tests are giving compatible results to the ones reported in the tables, thus confirming their robustness. As mentioned at the beginning of this chapter, this is the first set of scale factors for τ ID efficiency binned in τ p_T recommended by the TauWG for Run 2 analyses.
For the next iterations of this analysis, future improvement can be:

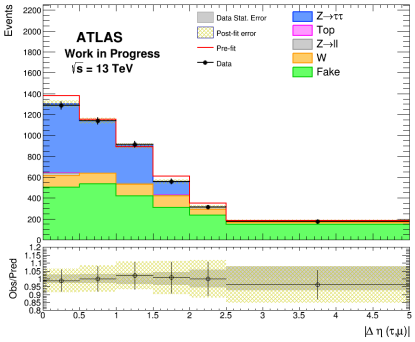
- addition of the $Z \to \tau\tau$ theory uncertainties for the Sherpa MC generator;
- improve model of W+jets background estimation or possibly increase the W+jets MC statistics;
- improve model of background rejection in the *notloose* ID category to improve fit stability.



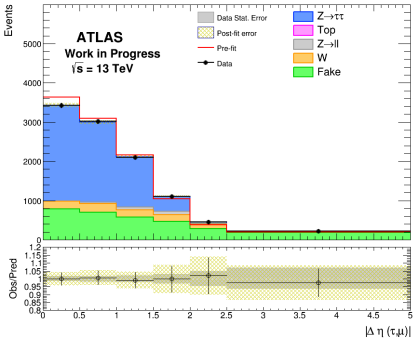
(a)



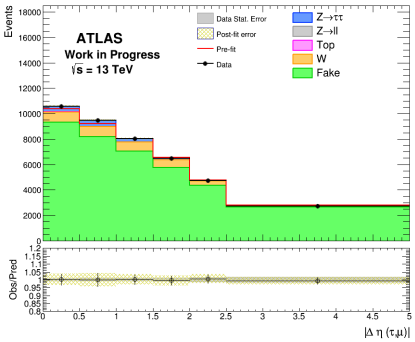
(b)



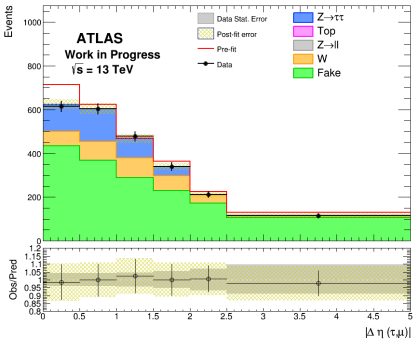
(c)



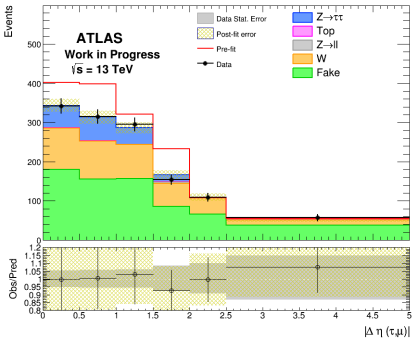
(d)



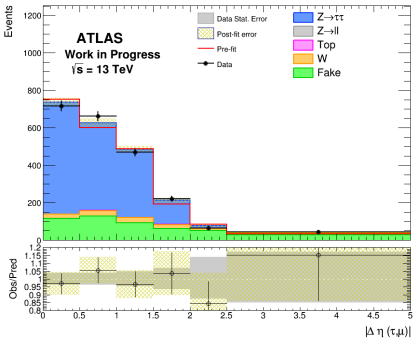
(e)



(f)

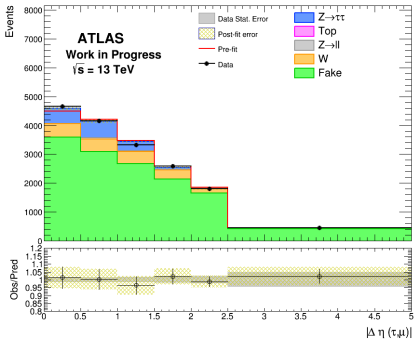


(g)

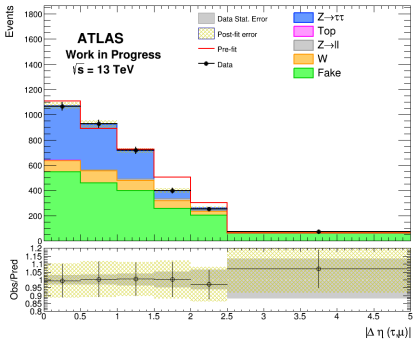


(h)

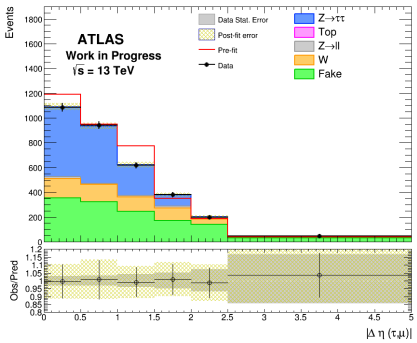
Figure 4.13.: $\Delta\eta(\mu,\tau)$ postfit distribution for 20 GeV $< p_T <$ 25 GeV, 1 prong : *a) not loose*, *b) loose not medium*, *c) medium not tight*, *d) tight* cat.; 3 prong : *e) not loose*, *f) loose not medium*, *g) medium not tight*, *h) tight* categories.



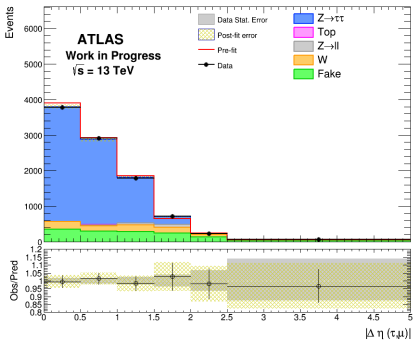
(a)



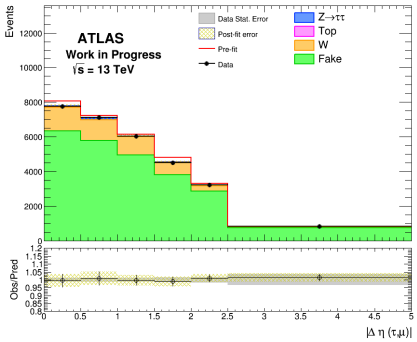
(b)



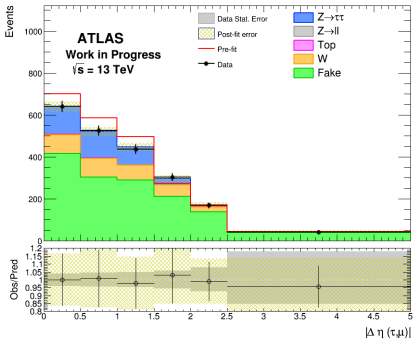
(c)



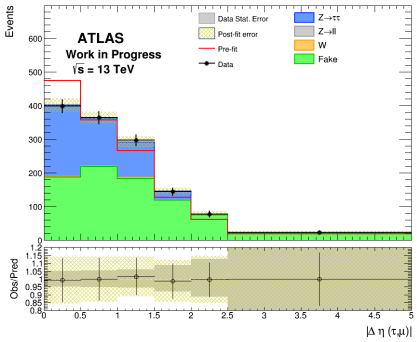
(d)



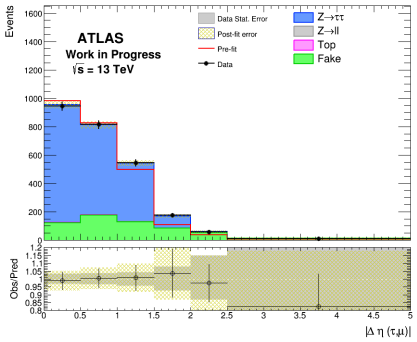
(e)



(f)

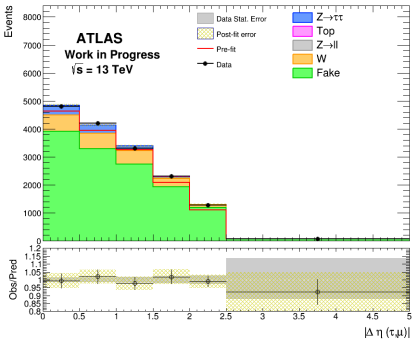


(g)

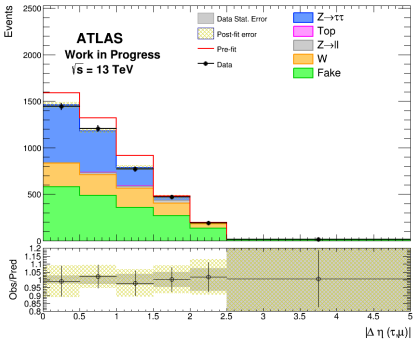


(h)

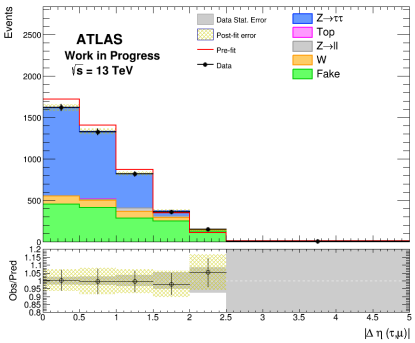
Figure 4.14.: $\Delta\eta(\mu,\tau)$ postfit distribution for 25 GeV $< p_T <$ 30 GeV, 1 prong : *a) not loose*, *b) loose not medium*, *c) medium not tight*, *d) tight* cat.; 3 prong : *e) not loose*, *f) loose not medium*, *g) medium not tight*, *h) tight* categories.



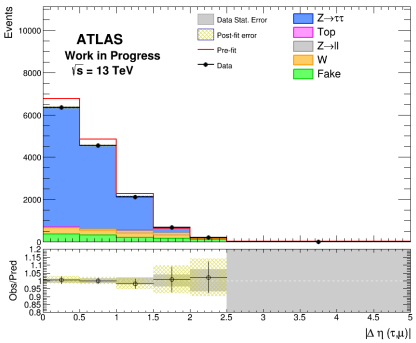
(a)



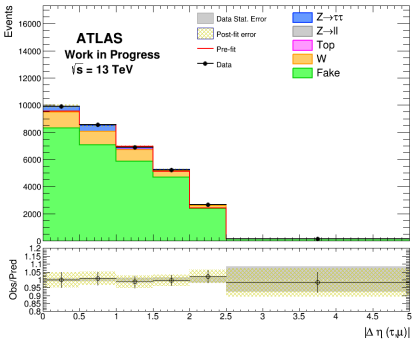
(b)



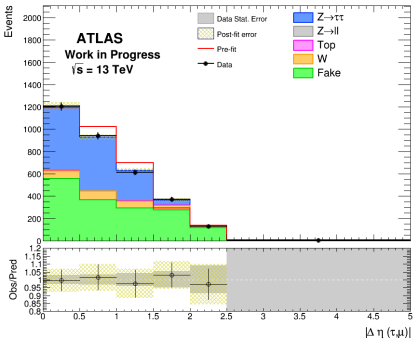
(c)



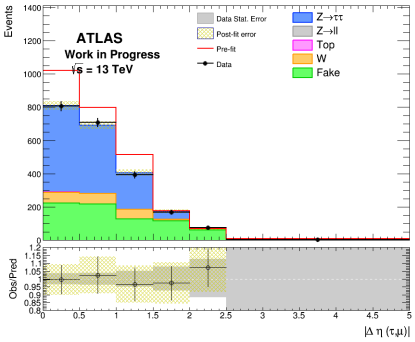
(d)



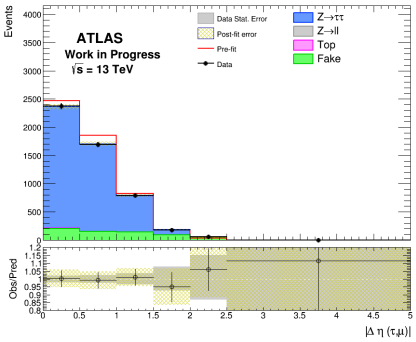
(e)



(f)

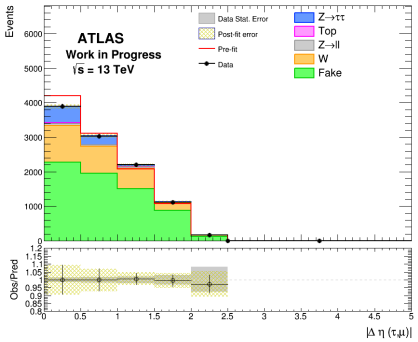


(g)

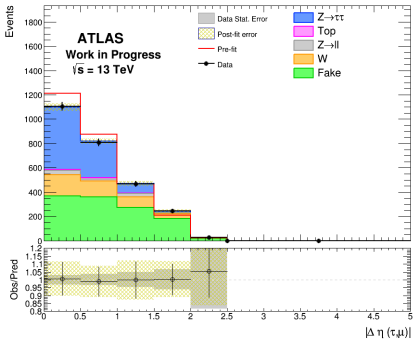


(h)

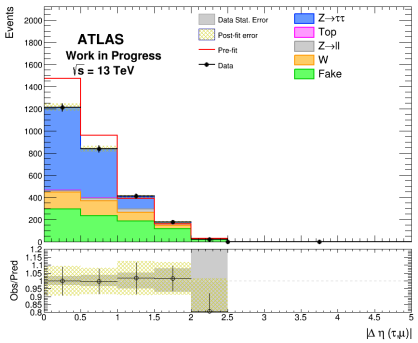
Figure 4.15.: $\Delta\eta(\mu,\tau)$ postfit distribution for 30 GeV $< p_T <$ 40 GeV, 1 prong : *a) not loose*, *b) loose not medium*, *c) medium not tight*, *d) tight* cat.; 3 prong : *e) not loose*, *f) loose not medium*, *g) medium not tight*, *h) tight* categories.



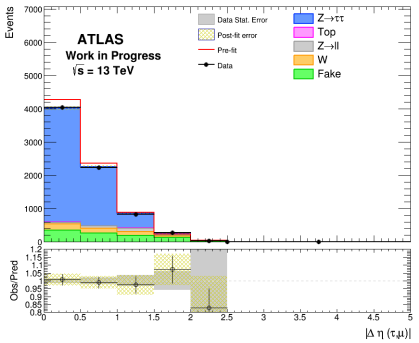
(a)



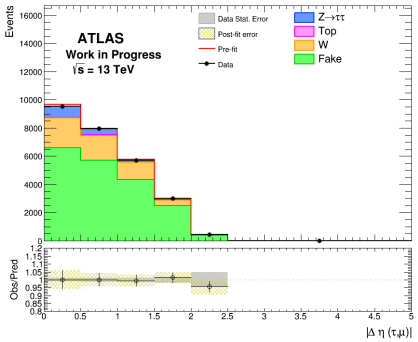
(b)



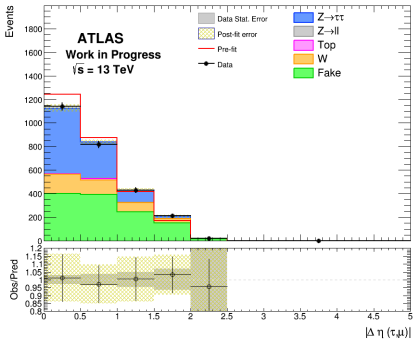
(c)



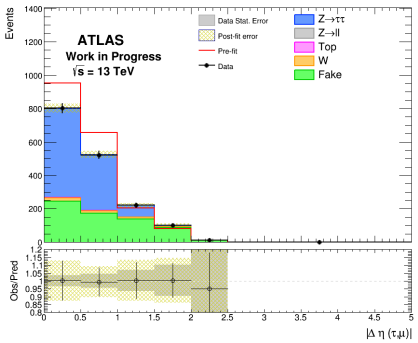
(d)



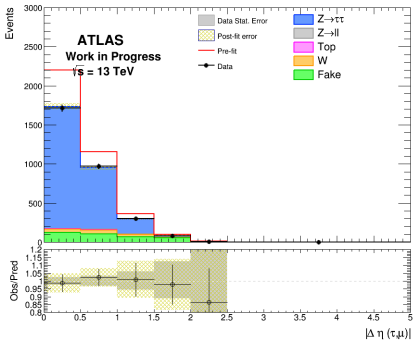
(e)



(f)



(g)



(h)

Figure 4.16.: $\Delta\eta(\mu,\tau)$ postfit distribution for $p_T > 40$ GeV, 1 prong : *a) not loose*, *b) loose not medium*, *c) medium not tight*, *d) tight* cat.; 3 prong : *e) not loose*, *f) loose not medium*, *g) medium not tight*, *h) tight* categories.

p_T (GeV)	ID	Prefit Eff 1P	Postfit Eff 1P	Prefit Eff 3P	Postfit Eff 3P
[20,25]	lse	0.893 ± 0.029	0.862 ± 0.055	0.869 ± 0.060	0.849 ± 0.176
	med	0.797 ± 0.026	0.796 ± 0.048	0.718 ± 0.050	0.688 ± 0.141
	tig	0.638 ± 0.021	0.629 ± 0.039	0.540 ± 0.040	0.634 ± 0.118
[25,30]	lse	0.901 ± 0.021	0.898 ± 0.059	0.824 ± 0.034	0.764 ± 0.101
	med	0.809 ± 0.019	0.795 ± 0.051	0.663 ± 0.027	0.664 ± 0.080
	tig	0.648 ± 0.016	0.669 ± 0.041	0.498 ± 0.022	0.532 ± 0.063
[30,40]	lse	0.903 ± 0.011	0.952 ± 0.035	0.816 ± 0.016	0.861 ± 0.066
	med	0.810 ± 0.010	0.864 ± 0.031	0.652 ± 0.013	0.693 ± 0.054
	tig	0.661 ± 0.008	0.717 ± 0.025	0.489 ± 0.010	0.545 ± 0.043
[40,-]	lse	0.869 ± 0.016	0.912 ± 0.056	0.789 ± 0.018	0.782 ± 0.106
	med	0.753 ± 0.014	0.807 ± 0.050	0.607 ± 0.015	0.615 ± 0.079
	tig	0.585 ± 0.011	0.657 ± 0.040	0.427 ± 0.011	0.449 ± 0.060

Table 4.3.: Pre/Postfit $Z \to \tau\tau$ efficiencies for 1 and 3 prong taus.

p_T (GeV)	ID	SF 1P	SF 3P
[20,25]	lse	0.965 ± 0.055 ± 0.040	0.980 ± 0.196 ± 0.085
	med	0.999 ± 0.056 ± 0.039	0.958 ± 0.191 ± 0.083
	tig	0.985 ± 0.056 ± 0.040	1.175 ± 0.225 ± 0.069
[25,30]	lse	0.996 ± 0.045 ± 0.053	0.926 ± 0.123 ± 0.037
	med	0.983 ± 0.043 ± 0.051	1.001 ± 0.122 ± 0.039
	tig	1.032 ± 0.045 ± 0.051	1.069 ± 0.129 ± 0.041
[30,40]	lse	1.053 ± 0.030 ± 0.027	1.055 ± 0.067 ± 0.050
	med	1.066 ± 0.030 ± 0.026	1.062 ± 0.067 ± 0.053
	tig	1.084 ± 0.030 ± 0.026	1.114 ± 0.069 ± 0.058
[40,-]	lse	1.050 ± 0.045 ± 0.050	0.991 ± 0.076 ± 0.113
	med	1.072 ± 0.046 ± 0.051	1.014 ± 0.077 ± 0.108
	tig	1.121 ± 0.048 ± 0.053	1.052 ± 0.080 ± 0.113

Table 4.4.: Scale factor values for 1 and 3 prong taus; uncertainty has been split into *Stat* and *Syst* component.

CHAPTER 5

Di-tau mass estimation

The sensitivity of a *mass bump-hunting* analysis heavily depends on how narrow the signal invariant mass distribution is compared to the distribution of the background processes. Unfortunately, this search strategy is less effective in searches of resonances decaying to a pair of τ leptons because of undetected τ energy associated with neutrinos. In proton-proton collisions, the full energy of neutrinos cannot be determined, since it is only possible to reconstruct a transverse energy imbalance in the calorimeter which is representative of the total transverse momentum of all neutrinos in the event, the E_T^{miss}. Therefore, when two or more neutrinos are produced in the same event, their individual transverse momenta and directions cannot be reconstructed. In the case of resonances decaying to a pair of τ, the two taus are often produced back-to-back and the missing momentum associated with their neutrinos partially cancels out. As a result, the invariant mass of a resonance cannot be directly reconstructed from the E_T^{miss} and visible decay products of the τ leptons. In the following chapter, two methods used to estimate the τ pair mass are described: the *Collinear Approximation* and the *Missing Mass Calculator* [48].

5.1. The Collinear Approximation Technique

This method is based on two important assumptions:

- the neutrinos from each τ decay are nearly collinear with the corresponding τ visible decay products ($\phi_\nu \simeq \phi_{vis}$ and $\theta_\nu \simeq \theta_{vis}$) as a result of a large boost;

- the E_T^{miss} in the event is due only to neutrinos.

In this case, the total invisible momentum carried away by neutrinos in each τ decay can be estimated by solving two equations:

$$\begin{aligned} E_{T_x}^{miss} &= p_{mis1}\sin(\theta_{vis1})\cos(\phi_{vis1}) + p_{mis2}\sin(\theta_{vis2})\cos(\phi_{vis2}) \\ E_{T_y}^{miss} &= p_{mis1}\sin(\theta_{vis1})\sin(\phi_{vis1}) + p_{mis2}\sin(\theta_{vis2})\sin(\phi_{vis2}) \end{aligned} \tag{5.1}$$

where $E_{T_x}^{miss}$ and $E_{T_y}^{miss}$ are the x- and y-components of the E_T^{miss} vector, p_{mis1} and p_{mis2} are the combined invisible momenta of each τ decay, and $\theta_{vis1,2}$ and $\phi_{vis1,2}$ are the polar and azimuthal angles of the visible products of each τ decay. From Eqn. 5.1, the invariant mass of the $\tau\tau$-system is given by:

$$M_{\tau\tau} = \frac{m_{vis}}{\sqrt{x_1 x_2}} \tag{5.2}$$

where m_{vis} is the visible mass of the τ decay products and

$$x_i = \frac{p_{vis-i}}{p_{vis-i} + p_{mis-i}} \tag{5.3}$$

are the momentum fractions carried away by the visible decay products. The Collinear Approximation gives a good mass resolution only for the fraction of events where the $\tau\tau$-system is boosted, that is produced in association with a large momentum jet, and the visible τ decay products are not back-to-back in the transverse plane to the beam line. The last requirement is needed because the system of equations becomes degenerate if $\phi_{vis1} = \phi_{vis2} + \pi$ and solutions $p_{mis1,2} \simeq \sin^{-1}(\phi_{vis1} - \phi_{vis2})$ diverge as $|\phi_{vis1} - \phi_{vis2}| \to \pi$. This technique is also very sensitive to the E_T^{miss} resolution and tends to over-estimate the $\tau\tau$ mass, leading to long tails in the reconstructed mass distribution.

5.2. Missing Mass Calculator (MMC)

Assuming perfect detector resolution and that there are no neutrinos in the $\tau\tau$ events except for those from the τ lepton decays, the full reconstruction of the event topology requires an estimation from six to eight unknowns: the x-, y- and z-components of the invisible momentum carried away by neutrino(s) for each of the two τ leptons in the event, and, if one or both τ's decay leptonically, the invariant mass of the neutrinos from each τ decay. The event kinematics is described by four equations connecting these unknowns:

$$\begin{aligned} E_{T_x}^{miss} &= p_{mis1}\sin(\theta_{mis1})\cos(\phi_{mis1}) + p_{mis2}\sin(\theta_{mis2})\cos(\phi_{mis2}) \\ E_{T_y}^{miss} &= p_{mis1}\sin(\theta_{mis1})\sin(\phi_{mis1}) + p_{mis2}\sin(\theta_{mis2})\sin(\phi_{mis2}) \\ M_{\tau 1}^2 &= m_{mis1}^2 + m_{vis1}^2 + 2\sqrt{p_{vis1}^2 + m_{vis1}^2}\sqrt{p_{mis1}^2 + m_{mis1}^2} - 2p_{vis1}p_{mis1}\cos\Delta\theta_{\nu m1} \\ M_{\tau 2}^2 &= m_{mis2}^2 + m_{vis2}^2 + 2\sqrt{p_{vis2}^2 + m_{vis2}^2}\sqrt{p_{mis2}^2 + m_{mis2}^2} - 2p_{vis2}p_{mis2}\cos\Delta\theta_{\nu m2} \end{aligned} \tag{5.4}$$

where $E_{T_x}^{miss}$ and $E_{T_y}^{miss}$ are the x- and y-components of the E_T^{miss} vector, $p_{vis1,2}$, $m_{vis1,2}$, $\theta_{vis1,2}$, $\phi_{vis1,2}$ are the momenta, invariant masses, polar and azimuthal angles of the visible τ decay products, and $M_\tau = 1.777$ GeV/c^2 is the τ invariant mass. The rest of the variables constitute the unknowns. Finally, $\Delta\theta_{\nu m1,2}$ is the angle between the vectors p_{mis} and p_{vis} for each of the two τ leptons, and it can be expressed in terms of the other variables. Considering hadronic τ decays, the m_{mis} is set to 0 as there is only one neutrino involved in the decay.
It should be noted that not all the solutions of the under-constrained system are equally probable, and additional knowledge of τ decay kinematics can be used to distinguish more likely solutions from less likely ones. This information can be provided knowing the distributions of (for example) the opening angle $\Delta\theta$. These distributions can only be produced based on MC samples due to mandatory knowledge of the *truth* information for the neutrinos, where truth means using generated particles before detector effects are simulated. This information about decay kinematics is incorporated as probability density functions in a global event likelihood to provide additional constraints and obtain a better estimator of $M_{\tau\tau}$:

$$L = -\log(P(\Delta\theta_1, p_{T_{\tau 1}}) \times P(\Delta\theta_2, p_{T_{\tau 2}})) \tag{5.5}$$

In the case in which one τ decays hadronically and the other one leptonically ($\tau_{lep}\tau_{had}$) there are three unknowns in the system of equations in Eqn. 5.4; thus the system can be solved exactly for any point in the ϕ_{miss1}, ϕ_{miss2} and m_{mis1} parameter space. In this way, the $p_{mis1,2}$ are fully defined and it is possible to determine $\Delta\theta_{1,2}$ and calculate the corresponding event weight based on the likelihood value. The most probable value of the $M_{\tau\tau}$ distribution is then used as best estimate of $M_{\tau\tau}$.
Another important dependence in the MMC calculation is the E_T^{miss} resolution, so the MMC implementation has to be adjusted to allow for possible E_T^{miss} mis-measurements. This is achieved by increasing the dimensionality of the parameter space in which the scanning is performed to include the two components of the E_T^{miss} resolution. In this case, the event likelihood becomes:

$$L = -\log(P(\Delta\theta_1, p_{T_{\tau 1}}) \times P(\Delta\theta_2, p_{T_{\tau 2}}) \times P(\Delta E_{T_x}^{miss}) \times P(\Delta E_{T_y}^{miss})) \tag{5.6}$$

where the probability density functions $P(\Delta E_{T_x}^{miss})$ and $P(\Delta E_{T_y}^{miss})$ are defined as:

$$P(\Delta E_{T_{x,y}}^{miss}) = e^{\left(-\frac{(\Delta E_{T_{x,y}}^{miss})^2}{\sigma^2}\right)} \tag{5.7}$$

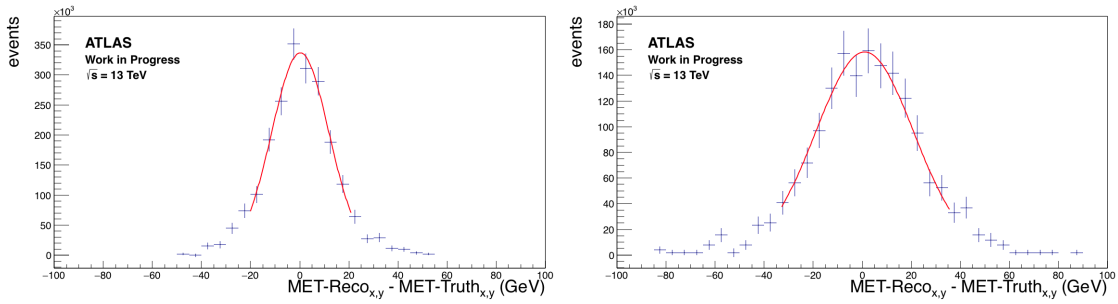
where σ is the resolution and $P(\Delta E_{T_{x,y}}^{miss})$ are the variations in the x- or y- components of the E_T^{miss}. The σ parameterisation has been determined in dedicated MMC retuning studies which are described in the next paragraph.

5.3. MMC retuning for the $\tau_{lep}\tau_{had}$ final state

The MMC was one of the most powerful tools used in $H \to \tau\tau$ searches in Run 1 at the LHC. Actually, in Run 2, the LHC collides proton-proton at a center of mass energy $\sqrt{s} = 13$ TeV and at higher luminosity, so the MMC is required to be retuned in order to continue to play a key role in the analysis. In particular the MMC dependency on the E_T^{miss} resolution needs to be studied due to changes in the ATLAS detector between Run 1 and Run 2 and the increase number of interactions per bunch crossing.
A new parameterisation for E_T^{miss} resolution as function of $\sqrt{\Sigma E_T}$ has been derived considering both $Z \to \tau\tau$+jets and gluon/vector boson fusion $H \to \tau\tau$ simulated samples, where ΣE_T is the scalar sum of all the objects which are used for the E_T^{miss} calculation (electrons, muons, taus, jets, E_T^{miss} soft term). Events have been selected using a looser selection criteria with respect to the ones described in Section 6.3 . The E_T^{miss} resolution is defined as:

$$\sigma_{x(y)} = \sigma(p^{reco}_{miss-x(y)} - p^{truth}_{miss-x(y)}) \tag{5.8}$$

and it is extracted fitting the $p^{reco}_{miss-x(y)} - p^{truth}_{miss-x(y)}$ distribution in bins of 25 GeV size in ΣE_T in the range 100 GeV $< \Sigma E_T <$ 350 GeV, which is the relevant ΣE_T range for this analysis. Examples of fitted distributions are shown in Fig. 5.1. The Gaussian fit has been restricted in the range of 1.5 times distribution RMS with respect to the mean, in order to avoid events in the tails which can bias the fit.



(a) (b)

Figure 5.1.: Fitted $p^{reco}_{miss-x(y)} - p^{truth}_{miss-x(y)}$ distributions for: *a)* 125 GeV$< \Sigma E_T <$ 150 GeV, *b)* 250 GeV $< \Sigma E_T <$ 275 GeV.

After having extracted the E_T^{miss} resolution from the Gaussian fit, this has been parameterised as function of $\sqrt{\Sigma E_T}$; the parameterisation for the $Z \to \tau\tau$+jets sample is shown in Fig. 5.2. Similar parameterisations have been extracted also for the $H \to \tau\tau$ samples mentioned before. It should be noted that the E_T^{miss} resolution can also depend on the final state topology, since the relative position of the taus in the final state can influence the hadronic activity of the event. This is particularly relevant for the Boosted final state, in which the analysis selection criteria require a jet which is recoiling against the $\tau\tau$ system. To take into account event topology, a correction term (*scale factor*)

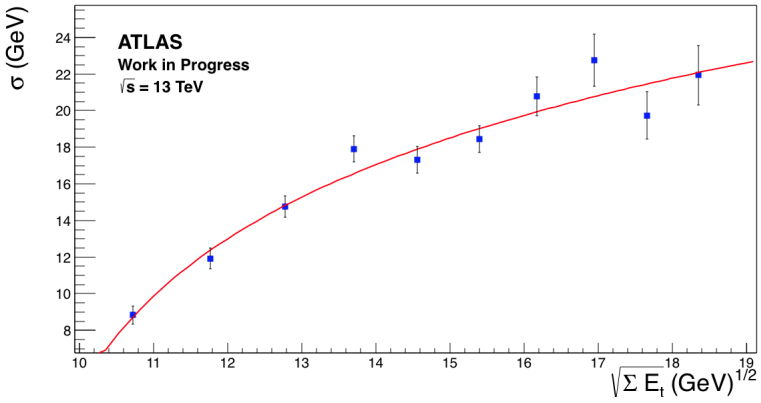


Figure 5.2.: The E_T^{miss} resolution parameterisation as function of $\sqrt{\Sigma E_T}$.

dependent on $\Delta\phi(\tau\tau)$ has been studied so that the final resolution is given by:

$$\sigma_{E_T^{miss}} = \sigma_{prev} * \underbrace{f\left[\left(\frac{\sigma_{x(y)}}{\sigma_{prev}}\left(\Delta\Phi(\tau 1, \tau 2)\right)\right)\right]}_{\text{scale factor}} \tag{5.9}$$

where σ_{prev} is the resolution parametrised as function of $\sqrt{\Sigma E_t}$. The correction term is derived comparing the observed and the parametrised E_T^{miss} resolution in bins of $\Delta\phi(\tau\tau)$. The ratio is then parametrised using a polynomial function in order to extract a functional form for the $\Delta\phi(\tau\tau)$ dependency, as shown in Fig. 5.3. Similar parameterisations have been extracted also for the $H \to \tau\tau$ samples. As cross check, the MMC shape using

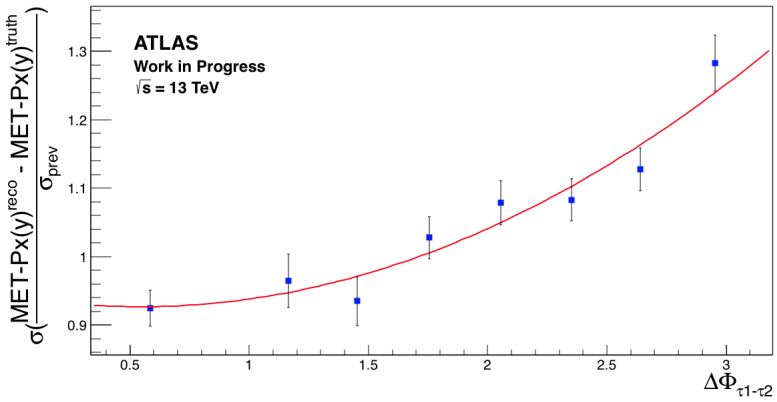


Figure 5.3.: The E_T^{miss} resolution correction factor as function of $\Delta\phi(\tau\tau)$.

the new E_T^{miss} resolution parameterisation has been checked in looser VBF and Boosted

region with respect to the ones described in section 6.5, as shown in Fig. 5.4 and 5.5. Looking at these plots, the Z and the Higgs peak positions match the expectations, giving confidence that the MMC can be used in the analysis. The parameterisation derived

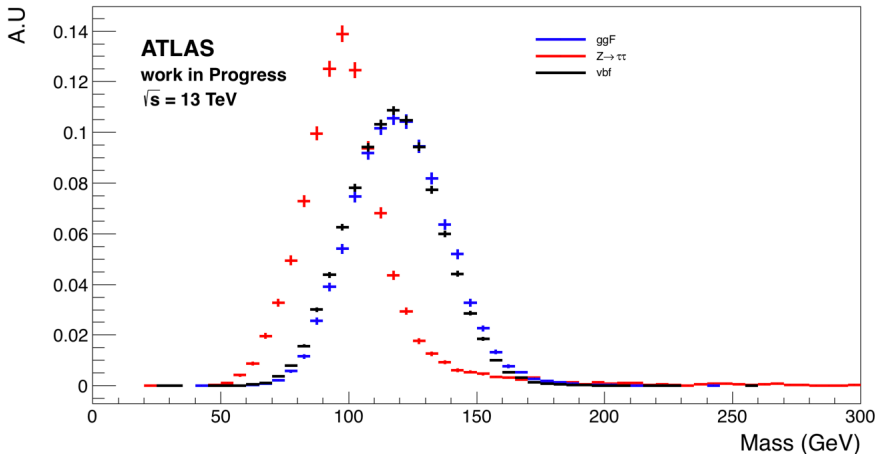


Figure 5.4.: The MMC distribution in Boosted-like region. The VBF and the ggF H→ $\tau\tau$ samples are generated with m_H = 125 GeV.

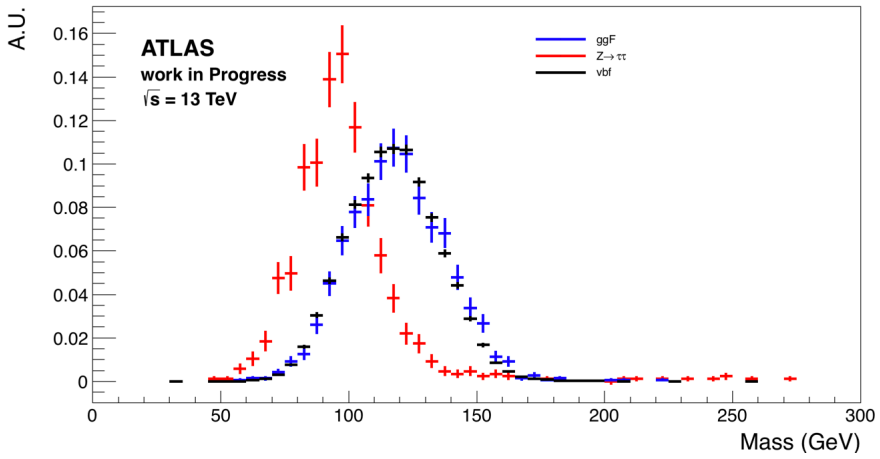


Figure 5.5.: The MMC distribution in VBF-like region. The VBF and the ggF H→ $\tau\tau$ samples are generated with m_H = 125 GeV.

using the Eqn. 5.9 has been inserted in the MMC package and it is currently used in the analysis described in the Chapter 6.

5.4. Comparison of the Collinear Approximation and MMC mass estimation

Since the mass of the $\tau\tau$ system is the main discriminant variable between the signal and the irreducible $Z \to \tau\tau$ background in the $H \to \tau\tau$ analysis, it is useful to understand which of the two mass estimation techniques described in the previous sections is providing the best separation power. In Fig. 5.6, the MMC and the Collinear Approximation mass distributions in the current $H \to \tau\tau$ signal regions are shown. The separation power between signal and background has been calculated normalising both histograms to 1 and using the formula:

$$Y = \frac{1}{2}\sum_{i=1}^{bins}\frac{(N_i^{sig} - N_i^{bkg})^2}{N_i^{sig} + N_i^{bkg}} \tag{5.10}$$

where N_i^{sig} and N_i^{bkg} are the number of signal/background events in the bin i. The separation is 0 for identical signal and background shapes, and it is 1 for shapes with no overlap. The separation power comparison between MMC and Collinear Approximation technique is shown in Table 5.1. From these results, the MMC seems to perform better in all the analysed signal regions and therefore it has been chosen as the mass estimation method for the $H \to \tau\tau$ analysis.

Region	MMC (%)	Coll. Approx (%)
Boosted	55.97	49.51
VBF	49.32	41.72

Table 5.1.: Separation power comparison between MMC and Coll. Approx mass.

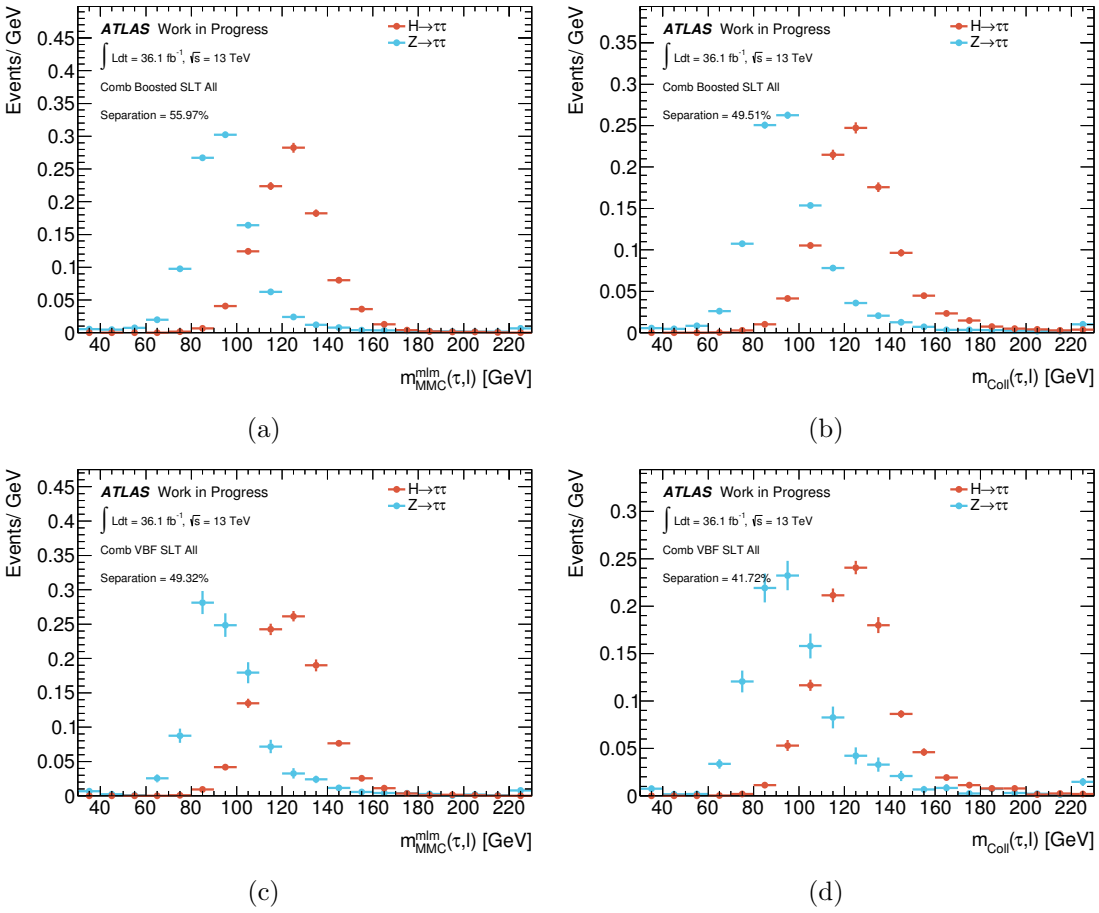


Figure 5.6.: MMC and Coll. Approx. signal/background comparison in $H \to \tau\tau \to \tau_{lep}\tau_{had}$ final state signal regions: *a),b)* Boosted SRs, *c),d)* VBF SRs. The H→ $\tau\tau$ sample is generated with $m_H = 125$ GeV.

CHAPTER 6

Search for $H \to \tau\tau$ decay in the $\tau_{lep}\tau_{had}$ final state

In this chapter, the search for the SM $H \to \tau\tau$ decay will be described. According to the τ decay, the analysis can be split into three final states (or channels):

- $\tau_{lep}\tau_{lep}$, in which both taus decay leptonically in electrons or muons;
- $\tau_{lep}\tau_{had}$, in which one τ decays leptonically and the other one decays hadronically;
- $\tau_{had}\tau_{had}$, in which both taus decay hadronically.

The branching ratios of the different final states are shown in Fig. 6.1.

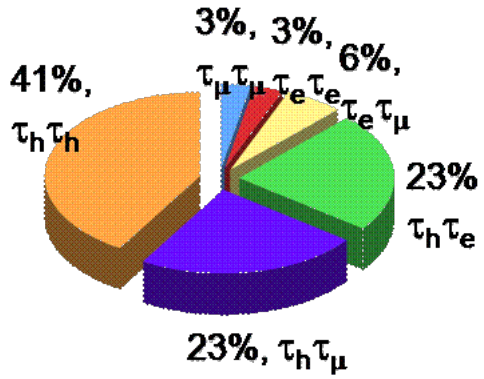


Figure 6.1.: Branching ratios of the different $H \to \tau\tau$ decay channels.

The main analysis selection criteria have been harmonised across the three channels in order to choose signal regions selecting similar phase space. The analysis is mainly

targeting vector-boson fusion and gluon-fusion Higgs production modes, so the analysis strategy has been optimised based on the kinematic features of these modes. In the following sections, only the $\tau_{lep}\tau_{had}$ final state will be discussed deeply: in particular, the analysis selection criteria and the background estimation method will be treated. At the end of the chapter, results based on a standalone $\tau_{lep}\tau_{had}$ fit on the MMC, introduced in Chapter 5, as well as results for the combined fit of all the three channels will be presented.

6.1. Data and Monte Carlo Samples

In this section, a brief description of the data and the Monte Carlo (MC) samples used in the analysis will be given.

6.1.1. Data samples

This analysis uses the full 2015 and 2016 datasets taken at 25 ns operation of the LHC, corresponding to a total integrated luminosity of 36.19 fb^{-1}. Both 2015 and 2016 datasets have been filtered according to the data quality requirements recommended by ATLAS. In particular, data taken in the period when the Insertable B-Layer (the innermost detector of the ATLAS experiment) was not fully operational have been discarded.

6.1.2. Signal samples

The VBF and associated VH production processes are simulated at NLO accuracy using Powheg-Box v2 ([49],[50],[51],[52]) with the MiNLO approach [53], interfaced to the Pythia 8.186 parton shower model [54]. The VBF samples are normalised to an approximate NNLO QCD cross section with NLO EW corrections applied ([55],[56],[57]), while VH samples are normalised to cross sections calculated at NNLO in QCD, with NLO EW radiative corrections applied ([58],[59],[60]).
Production via gluon fusion (ggH or ggF) is simulated at NNLO accuracy in QCD using the Powheg NNLOPS program [61]. The PDF4LHC15 [62] parametrisation of the parton distribution functions (PDFs) is used in the matrix element of all the production processes. The AZNLO [63] tune is used, with PDF set CTEQ6L1 [64], for the modelling of the non-perturbative effects. Photos++ version 3.52 [65] is used for QED emissions from electroweak vertices and charged leptons ([66],[67],[68],[69]). The overall normalisation of the gluon fusion process is taken from the N^{3}LO QCD calculation with NLO EW corrections applied.

6.1.3. V+jets and Diboson samples

Background samples of electroweak production of W/Z bosons from VBF, W/Z bosons associated with jets and Diboson (VV) processes are simulated with the Sherpa [70] generator. Matrix elements are calculated at leading-order using the Comix [71] and Open Loops [72] matrix element generators and merged with the Sherpa parton shower [73]

using the ME+PS@NLO prescription [74]. The NNPDF30NNLO [75] PDF set is used in conjunction with dedicated parton shower tuning developed by the Sherpa author.

6.1.4. Top samples

For the generation of $t\bar{t}$ production, the Powheg-Box v2 generator with CT10 PDF sets in the matrix element calculations is used. Electroweak top quark production is generated using the Powheg-Box v1 [76],[77] generator. This generator uses the 4-flavour scheme for the NLO matrix elements calculations together with the fixed four-flavour PDF set CT10$_F$4. For all top processes, top-quark spin correlations are preserved. The parton shower, fragmentation, and the underlying event are simulated using Pythia 6.428 [78] with the CTEQ6L1 PDF set and the corresponding Perugia 2012 tune [79].

An overview of the MC generators used in this analysis is shown in Table 6.1.

Process	Generator		PDF set		Order
	ME	PS	ME	PS	
$H \to \tau\tau$					
ggF	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	NNLO+NNLL
VBF	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	(N)NLO
VH	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	NNLO
Background					
$V + jets$	SHERPA 2.2.1		NNPDF30		NNLO
$t\bar{t}$	Powheg	Pythia6	CT10	CTEQ6L1	NNLO+NNLL
Single top	Powheg	Pythia6	CT10	CTEQ6L1	NNLO
Di-Boson	SHERPA 2.2.1		NNPDF30		NNLO

Table 6.1.: Overview over the MC generators used to simulate signal and background samples.

For all samples, a full simulation of the ATLAS detector response [80] using the Geant4 program [81] was performed. The effect of multiple pp interactions in the same and neighbouring bunch-crossing is included by overlaying minimum-bias events simulated with Pythia 8 using the MSTW2008LO [82] PDF and the A2 [83] set of tuned parameters on each generated signal and background event. The number of overlaid events is such that the distribution of the average number of interactions per pp bunch crossing in the simulation matches the one observed in data. The resulting simulated events are processed through the same reconstruction code as the data.

6.2. Object Reconstruction

The topology of the $H \to \tau\tau$ events makes it necessary to reconstruct τ leptons, electrons, muons and missing transverse energy (E_T^{miss}). Additional jets in the event are used in the categorisation of VBF and boosted topologies and to suppress backgrounds.

Electron candidates are reconstructed from energy deposits in the cells of the electromagnetic calorimeter associated with a charged particle track measured in the inner detector [84], [85], [86]. The candidates are reconstructed with a likelihood-based identification selection using the electromagnetic shower shape, track quality and track-cluster association, and they are required to have a *loose* identification criteria with a transverse momentum $p_T > 15$ GeV and to be in the fiducial volume of the inner detector, $|\eta| \leq 2.47$. The transition region between the barrel and the end-cap calorimeters ($1.37 \leq |\eta| \leq 1.52$) is excluded.
Muon candidates are reconstructed from track segments in the muon spectrometer matched with tracks found in the inner detector within $|\eta| \leq 2.5$. The tracks of the muon candidates are refitted using the complete track information from both detector systems [87]. The candidates are required to have a transverse momentum $p_T > 10$ GeV and to pass a *loose* muon identification requirement.
Jets are reconstructed from topological clusters of energy depositions in the calorimeter using the anti-k_t algorithm with a distance parameter $R = 0.4$. The jet energy is calibrated using the electromagnetic+JES scaling scheme. Only jets with $p_T > 20$ GeV and $|\eta| \leq 4.5$ are considered. In order to suppress jets originating from pile-up, jets with $p_T \leq 50$ GeV and $|\eta| < 2.4$ are required to have $|JVT| > 0.59$, where JVT is the output of the jet vertex tagger algorithm [88] which is used to identify and select jets originating from hard-scatter interaction through the use of tracking and vertex information. In the forward region, pile-up jets are suppressed by requiring that jets with $p_T \leq 50$ GeV and $|\eta| > 2.5$ will pass the selection of the fJVT algorithm, fJVT > 0.4 [89].
In order to identify the jets initiated by b-quarks, the *MV2c10* tagger is used [90]. A working point which corresponds to an average efficiency of 85 % for b-jets in $t\bar{t}$ events is chosen. Tagging and mis-tagging efficiency scale factors have been used to correct simulated events after having applied the b-tagging algorithms to the jets. The b-tagged jets are required to pass the $p_T > 20$ GeV and $|\eta| \leq 2.4$ requirements.
Tau reconstruction is discussed in detail in Chapter 4. An additional dedicated likelihood-based veto is used to reduce the electron mis-identified as a $\tau_{had-vis}$, providing about 95% selection efficiency for true 1-track τ candidates and a reduction of electrons faking τ leptons by a factor of 20 to 200. The hadronic τ candidates are required to have $p_T > 20$ GeV and $|\eta| \leq 2.5$ with *loose* identification requirement. The transition region between the barred and end-cap calorimeters is excluded.
The E_T^{miss} is calculated as the module of the negative vectorial sum of the p_T of all fully reconstructed and calibrated physics objects described before. The *soft term* which is formed by the inner-detector tracks originating from the hard-scattering vertex that are not associated to the reconstructed objects is added in the E_T^{miss} calculation [91]. All the selected objects have later been passed through the standard overlap removal procedure as recommended by ATLAS; this is ensuring that the various objects considered in the analysis have been well reconstructed and not overlapped.

6.3. Preselection Region

The analysis defines a *Preselection* region which is used to validate the background modelling of all relevant processes. The $\tau_{lep}\tau_{had}$ channel uses mainly single lepton triggers (SLT). A low electron/muon p_T region can also be defined using τ+lepton trigger (TLT). This region is currently not fully used in the analysis and it will be described later in Chapter 7. To select the SLT region, the lowest unprescaled triggers have been used, as summarised in Table 6.2. Electron/muon p_T thresholds of 25 (27) GeV and 21 (27.3) GeV are required for trigger matched lepton candidates in the 2015 (2016) dataset, respectively. The different phase spaces which can be selected according to the used triggers and based on electron/muon and τ p_T requirements are shown in Fig. 6.2.

Trigger menu	Data period	Chain name (in the menu)
Single electron	2015	HLT_e24_lhmedium_L1EM20VH
		HLT_e60_lhmedium
		HLT_e120_lhloose
	2016	HLT_e26_lhtight_nod0_ivarloose
		HLT_e60_lhmedium_nod0
		HLT_e140_lhloose_nod0
Single muon	2015	HLT_mu20_iloose_L1MU15
		HLT_mu50
	2016	HLT_mu26_ivarmedium
		HLT_mu50

Table 6.2.: Single lepton triggers used in the $\tau_{lep}\tau_{had}$ channel.

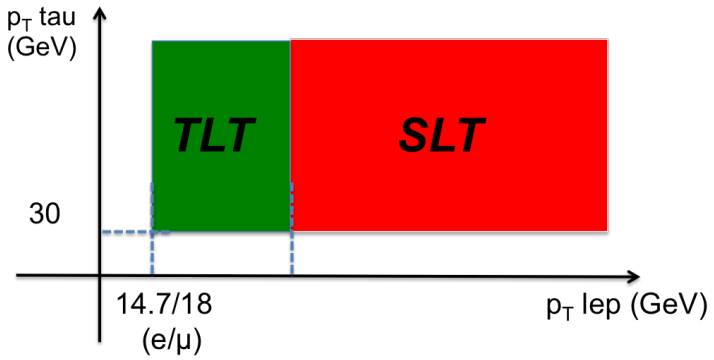


Figure 6.2.: Analysis phase space selection based on used triggers.

Selected electrons and muons must pass the *Medium* identification quality and *Gradient* isolation requirements.
After trigger and lepton selection, exactly one hadronic τ candidate with $p_T > 30$ GeV

and *Medium* identification quality is required. In the $\tau_e\tau_{had}$ channel, the BDT-based electron veto is further imposed to 1-prong τ candidates in addition to the likelihood-based electron veto. This veto is necessary to suppress $Z \to ee$ events. Events are selected if exactly one lepton and one hadronic τ candidate with opposite charge are identified. Moreover, at least one jet with $p_T > 40$ GeV is required to suppress the irreducible Drell-Yan Z/γ^* background; the transverse mass as a function of the lepton and E_T^{miss} (m_T) is required to be less than 70 GeV to reduce the W+jets events in the signal region. A *b jet*-veto for all jets with $p_T > 20$ GeV is applied to suppress the background originating from $t\bar{t}$ production and electroweak single top quark production. Additional cuts are introduced to further purify the signal event selection. These cuts consist of topological selections between two τ candidates and correlation between the direction of the E_T^{miss} and the taus. Angular cuts are required on the two τ candidates both in the difference in pseudo-rapidity $|\Delta\eta| \leq 1.5$ and the opening angle $\Delta R \leq 2.5$. At this stage, also a cut on $E_T^{miss} > 20$ GeV is applied to reject background with no E_T^{miss} in the final state (mainly $Z \to ll$, with $l = e, \mu$). Fractions of τ lepton momenta carried by visible decay products are calculated using the collinear approximation (Section 5.1), assuming that the neutrinos have the same direction as the original τ leptons. These fractions must be within a range $0.1 \leq x_1 \leq 1.4$ and $0.1 \leq x_2 \leq 1.2$, where the index 1 and 2 in x represent the electron (muon) and the τ candidate. This condition helps to reduce the contamination of the irreducible backgrounds where E_T^{miss} is not aligned with the direction of the leptons.
Events that fail the MMC mass reconstruction are removed; in general a very high efficiency ($\simeq 99\%$) is achieved for signal and $Z \to \tau\tau$ events, while a lower efficiency is observed for background samples, providing a good handle to reject them. The reason why MMC rejects a small fraction of events is related to the fact that true values of the scanned variables can be outside the scan range or that E_T^{miss} fluctuates outside its resolution and be outside the scan range. Fig. 6.3 and 6.4 and Table 6.3 show the modelling of the variables used to define analysis Preselection stage and yields of signal and background processes at this stage. Good Data/MC agreement and good modelling has been observed for all the considered variables. Background has been estimated following the description in Section 6.4.

Sample/Region	$\tau_e\tau_{had}$ Presel	$\tau_\mu\tau_{had}$ Presel
Fake	2819.5 ± 27.3	2478.2 ± 23.8
VV	169.8 ± 3.8	177.9 ± 4.2
Top	143.4 ± 7.0	144.1 ± 7.1
Zll	307.8 ± 88.2	437.7 ± 65.2
Ztt	7099.0 ± 77.6	8171.1 ± 88.3
Total Bkg	10539.5 ± 120.9	11409.0 ± 112.6
ggH	103.7 ± 1.3	109.7 ± 1.3
VBF	33.6 ± 0.3	36.1 ± 0.3
VH	9.31 ± 0.26	9.79 ± 0.26
Total Sgn	146.6 ± 1.3	155.6 ± 1.4
Data	10840.0 ± 104.1	11996.0 ± 109.5

Table 6.3.: Yields at Preselection level.

6.4. Background estimation

The main backgrounds for the $H \to \tau\tau \to \tau_{lep}\tau_{had}$ final state are $Z \to \tau\tau$, $Z \to ll$, W+jets, Top and Dibosons. Backgrounds can be classified in three categories:

- events with true electron or muon and τ_{had} signatures, like Z/γ^*, VV, $t\bar{t}$, single-top;
- events where a jet fakes a τ_{had} signature, like QCD and W+jets;
- events where a light charged lepton fakes a τ_{had} signature, like $Z \to ll$.

To validate and control the backgrounds, several control regions (CRs) are defined by inverting the requirements used to remove these backgrounds from the signal region (SRs), as shown in Table 6.4.
The following regions have been defined to study the most important sources of background:

- the W+jets control region is the same as the signal region, with an inverted cut on the transverse mass so that $m_T > 70$ GeV;
- the Top control region takes the same selection as the signal region, but with the b-veto requirement inverted to require at least one b-tagged jet and a requirement of $m_T > 40$ GeV. The inverted m_T requirement is needed to achieve a purer region by rejecting $Z \to \tau\tau$ contamination;
- the QCD control region is defined by inverting the lepton isolation. This region is enriched of events where one jet can fake a hadronic τ and another jet fakes an electron or muon.

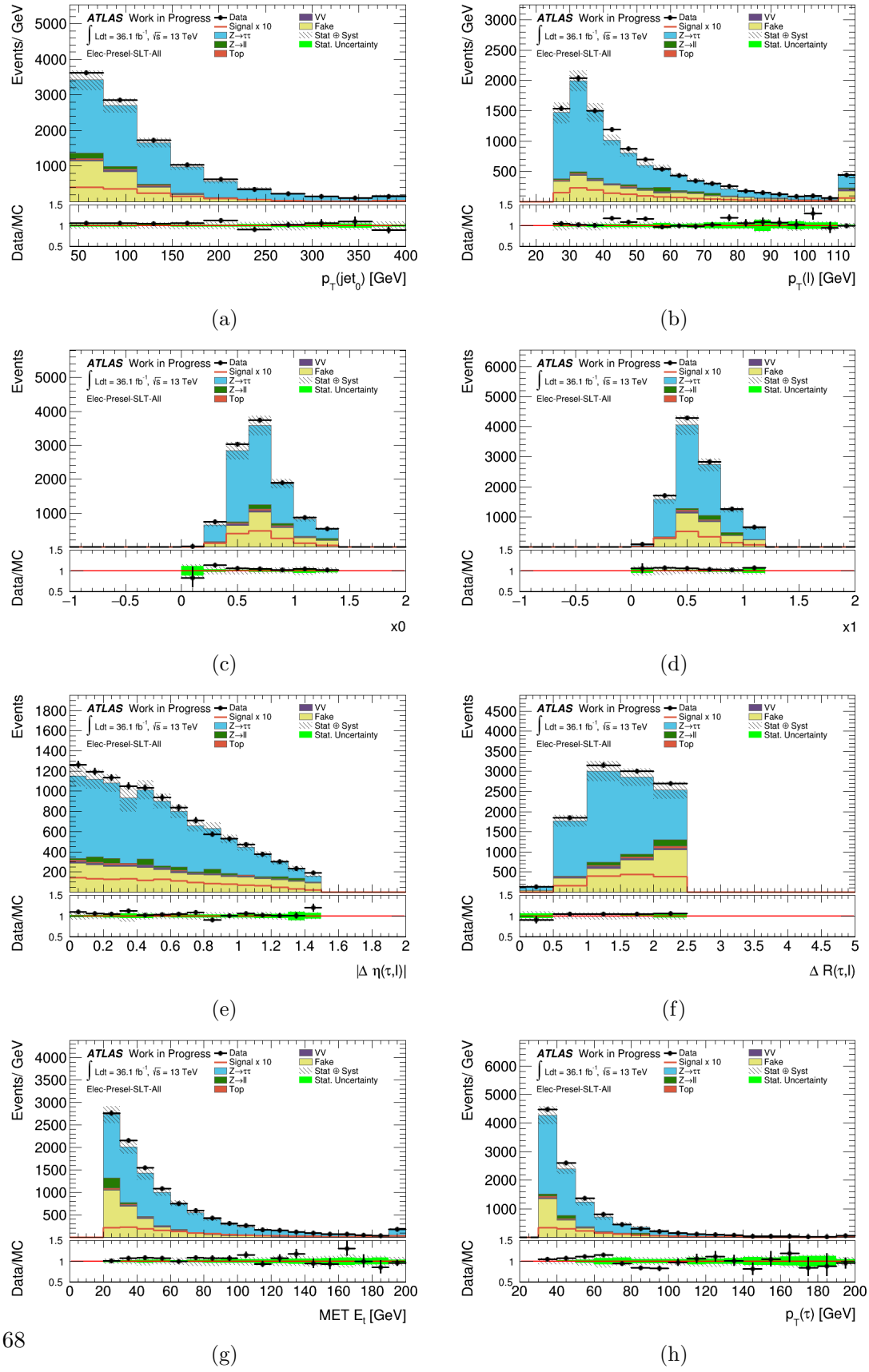
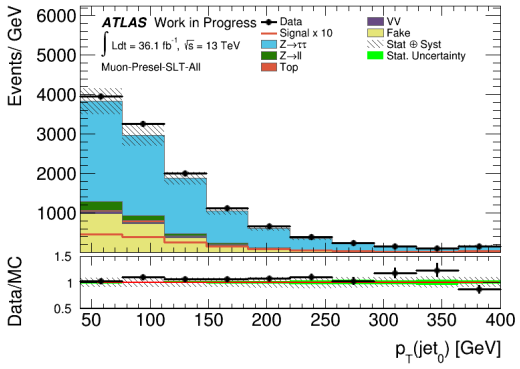
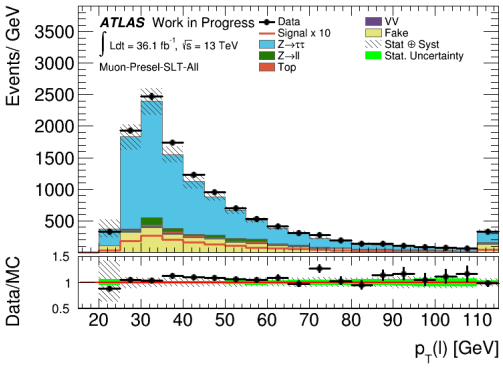


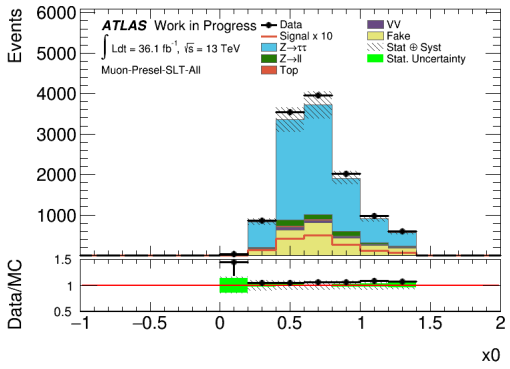
Figure 6.3.: Distributions at Preselection for $\tau_e\tau_{had}$ channel : *a)* lead jet p_T, *b)* elec. p_T, *c)* coll. approx. x_0, *d)* coll. approx. x_1, *e)* $\Delta\eta(lep,\tau)$, *f)* $\Delta R(lep,\tau)$, *g)* E_T^{miss}, *h)* τ p_T.



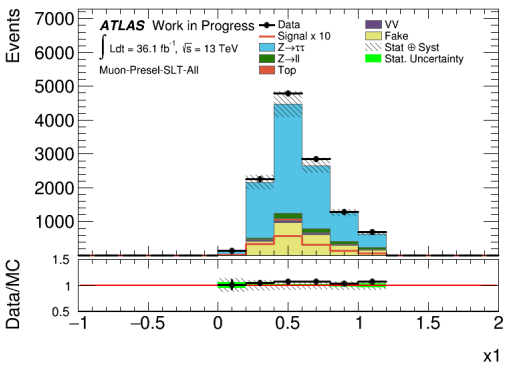
(a)



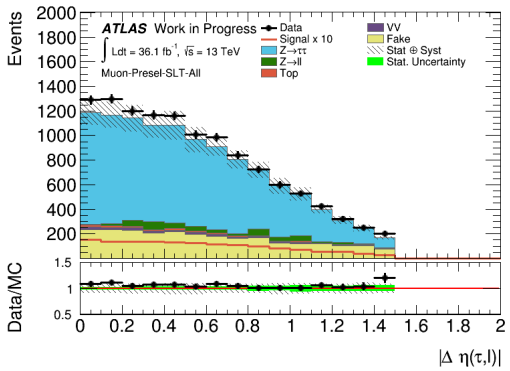
(b)



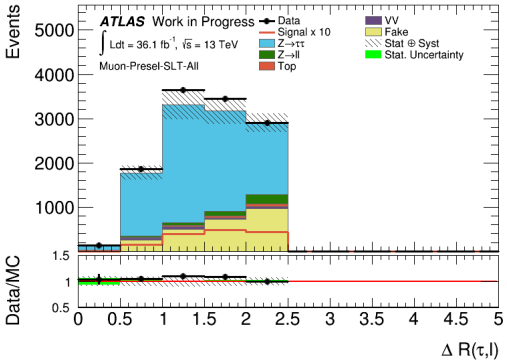
(c)



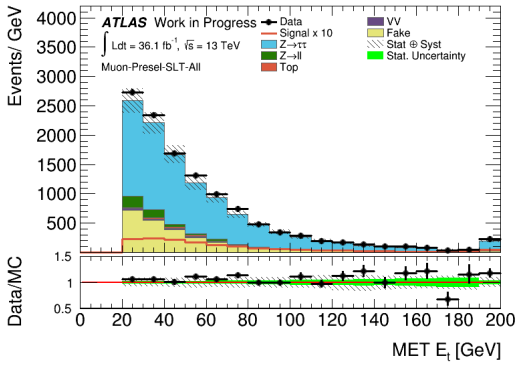
(d)



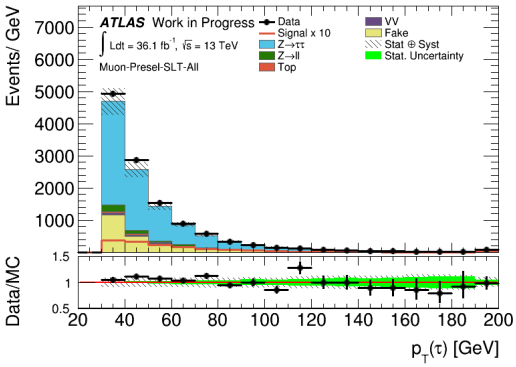
(e)



(f)



(g)



(h)

Figure 6.4.: Distributions at Preselection for $\tau_\mu\tau_{had}$ channel : *a)* lead jet p_T, *b)* muon p_T, *c)* coll. approx. x_0, *d)* coll. approx. x_1, *e)* $\Delta\eta(lep,\tau)$, *f)* $\Delta R(lep,\tau)$, *g)* E_T^{miss}, *h)* τ p_T.

Sample	CR selection = SR, except
Top CR	Invert b-veto and $m_T > 40$ GeV
W + jets CR	Invert m_T ($m_T > 70$ GeV)
QCD CR	Invert lepton isolation

Table 6.4.: Definition of the control regions.

The estimation of the jets faking taus in the signal region is done using the fake-factor method, while the jet faking leptons are estimated by the MC simulation since the fraction of this type of background is very small. Distributions of the main analysis variables for the different CRs at Preselection level are shown in Fig. 6.5-6.10, while yields are reported in Table 6.5. Also for the CRs, a good Data/MC agreement and a good modelling has been observed for all the considered variables.

Sample/Region	$\tau_e\tau_{had}$ QCD CR	$\tau_\mu\tau_{had}$ QCD CR	$\tau_e\tau_{had}$ Top CR	$\tau_\mu\tau_{had}$ Top CR	$\tau_e\tau_{had}$ W CR	$\tau_\mu\tau_{had}$ W CR
Fake	948.3 ± 12.2	709.1 ± 10.9	1165.8 ± 21.1	1165.4 ± 15.9	1278.5 ± 16.1	1096.9 ± 14.5
VV	4.0 ± 0.9	7.2 ± 0.8	43.7 ± 2.4	37.5 ± 2.0	84.8 ± 2.7	80.8 ± 2.8
Top	11.8 ± 1.9	12.2 ± 2.0	2027.1 ± 29.8	1930.4 ± 29.5	106.4 ± 6.7	97.3 ± 5.5
Zll	16.8 ± 16.3	11.1 ± 18.8	56.4 ± 28.2	76.6 ± 18.2	114.8 ± 46.2	87.9 ± 25.3
Ztt	262.0 ± 14.8	427.0 ± 20.6	281.6 ± 17.8	325.5 ± 27.1	138.3 ± 9.7	145.7 ± 9.8
Total Bkg	1242.9 ± 25.3	1166.5 ± 30.0	3574.6 ± 49.5	3535.4 ± 46.8	1722.7 ± 50.4	1508.7 ± 31.4
ggH	3.8 ± 0.2	5.8 ± 0.3	6.0 ± 0.4	5.1 ± 0.4	5.3 ± 0.3	4.9 ± 0.3
VBF	1.1 ± 0.1	1.9 ± 0.1	1.4 ± 0.1	1.4 ± 0.1	1.5 ± 0.1	1.5 ± 0.1
VH	0.4 ± 0.1	0.5 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.0 ± 0.1	0.8 ± 0.1
Total Sgn	5.3 ± 0.3	8.3 ± 0.3	8.7 ± 0.4	7.7 ± 0.4	7.8 ± 0.3	7.2 ± 0.3
Data	1211.0 ± 34.8	1212.0 ± 34.8	3622.0 ± 60.2	3569.0 ± 59.7	1739.0 ± 41.7	1561.0 ± 39.5

Table 6.5.: Yields in the *QCD*/*Top* and *W* CRs at Preselection level.

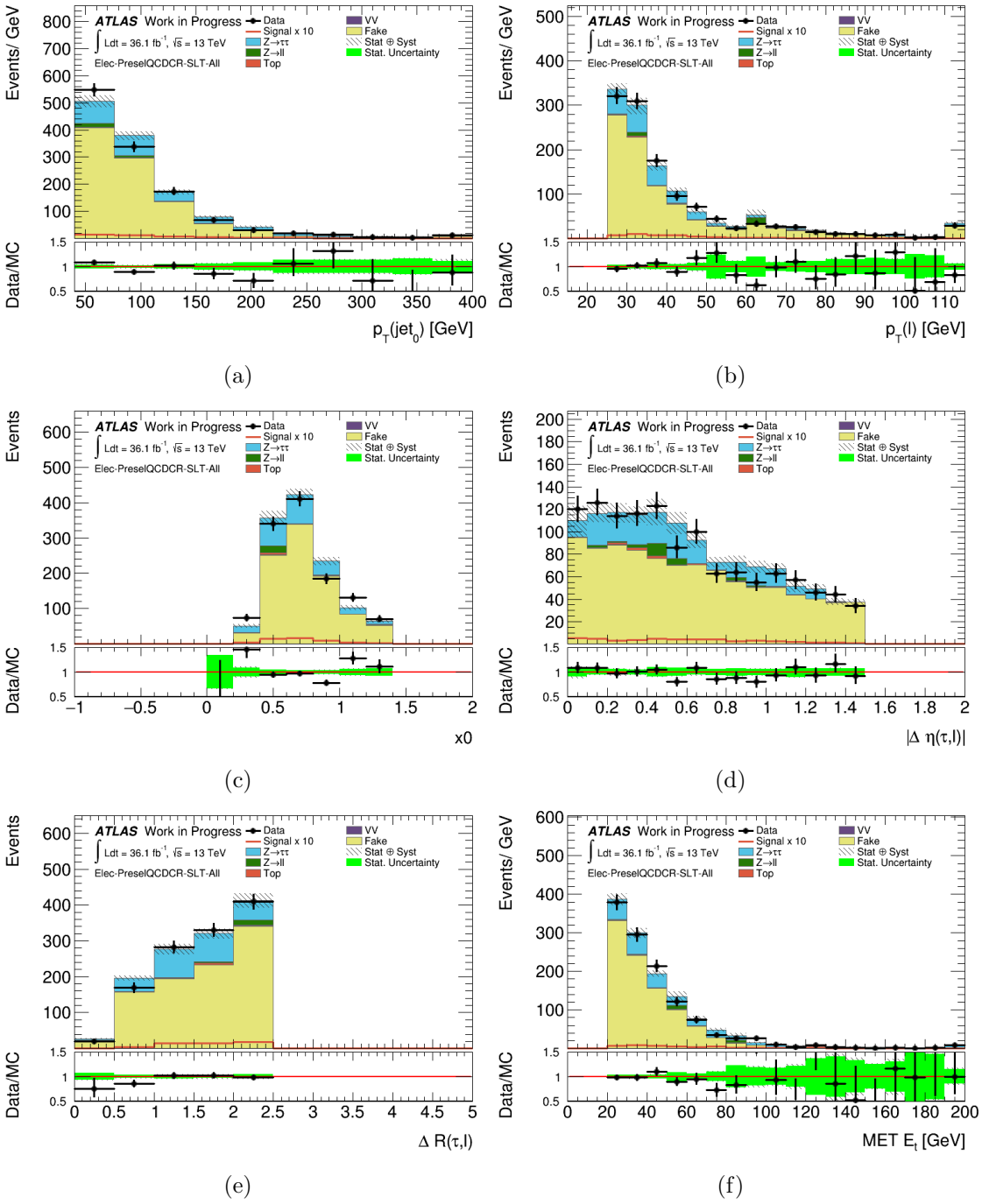
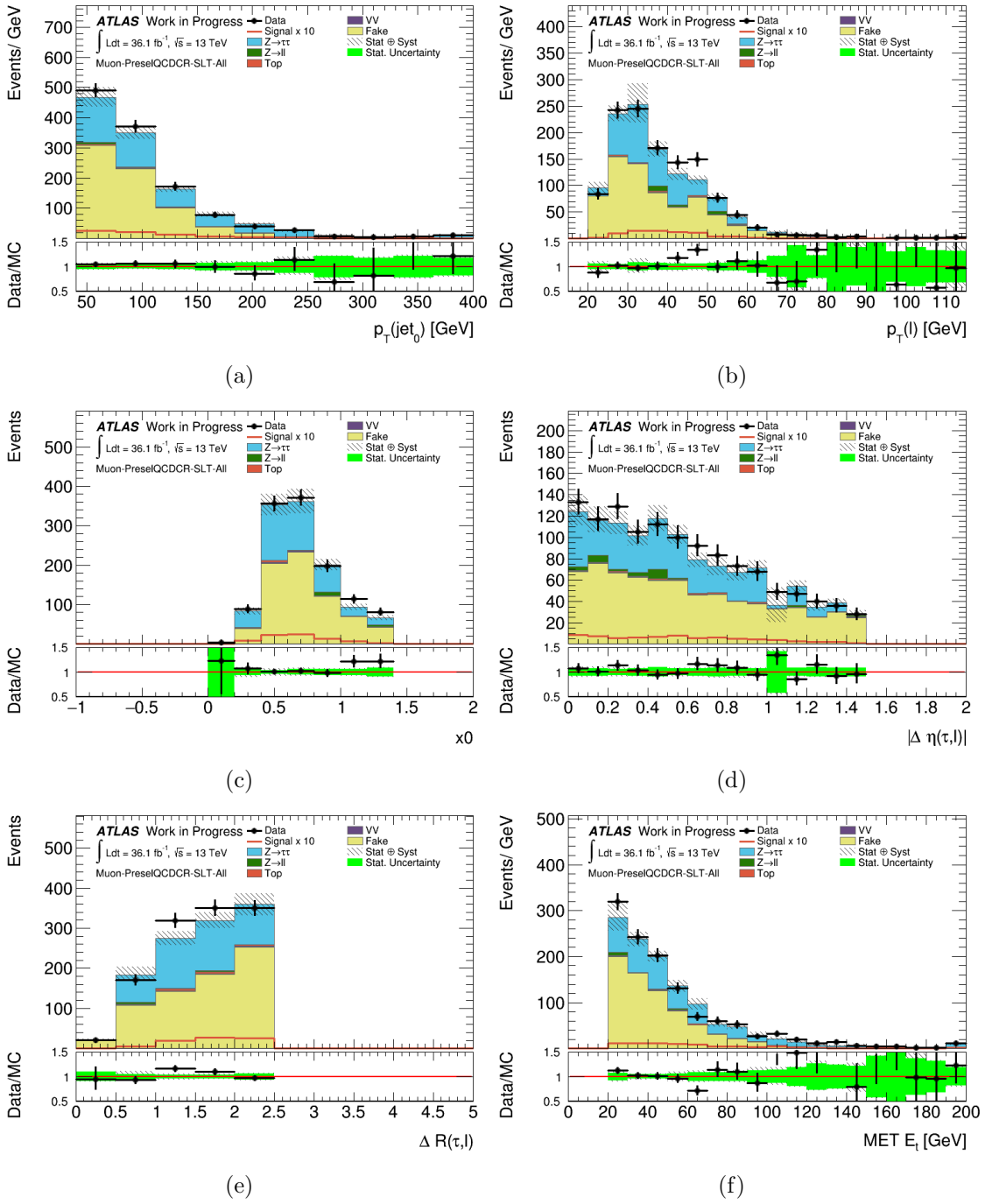


Figure 6.5.: Distributions in QCD CR at Preselection level for $\tau_e\tau_{had}$ channel : *a)* lead jet p_T, *b)* elec. p_T, *c)* coll. approx. x_0, *d)* $\Delta\eta(lep,\tau)$, *e)* $\Delta R(lep,\tau)$, *f)* E_T^{miss}.



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

Figure 6.6.: Distributions in QCD CR at Preselection level for $\tau_\mu\tau_{had}$ channel : *a)* lead jet p_T, *b)* muon p_T, *c)* coll. approx. x_0, *d)* $\Delta\eta(lep,\tau)$, *e)* $\Delta R(lep,\tau)$, *f)* E_T^{miss}.

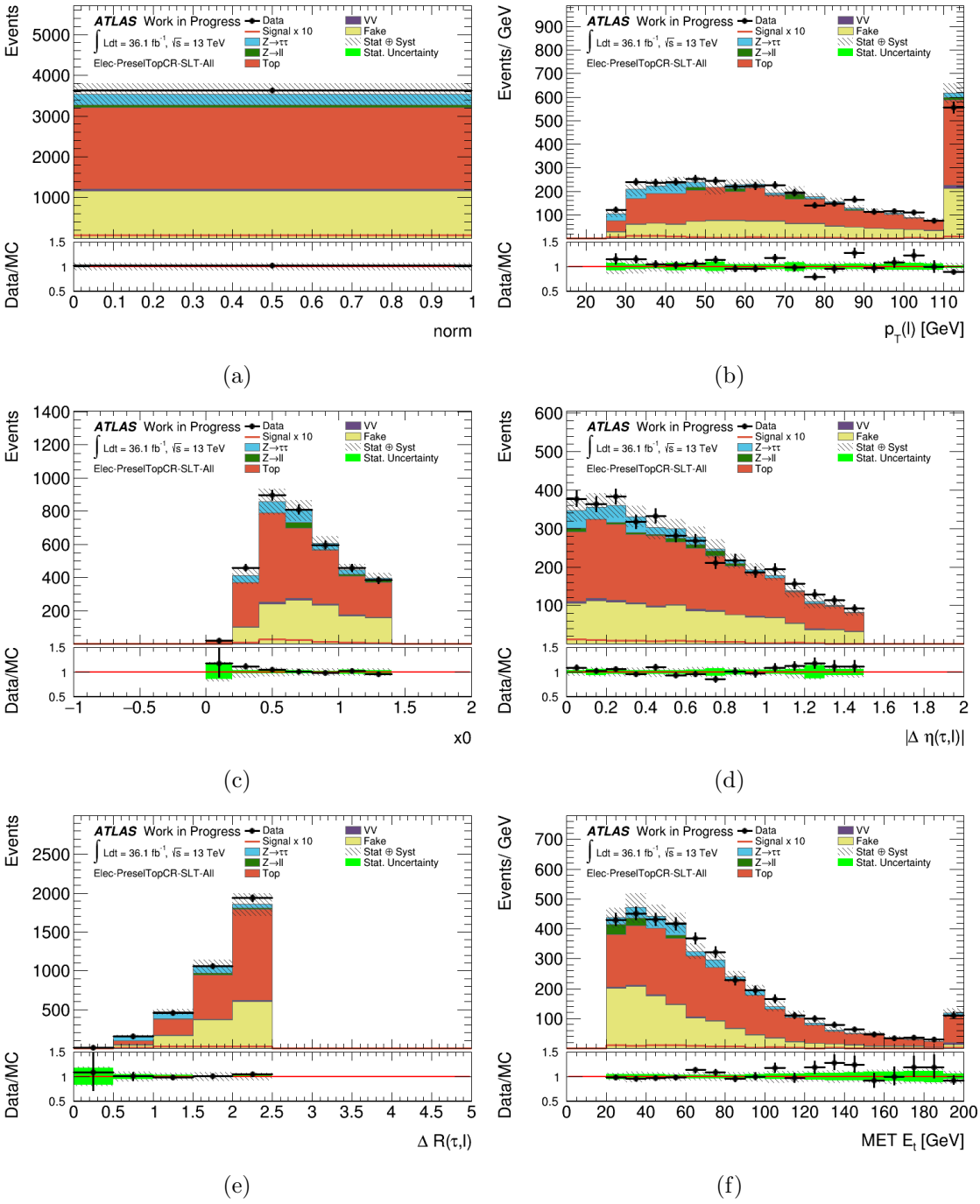
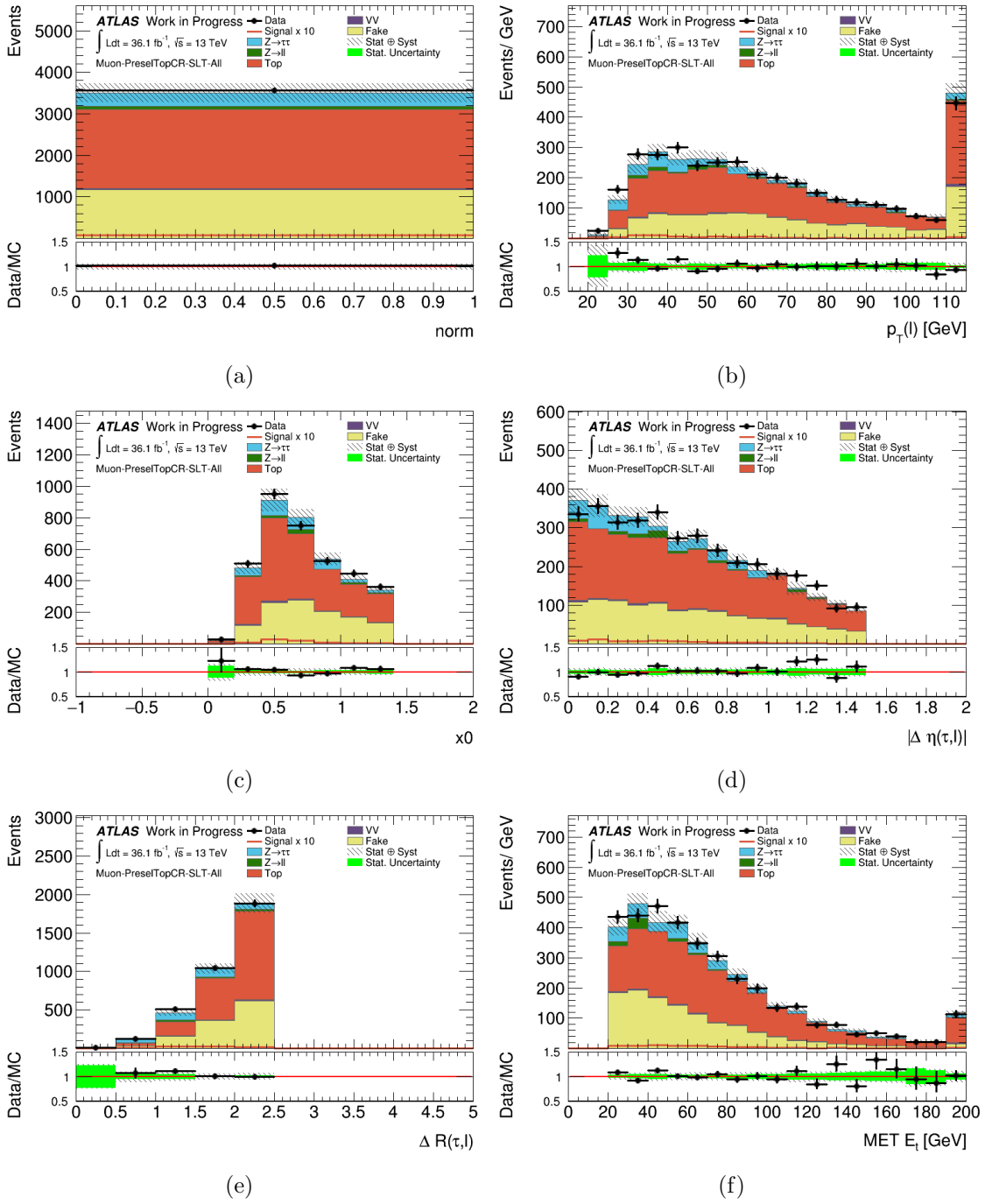


Figure 6.7.: Distributions in Top CR at Preselection level for $\tau_e\tau_{had}$ channel : *a)* norm, *b)* elec. p_T, *c)* coll. approx. x_0, *d)* $\Delta\eta(lep,\tau)$, *e)* $\Delta R(lep,\tau)$, *f)* E_T^{miss}.



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

Figure 6.8.: Distributions in Top CR at Preselection level for $\tau_\mu\tau_{had}$ channel : *a)* norm, *b)* muon p_T, *c)* coll. approx. x_0, *d)* $\Delta\eta(lep,\tau)$, *e)* $\Delta R(lep,\tau)$, *f)* E_T^{miss}.

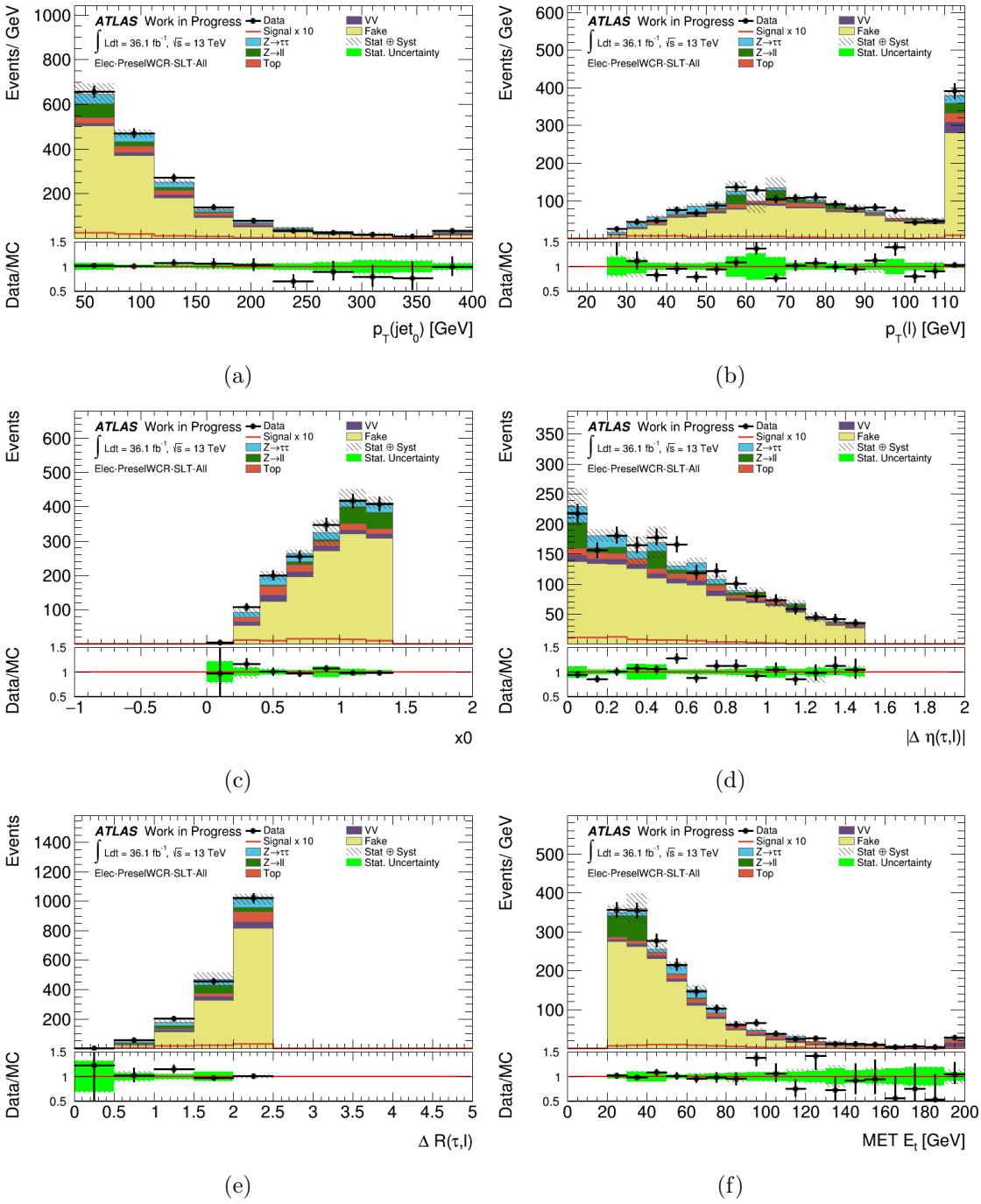
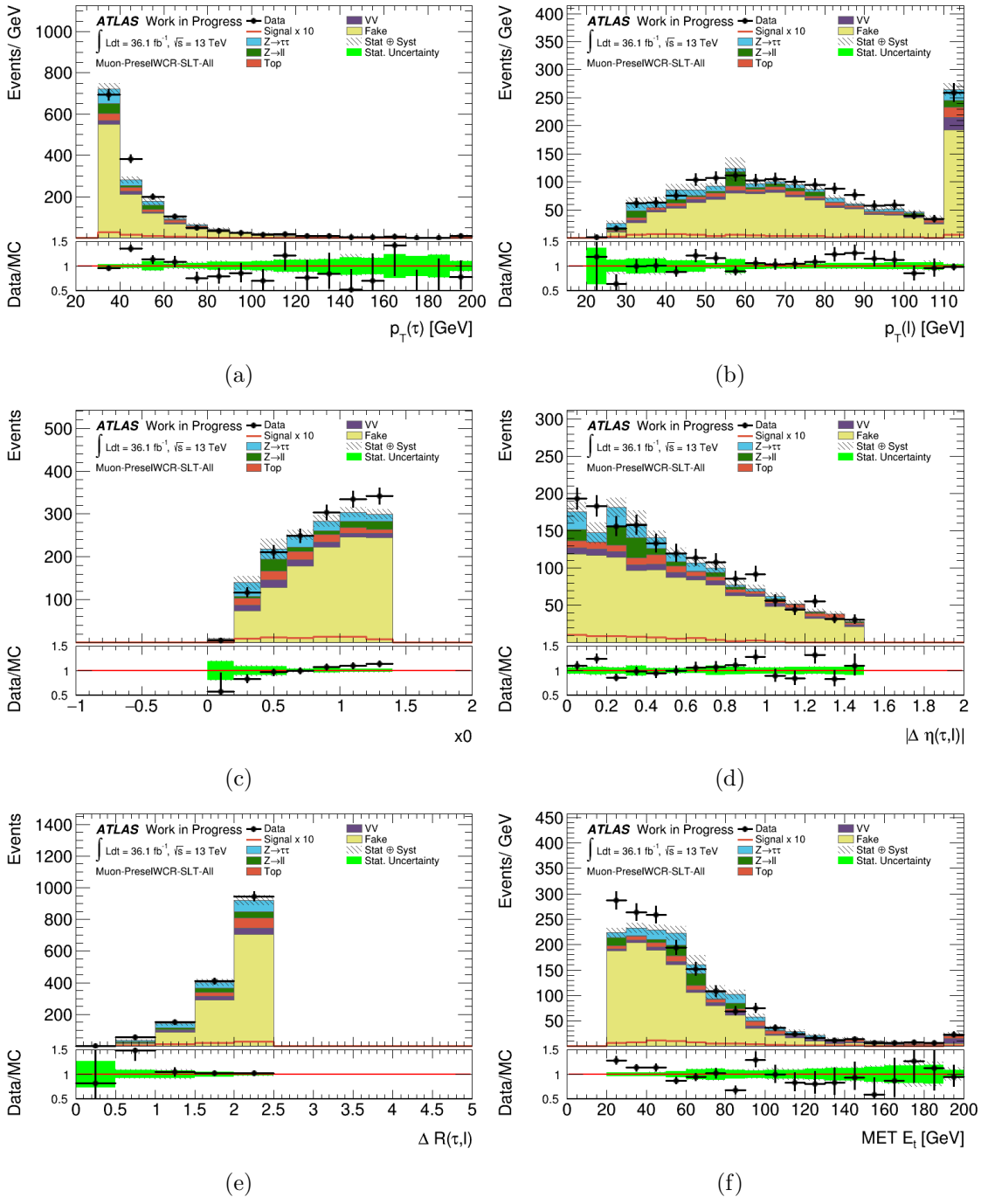


Figure 6.9.: Distributions in W CR at Preselection level for $\tau_e\tau_{had}$ channel : *a)* lead jet p_T, *b)* elec. p_T, *c)* coll. approx. x_0, *d)* $\Delta\eta(lep,\tau)$, *e)* $\Delta R(lep,\tau)$, *f)* E_T^{miss}.



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

Figure 6.10.: Distributions in W CR at Preselection level for $\tau_\mu\tau_{had}$ channel : *a)* τ p_T, *b)* muon p_T, *c)* coll. approx. x_0, *d)* $\Delta\eta(lep,\tau)$, *e)* $\Delta R(lep,\tau)$, *f)* E_T^{miss}. The mis-modelling visible in the low E_T^{miss} region can be addressed to lack of $Z \to \mu\mu$ MC and to Sherpa generator negative event weights.

6.4.1. Fake Factor method

The background from jets mis-identified as hadronically decaying τ leptons is a dominant background for the $\tau_{lep}\tau_{had}$ final state. It consists mostly of W+jets events, but fake τ candidates also come from multi-jet production and $t\bar{t}$. The method which is used to determine this background is a data-driven method called *Fake Factor* which is described in the next paragraph.

Method description

For the purposes of this method, an $anti-\tau$ region is defined taking a τ candidate which passes all analysis requirements except for failing the *medium* τ identification requirement. A lower threshold on the τ identification score of 0.25 is applied, since candidates with very low score are dominated by gluon-induced jets and jets arising from pile-up, while in the signal regions jets are mainly coming from quarks. Since only one τ candidate is selected for each event, it is possible to construct, for each signal region, a corresponding $anti-\tau$ control region, containing the events passing exactly the full selection except that the τ candidate is an $anti-\tau$.
The estimate of the fake background, both shape and normalisation, in each signal region can then be determined by using the data events in the corresponding $anti-\tau$ region and multiplying them with a transfer factor, called *combined fake-factor* ($\mathcal{F}$), to correct for the different selection efficiency between $pass-\tau$ and $anti-\tau$ (Fig. 6.11). Events in the $anti-\tau$ region not corresponding to fakes from jets are subtracted using simulated event samples:

$$N_{fakes}^{SR} = (N_{Data}^{anti-\tau} - N_{MC,\ no\ jet\to\tau}^{anti-\tau}) \times \mathcal{F} \tag{6.1}$$

The combined fake-factor for each signal region is binned in τ p_T and number of τ tracks.

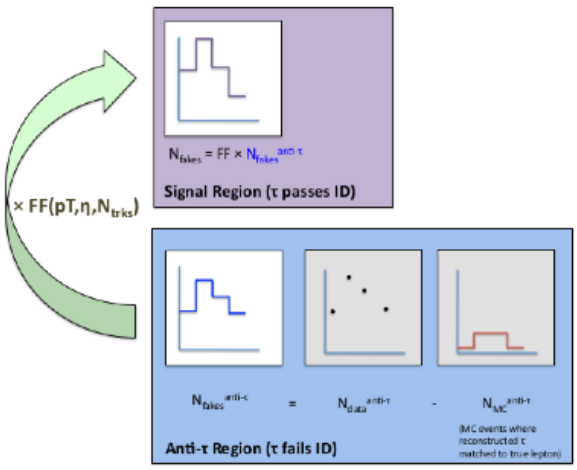


Figure 6.11.: Fake factor method application.

Fake-factors depend on the quark/gluon composition of a given sample and are therefore different for each possible background source. The combined fake-factor should therefore in principle be constructed as the sum of the individual fake-factors for each relevant process, weighted by the expected fractional contribution for the considered process in the anti-τ region:

$$\mathcal{F} = R_W F_W + R_{Top} F_{Top} + R_{QCD} F_{QCD} \tag{6.2}$$

However, the small background from Top-quark production does not play an important role and can be expected to have reasonably similar fake-factors to the W+jets background, with a preponderance of quark-initiated jets. This can be justified by checking the MC predictions at various analysis levels, which are usually less than 2%. Therefore, it is assumed that all the processes except multi-jet production can be described using the fake-factors derived for W+jet events. Thus the combined fake-factor is given by:

$$\mathcal{F} = R_{QCD} F_{QCD} + R_W F_W \tag{6.3}$$

The fraction of multi-jet events in each region, R_{QCD}, is obtained from data as described below. The fractional contribution from W+jets production is given by $R_W = 1 - R_{QCD}$. The individual fake-factors F_i, with $i = W, QCD$, are obtained in the dedicated W and QCD control regions as the ratio of data events in the $pass - \tau$ events over those in the $anti - \tau$. Contributions from events where the τ is not faked by a jet are subtracted from data yields using MC simulation:

$$F_i = \frac{N_{Data}^{pass,CR_i} - N_{MC,\ no\ jet\to\tau}^{pass,CR_i}}{N_{Data}^{fail,CR_i} - N_{MC,\ no\ jet\to\tau}^{fail,CR_i}} \tag{6.4}$$

The fraction of QCD multi-jet events in each SR $anti - \tau$ region is given by:

$$R_{QCD} = \frac{N_{QCD,Data}^{anti-\tau}}{N_{SR,Data}^{anti-\tau} - N_{SR,MC\ no\ jet\to\tau}^{anti-\tau}} \tag{6.5}$$

The number of QCD events in the $anti-\tau$ region, $N_{QCD,Data}^{anti-\tau}$, is estimated from data by multiplying the events in the QCD $anti - \tau$ CR with a transfer factor, called *Isolation Factor* ($\mathcal{I}$), accounting for the difference between failing and passing the lepton isolation. Events with a true lepton in the QCD $anti - \tau$ CR are subtracted using MC :

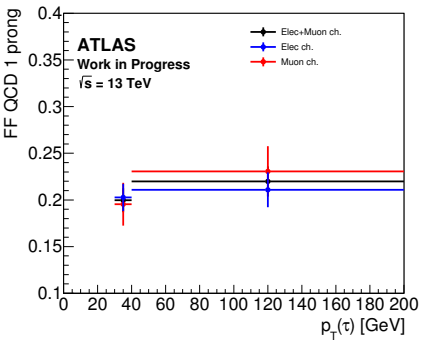
$$N_{QCD,Data}^{anti-\tau} = (N_{Data}^{QCDCR,anti-\tau} - N_{MC,true\ lepton}^{QCDCR,anti-\tau}) \times \mathcal{I} \tag{6.6}$$

These isolation factors are calculated (separately for electron and muon final state) by taking the ratio of the events passing the lepton isolation requirement over those that fail it in a dedicated control region. This control region is defined using exactly the same cuts as for the *Preselection* stage with the only modification that the charge sign of the τ candidate and the lepton is required to be the same (*SSP* region). This region definition ensures the orthogonality between the regions where the isolation factors are calculated and then subsequently applied. It is also assumed that there is no significant

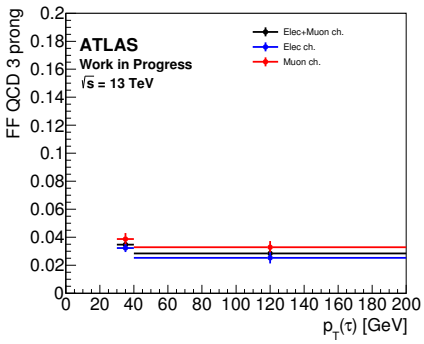
difference in the isolation factors between the Preselection stage and the signal regions where they are applied. Events with true leptons in the SSP are subtracted using MC simulation:

$$\mathcal{I} = \frac{N_{data}^{iso,SSP} - N_{MC,true\ lepton}^{iso,SSP}}{N_{data}^{non-iso,SSP} - N_{MC,true\ lepton}^{non-iso,SSP}} \tag{6.7}$$

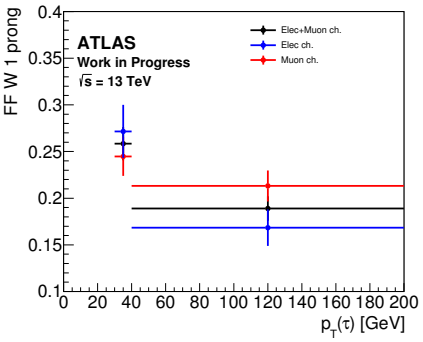
R_{QCD} can then be calculated in each region separately for electron and muon final states, for 1- and 3-prong τ candidates, and it is binned according to the lepton and the azimuthal angle between the τ candidate and the missing transverse momentum, $\Delta\phi(\tau, E_T^{miss})$, which was found to be the most sensitive among a large number of considered variables for this purpose. The binning has been optimised according to the available statistics.
The same procedure previously described has been followed also to obtain the fake estimate not only in the signal region but also in the other regions used in the analysis (such as the Top control region which is used in the fit). However, in this case it was assumed that $R_{QCD} = 0$ for the Top(W)-CR, while in QCD CR it was assumed $R_{QCD} = 1$.
In Fig. 6.12 the values of F_{QCD}, F_W and the Isolation factors at Preselection level for 1/3 prong τ selection are shown.



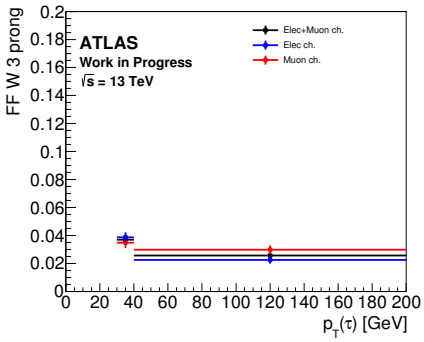
(a)



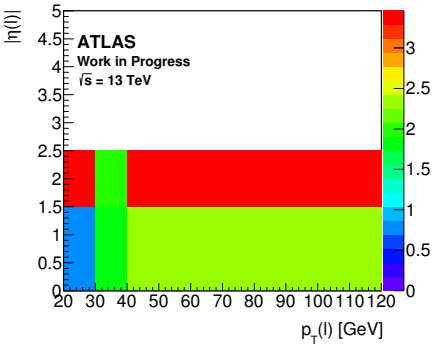
(b)



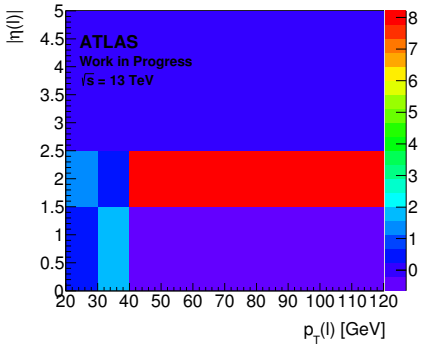
(c)



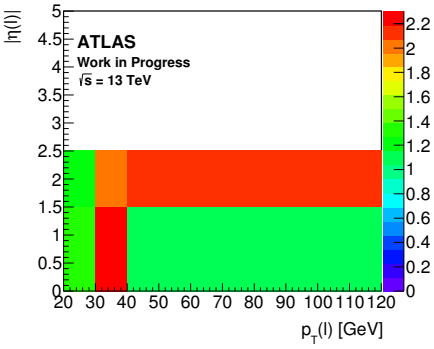
(d)



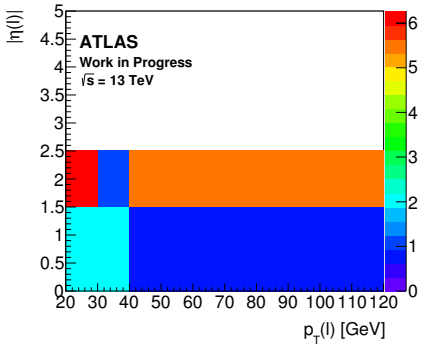
(e)



(f)



(g)



(h)

Figure 6.12.: Fake factors values for: *a) (b)* F_{QCD} 1 (3) prong, *c) (d)* F_W 1 (3) prong. Isolation factors for: *e) (f)* $\tau_e\tau_{had}$ 1 (3) prong, *g) (h)* $\tau_\mu\tau_{had}$ channel 1 (3) prong.

Assumptions and closure tests

The following assumptions to use fake-factor method has been made and verified:

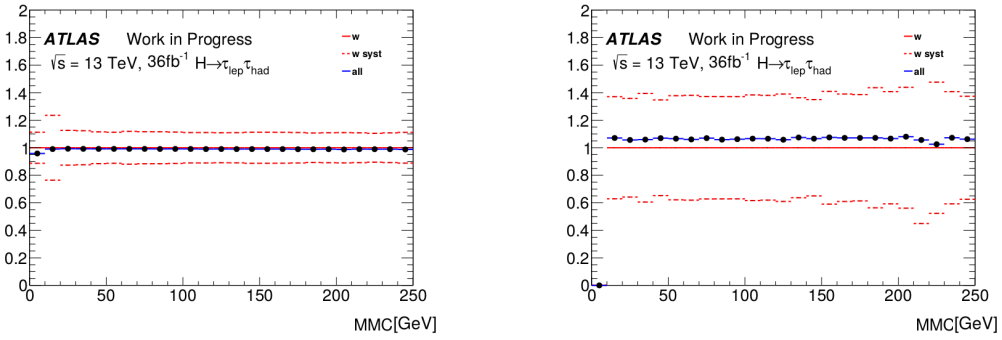
- *all background processes except QCD can be described using the W fake-factors.* In Fig. 6.13 a comparison of the final fake estimate made using this assumption (and all the related uncertainties) to a fake estimate where all processes have been included separately is shown; as can be observed, the difference between the two calculations is negligible compared to the uncertainties; thus no further uncertainties have been assigned;

- *the 2015 and 2016 datasets can be combined.* In Fig. 6.14 a comparison of the final fake estimate using this assumption (and all the related uncertainties) to a fake estimate where the fake-factors have been calculated and applied separately for the 2015 and 2016 datasets is show; as can be observed, the difference between the two calculations is negligible compared to the uncertainties; thus no further uncertainties have been assigned;

- *fake-factors calculated in the different CRs can be used in the SR.* This assumption cannot be directly tested; however, a same-sign closure test has been performed to show that the fake-factor method performs as expected. In this closure test, the full procedure was repeated using the same setup and analysis selection, except that the opposite sign requirement for the τ decay products was changed to a same-sign requirement; this enhances the fake background component and suppresses other types of background. MMC distributions in these same sign regions are shown in Fig. 6.15 for the main analysis categories. In general closure tests show good agreement between data and predictions, except for a visible non-closure effect in the case of 1-prong $\tau_e\tau_{had}$ events in the boosted category which is not compensated by the uncertainties described in Section 6.4.1 (differently from the VBF region in which larger systematics cover non closure effects). To further investigate the origin of this effect, many investigations have been done (like variation of the R_{QCD} calculation/application, study of other distributions to check if the mis-modelling comes from a particular phase space region, etc...) but none of them gave useful hints; therefore a Fake estimate non-closure systematic has been assigned as further uncertainty and propagated in the final fit. The impact of this systematic on the expected significance in a standalone $H \to \tau\tau \to \tau_{lep}\tau_{had}$ final state was found to be at the 1 % level and of course it would be even less when the fit will be performed adding the other decay channels (full leptonic and full hadronic final state). Therefore it is possible to conclude that this non-closure systematic which is applied only on one sub-channel will not be significant for the full analysis.

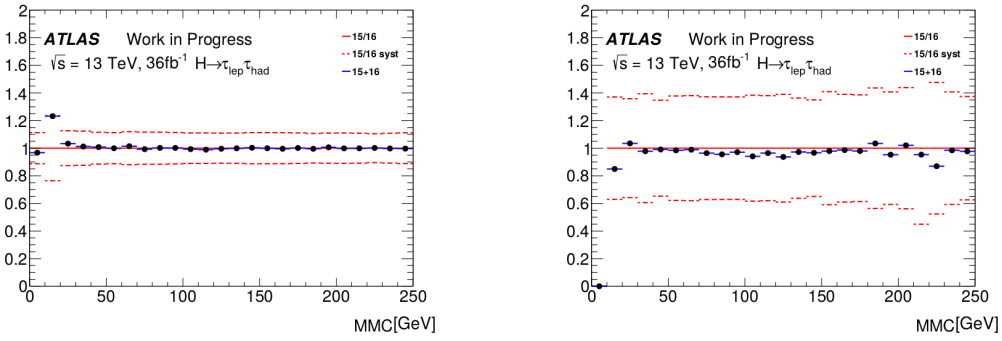
Uncertainties

Uncertainties on the background estimate using the fake-factor method arise from different sources. The following uncertainties have been considered:



(a) (b)

Figure 6.13.: The final fake estimate (with all uncertainties) made using the assumption that all background processes except QCD can be described using the W fake-factors, compared to a fake estimate where all processes were separately included *a)* Boosted region; *b)*: VBF region.



(a) (b)

Figure 6.14.: The final fake estimate (with all uncertainties) made using a combined 2015 and 2016 dataset for the fake-factor calculation, compared to a fake estimate where the fake-factors were calculated and applied separately for the dataset from 2015 and from 2016. *a)* Boosted region; *b)* VBF region.

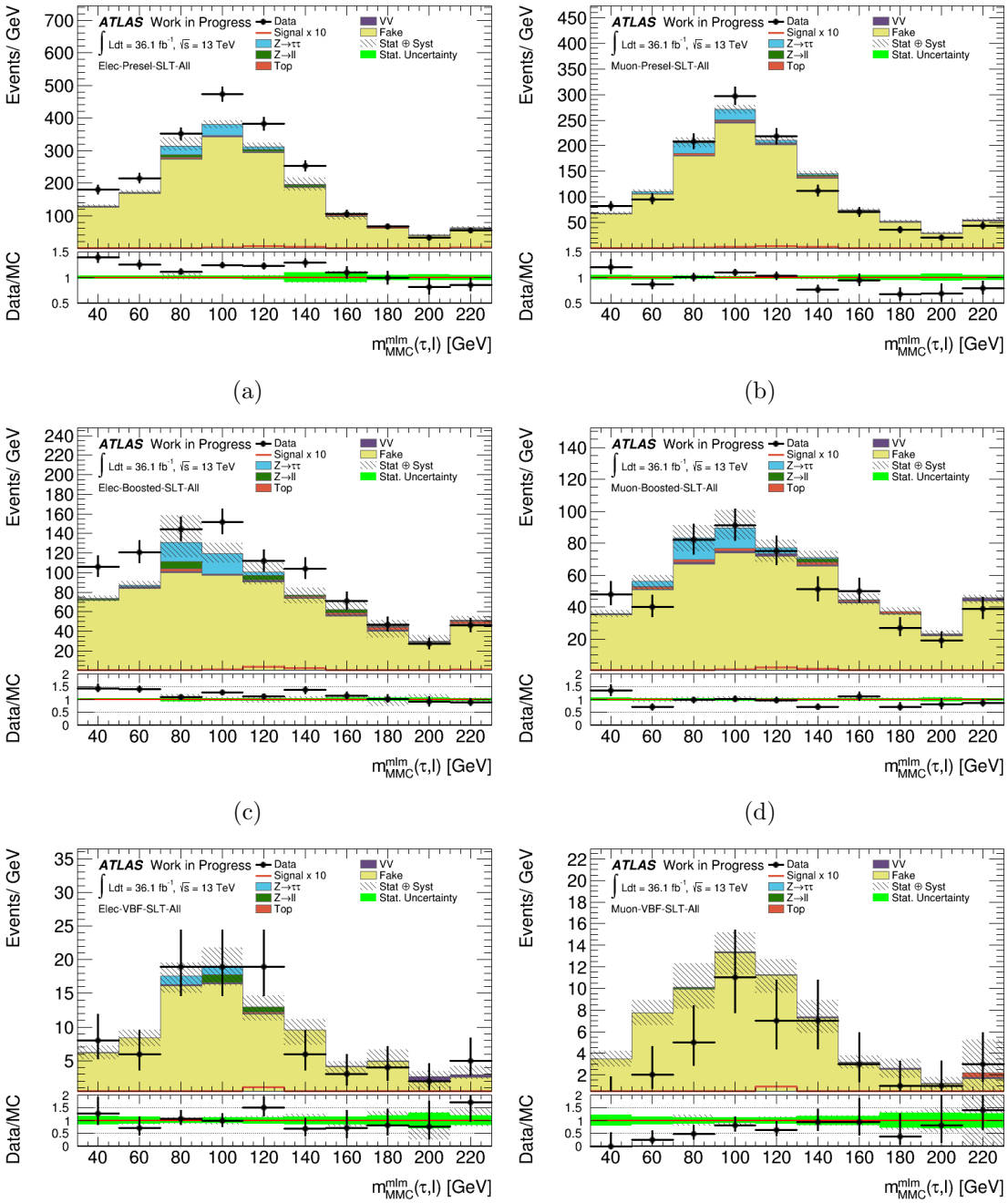


Figure 6.15.: MMC distributions in the same sign closure region: *a) (b)* $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) Preselection, *c) (d)* $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) Boosted region, *e) $\tau_e\tau_{had}$ (f)* ($\tau_\mu\tau_{had}$) VBF region. Figures do not show Fake-Factor systematic uncertainties discussed in Section 6.4.1. For the signal, $m_H = 125$ GeV is chosen.

- the uncertainty on the individual fake-factors, F_{QCD} and F_W. These uncertainties come from the limited statistics in the corresponding control regions where they are calculated. To simplify uncertainty propagation in the final fit, these uncertainties are all considered as correlated across all bins inside a given category;
- the uncertainty on the estimation of R_{QCD} from data. In this case, uncertainties arise from the measurement of the isolation factor:
  - the statistical uncertainty on the isolation factors, again arising from the limited statistics in the control regions where these are calculated;
  - uncertainty based on the difference between isolation factors calculated using same-sign and opposite sign event selection;
  - contamination of the control region with true lepton events which are subtracted using MC estimation. To estimate this effect, the subtracted MC was varied by ± 35 % (corresponding to the maximum effect of systematic uncertainties on the MC estimate).
- the non-closure systematics in the SS region for the $\tau_e\tau_{had}$ 1 prong final state;
- the statistical uncertainty in the anti-τ region where the fake-factors are applied.

The impact of these uncertainties is shown in Table 6.6 for the Boosted and VBF region; from these results, it is possible to conclude that the uncertainty on the individual fake-factors gives the larger contribution to the total error on the fake-estimate.

	VBF	Boosted
	Rel. diff. to nom. (%)	
Individual FF Stat. Unc. (up)	31	9
Individual FF Stat. Unc. (down)	-32	-8
Uncertainty on R_{QCD} (up)	4	-4
Uncertainty on R_{QCD} (down)	-3	2
SS non-closure Unc.	5	3
anti-τ region Stat. Unc. (up)	4	1
anti-τ region Stat. Unc. (down)	-4	-1

Table 6.6.: Summary of the fake-estimate relative uncertainties on the yields in the VBF and Boosted (inclusive) categories.

6.5. Signal Regions definition

The analysis target is mainly to identify the events produced according to the highest Higgs boson production cross-section mode: the gluon-fusion and the vector-boson fusion. In order to enhance the signal over the background contribution, dedicated signal regions are defined.

6.5.1. VBF region selection

The VBF selection is designed to enrich signal events from the Higgs vector boson fusion production process. The requirements are

- at least 2 jets in the event with $p_T > 30$ GeV;
- the pseudo-rapidity difference between the two leading jets must be $|\Delta\eta|_{jj} > 3.0$;
- the invariant mass of the two leading jets must be $M_{jj} > 400$ GeV;
- the product of pseudorapidities of the two jets must be negative $\eta_{j1} \times \eta_{j2} \leq 0$;
- both leptonic and hadronic tau candidates must lie between the two leading jets in pseudorapidity (*centrality*).

This region is further split into two regions to separate phase space regions with different signal-to-background ratios and thus enhance the sensitivity. The events with $M_{jj} > 500$ GeV and $p_T^H > 100$ GeV are selected in the *Tight* region, all other events go into the *Loose* region. p_T^H is the *reconstructed* Higgs boson momentum, which is given by the sum of the four-momenta of the electron/muon, the τ and the E_T^{miss}.
In Tables 6.7 and 6.8, event yields in VBF signal region and relative control regions are reported, while in Fig. 6.16-6.19 the modelling of the main variables used to the define this signal regions are shown. Finally, Fig. 6.20 shows the MMC distribution in the split VBF signal regions. In general, good modelling and Data/MC agreement has been observed, considering also the large uncertainties due to the low number of events selected in this particular phase space region. A mis-modelling has been observed in the Top CR for the $\tau_e\tau_{had}$ channel, but it should be considered that the normalisation for the Top process will be determined later by the likelihood fit, so the MC estimation which is shown in the plots is used only as input for the fit.

Sample/Region	$\tau_e\tau_{had}$ SR	$\tau_\mu\tau_{had}$ SR
Fake	91.6 ± 4.9	66.0 ± 3.3
VV	7.0 ± 1.1	7.0 ± 1.2
Top	10.6 ± 1.8	8.2 ± 1.7
Zll	11.8 ± 3.9	9.3 ± 2.5
Ztt	225.6 ± 10.2	250.1 ± 12.7
Total Bkg	346.6 ± 12.2	340.6 ± 13.6
ggH	5.4 ± 0.3	5.5 ± 0.3
VBF	15.8 ± 0.2	16.6 ± 0.2
VH	0.1 ± 0.1	0.2 ± 0.1
Total Sgn	21.3 ± 0.3	22.3 ± 0.3
Data	396.0 ± 19.9	418.0 ± 20.4

Table 6.7.: Yields in the VBF signal region.

Sample/Region	$\tau_e\tau_{had}$ QCD CR	$\tau_\mu\tau_{had}$ QCD CR	$\tau_e\tau_{had}$ Top CR	$\tau_\mu\tau_{had}$ Top CR	$\tau_e\tau_{had}$ W CR	$\tau_\mu\tau_{had}$ W CR
Fake	37.5 ± 2.3	31.6 ± 2.1	34.1 ± 2.3	30.7 ± 2.2	32.2 ± 4.8	27.6 ± 2.0
VV	0.0 ± 0.1	0.1 ± 0.1	3.0 ± 0.8	2.1 ± 0.4	3.1 ± 0.3	3.2 ± 0.6
Top	0.7 ± 0.5	0.8 ± 0.5	77.9 ± 6.4	78.7 ± 6.1	5.8 ± 1.2	6.9 ± 1.4
Zll	0.3 ± 0.3	0.3 ± 0.2	4.1 ± 1.3	15.2 ± 10.4	9.8 ± 4.8	2.0 ± 1.0
Ztt	12.8 ± 2.4	14.8 ± 2.4	6.0 ± 3.1	19.1 ± 4.4	3.9 ± 1.7	8.9 ± 2.5
Total Bkg	51.3 ± 3.3	47.6 ± 3.3	125.1 ± 7.6	145.9 ± 13.0	54.7 ± 7.1	48.6 ± 3.7
ggH	0.2 ± 0.1	0.3 ± 0.1	0.6 ± 0.1	0.5 ± 0.1	0.4 ± 0.1	0.3 ± 0.1
VBF	0.5 ± 0.0	0.8 ± 0.1	0.6 ± 0.0	0.5 ± 0.0	0.6 ± 0.0	0.6 ± 0.0
VH	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total Sgn	0.8 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.0 ± 0.1	1.1 ± 0.1	1.0 ± 0.1
Data	53.0 ± 7.3	48.0 ± 6.9	189.0 ± 13.7	135.0 ± 11.6	73.0 ± 8.5	38.0 ± 6.2

Table 6.8.: Yields in the *QCD/ Top* and *W* CRs in VBF region.

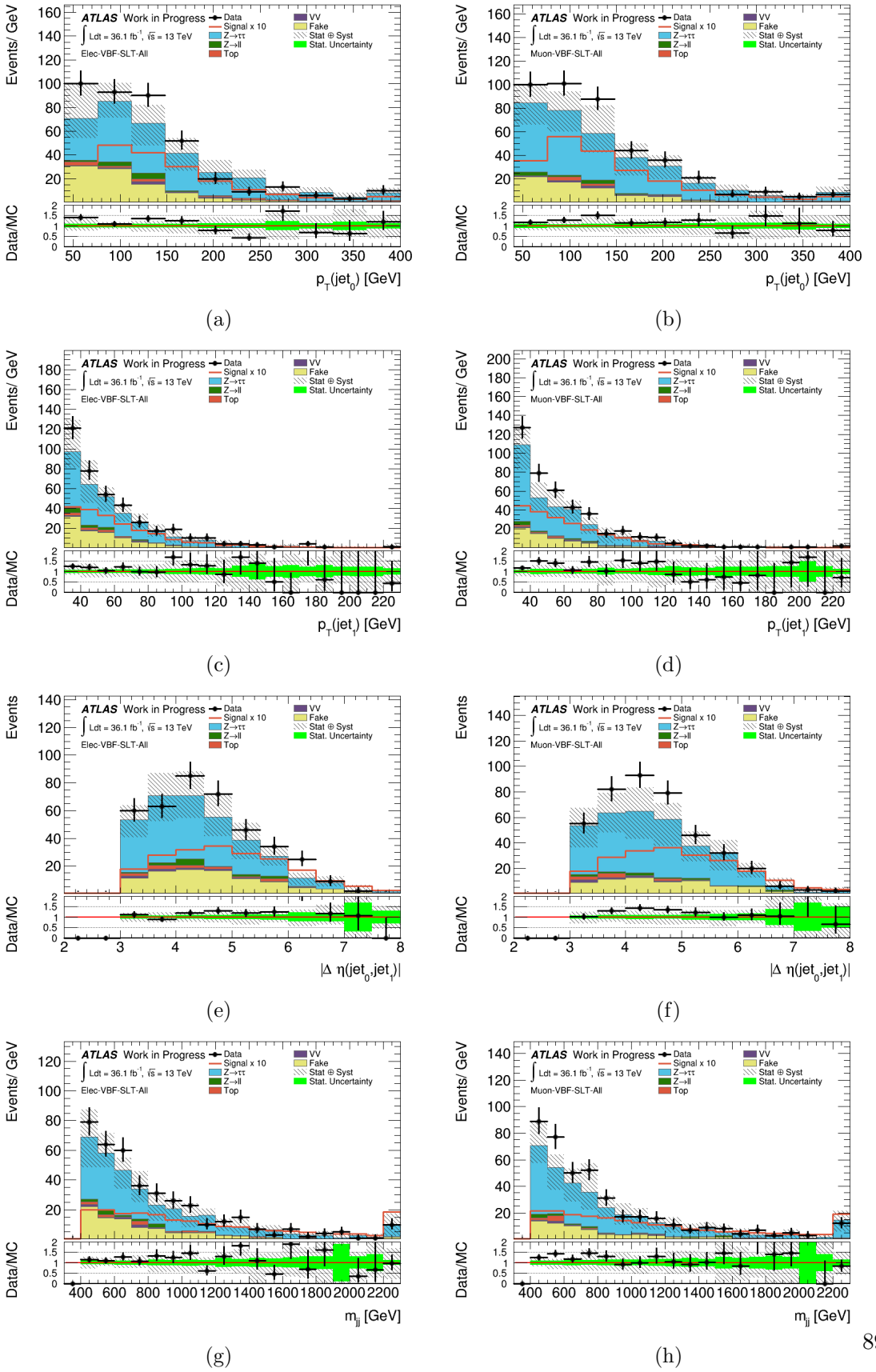


Figure 6.16.: Distributions in VBF signal region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* lead jet p_T, *c) (d)* sub-lead jet p_T, *e) (f)* $|\Delta\eta|_{jj}$, *g) (h)* M_{jj}.

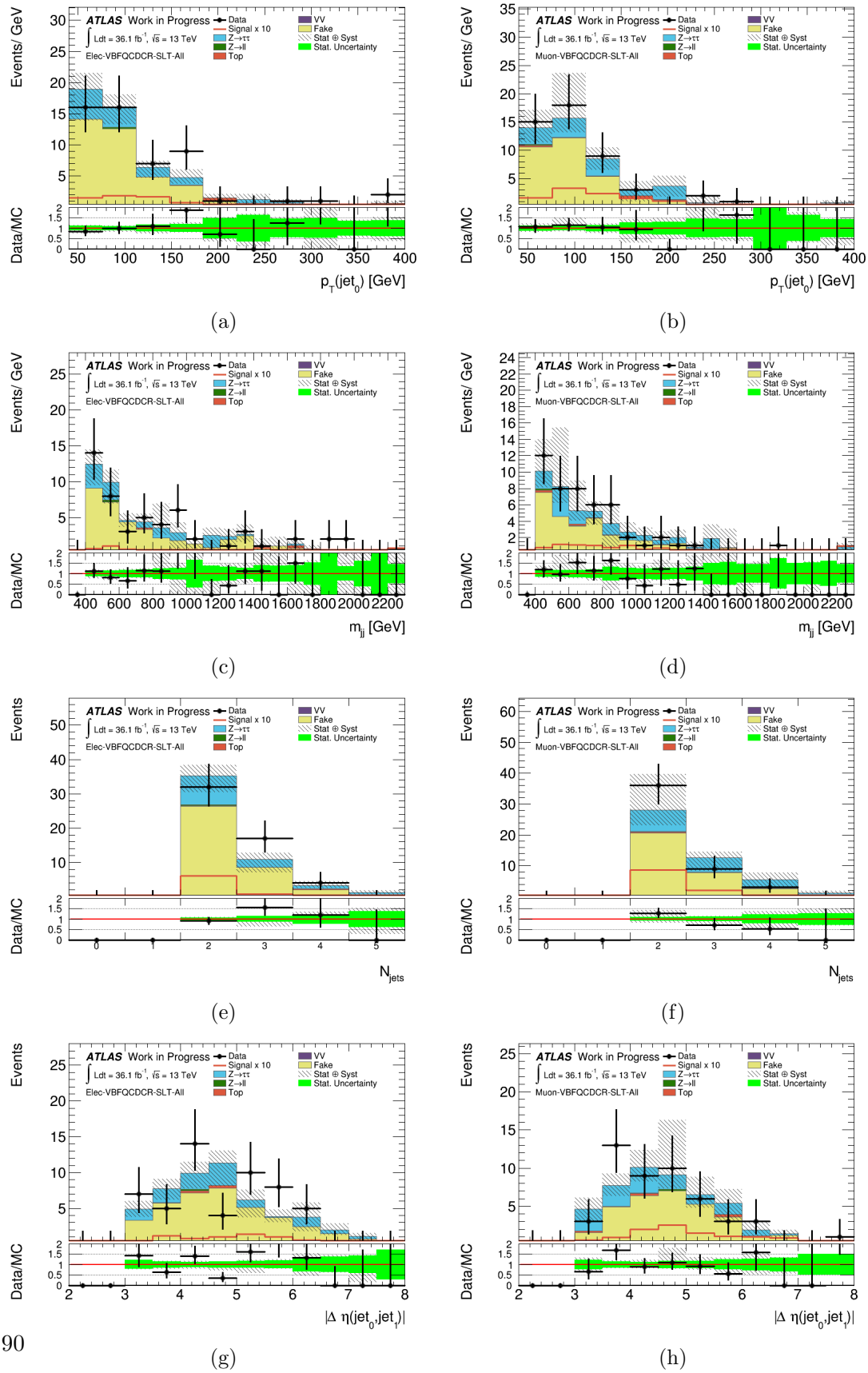


Figure 6.17.: Distributions in QCD CR in VBF region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a),b)* lead jet p_T, *c),d)* M_{jj}, *e), f)* number of jets *g), h)* $|\Delta\eta|_{jj}$.

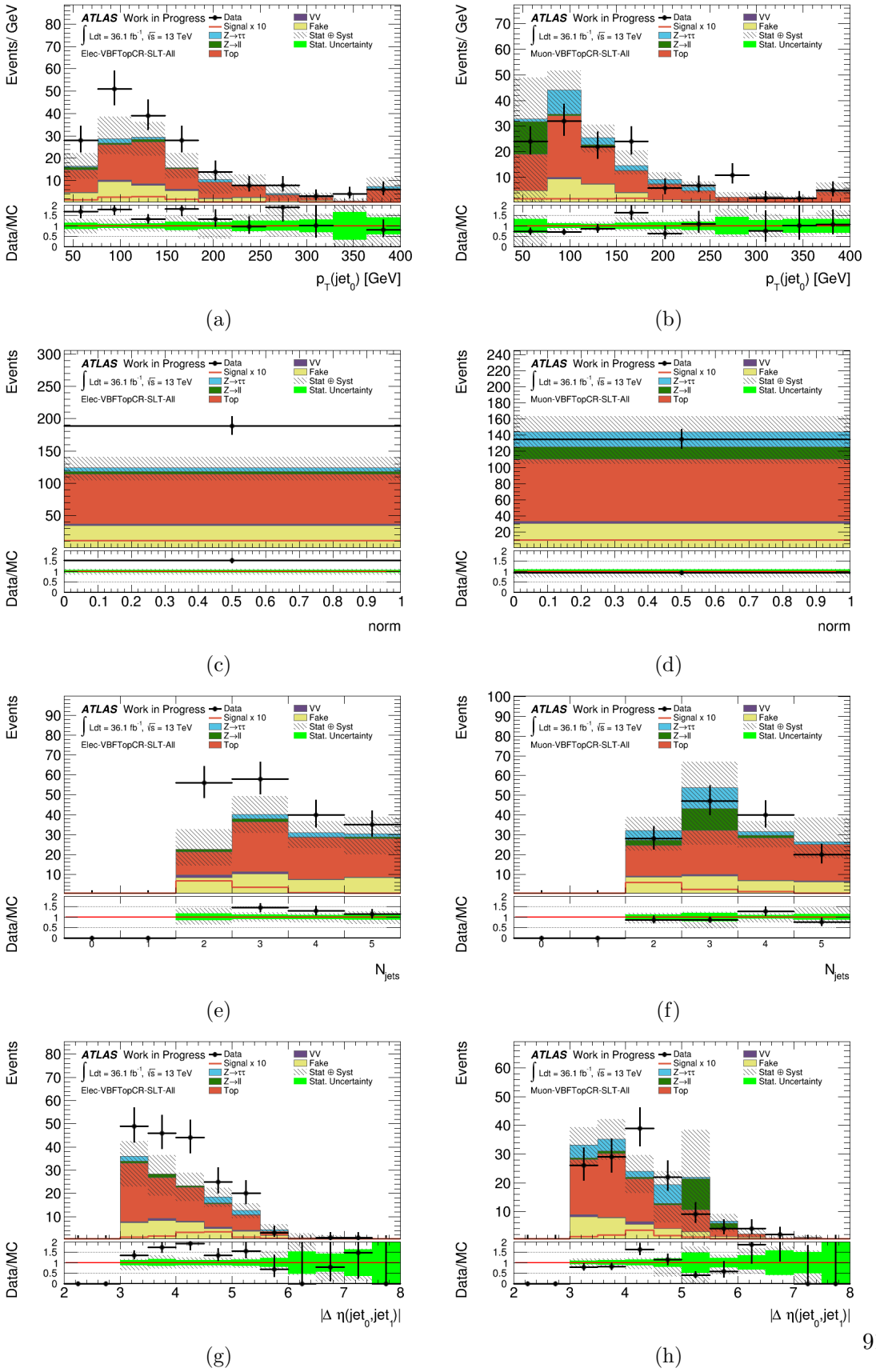


Figure 6.18.: Distributions in Top CR in VBF region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a),b)* lead jet p_T, *c),d)* norm, *e), f)* number of jets *g), h)* $|\Delta\eta|_{jj}$. The mis-modelling observed in the Top CR for the $\tau_e\tau_{had}$ channel will be later *fixed* by the dedicated Top normalisation factor in the final fit.

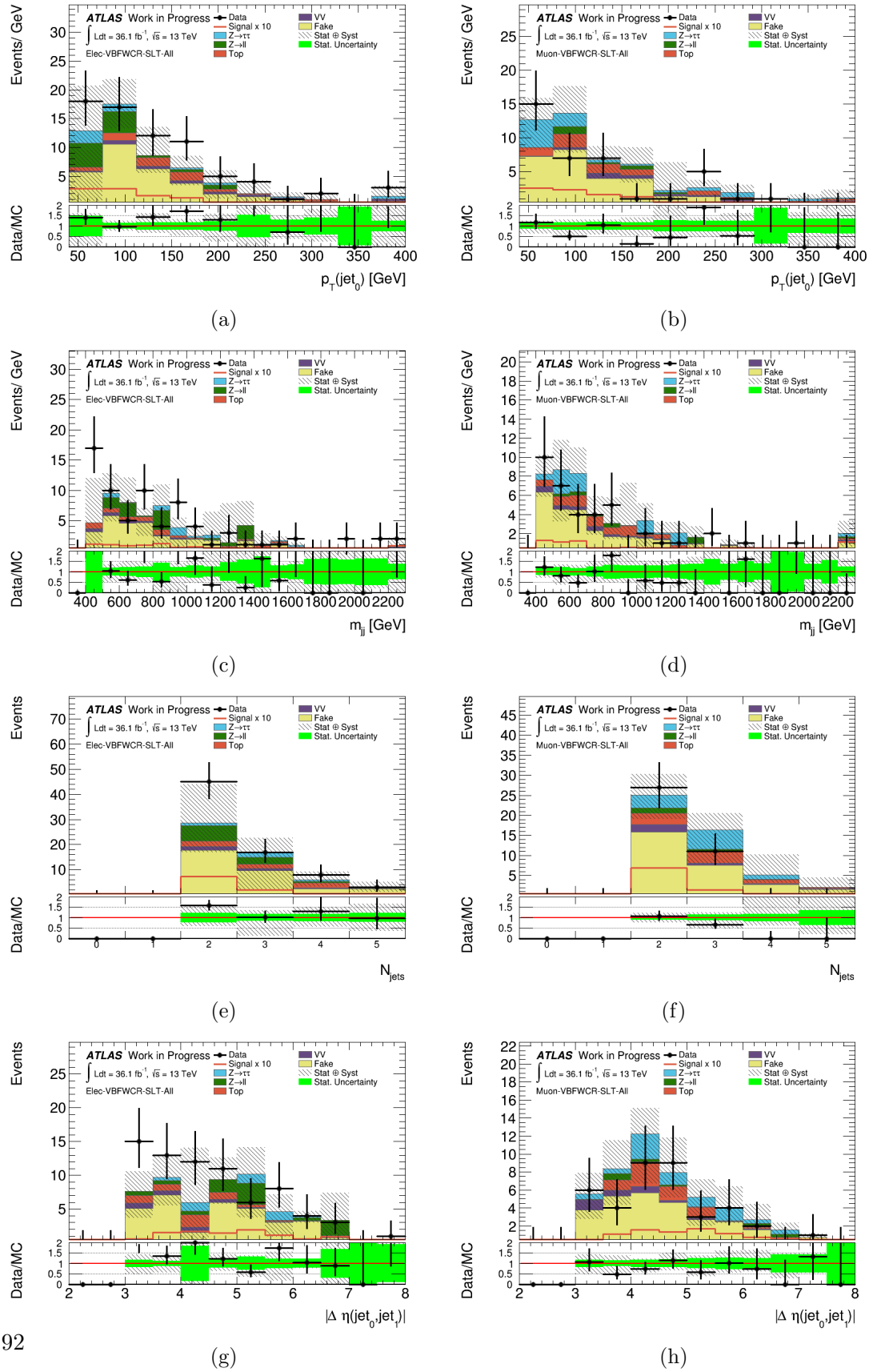


Figure 6.19.: Distributions in W CR in VBF region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a),b)* lead jet p_T, *c),d)* M_{jj}, *e), f)* number of jets *g), h)* $|\Delta\eta|_{jj}$.

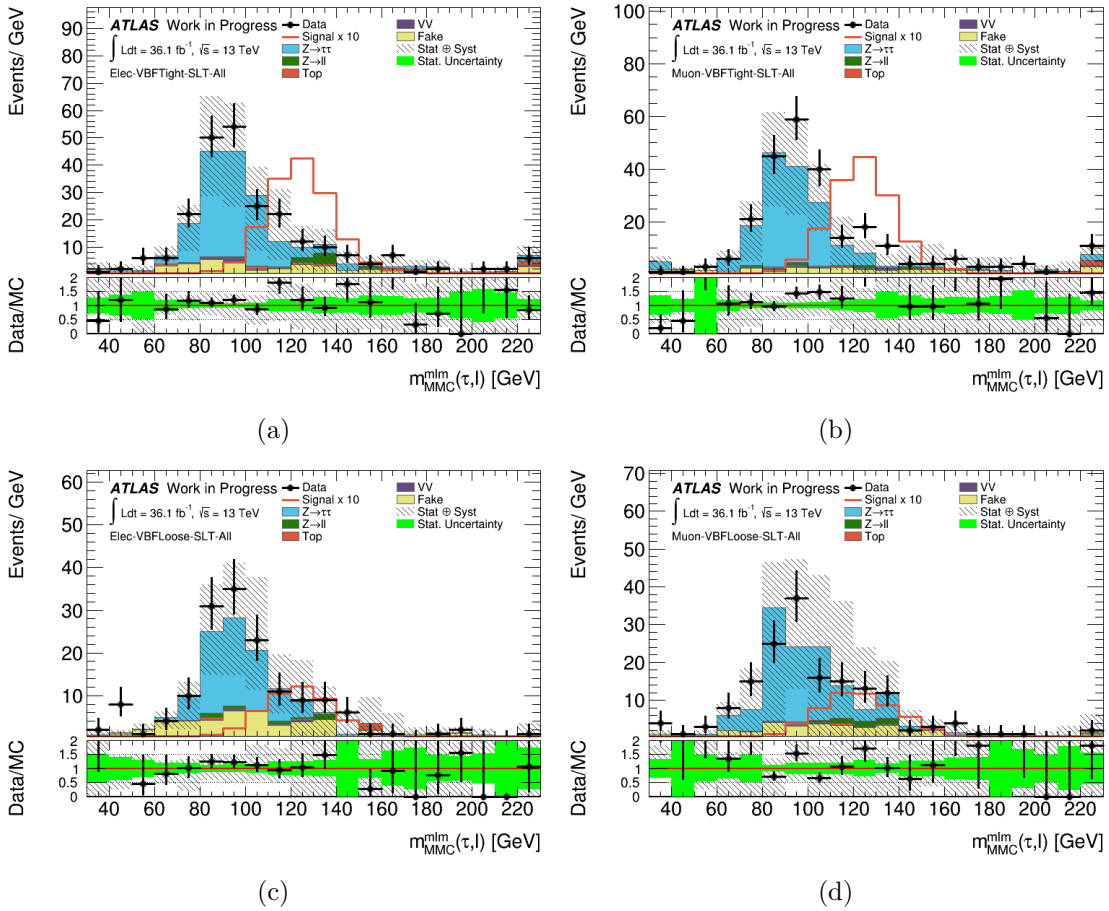


Figure 6.20.: MMC distributions in the final signal regions for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel: *a) (b)* VBF Tight, *c) (d)* VBF Loose. For the signal, $m_H = 125$ GeV is chosen.

6.5.2. Boosted region selection

The *Boosted* selection enriches signal from Higgs boson production with large p_T, typically in the gluon fusion process if a hard jet is produced together with the Higgs boson. The Boosted region is defined by the following requirement:

- pass all the requirements of the preselection, but the VBF selection;
- the transverse momentum of the Higgs boson $p_T^H > 100$ GeV.

This region is further split into two regions to separate phase space regions with different signal-to-background ratios and thus enhance the sensitivity. The events with $\Delta R_{\tau\tau} \leq 1.5$ and $p_T^H > 140$ GeV are selected in *Tight* region, all other events go into the *Loose* region.
In Tables 6.9 and 6.10, event yields in Boosted signal region and relative control regions are reported, while in Fig. 6.21-6.24 the modelling of the main variables used to define

this signal regions are shown. Finally, Fig. 6.25 shows the MMC distribution in the split Boosted signal regions. In general, good Data/MC agreement and good modelling has been observed for all the considered variables.

Sample/Region	$\tau_e\tau_{had}$ SR	$\tau_\mu\tau_{had}$ SR
Fake	1325.1 ± 16.6	1210.2 ± 16.5
VV	126.4 ± 3.2	126.9 ± 3.4
Top	97.0 ± 6.0	92.1 ± 5.5
Zll	131.4 ± 31.2	142.5 ± 19.0
Ztt	4279.3 ± 53.0	4838.8 ± 57.8
Total Bkg	5959.2 ± 64.1	6410.6 ± 63.4
ggH	62.2 ± 0.9	63.3 ± 1.0
VBF	12.6 ± 0.2	13.3 ± 0.2
VH	6.9 ± 0.2	7.01 ± 0.2
Total Sgn	81.8 ± 1.0	83.6 ± 1.0
Data	6138.0 ± 78.3	6765.0 ± 82.2

Table 6.9.: Yields in the Boosted signal region.

Sample/Region	$\tau_e\tau_{had}$ QCD CR	$\tau_\mu\tau_{had}$ QCD CR	$\tau_e\tau_{had}$ Top CR	$\tau_\mu\tau_{had}$ Top CR	$\tau_e\tau_{had}$ W CR	$\tau_\mu\tau_{had}$ W CR
Fake	366.7 ± 7.3	274.0 ± 6.6	728.3 ± 12.9	742.1 ± 12.7	621.1 ± 10.5	547.6 ± 10.1
VV	3.7 ± 0.4	4.9 ± 0.6	35.3 ± 2.1	29.7 ± 1.8	66.9 ± 2.3	65.5 ± 2.4
Top	8.5 ± 1.6	7.5 ± 1.6	1494.4 ± 25.7	1401.5 ± 25.1	78.6 ± 5.7	69.8 ± 4.7
Zll	7.0 ± 8.6	2.8 ± 3.6	21.6 ± 4.4	28.4 ± 6.1	27.8 ± 10.8	51.0 ± 20.
Ztt	142.9 ± 9.8	224.4 ± 13.5	194.5 ± 12.9	217.0 ± 14.6	85.4 ± 7.1	89.0 ± 7.5
Total Bkg	528.9 ± 15.0	513.5 ± 15.6	2474.1 ± 31.9	2418.7 ± 32.4	879.9 ± 17.7	822.9 ± 24.3
ggH	2.1 ± 0.2	3.4 ± 0.3	4.1 ± 0.3	3.1 ± 0.3	2.9 ± 0.2	2.7 ± 0.2
VBF	0.4 ± 0.0	0.7 ± 0.0	0.6 ± 0.0	0.7 ± 0.0	0.6 ± 0.0	0.6 ± 0.0
VH	0.2 ± 0.1	0.3 ± 0.1	0.9 ± 0.1	0.9 ± 0.1	0.7 ± 0.1	0.6 ± 0.1
Total Sgn	2.7 ± 0.2	4.4 ± 0.3	5.7 ± 0.3	4.8 ± 0.3	4.3 ± 0.3	3.8 ± 0.2
Data	495.0 ± 22.2	556.0 ± 23.6	2501.0 ± 50.0	2433.0 ± 49.3	915.0 ± 30.2	822.0 ± 28.7

Table 6.10.: Yields in the *QCD*/*Top* and *W* CRs in Boosted region.

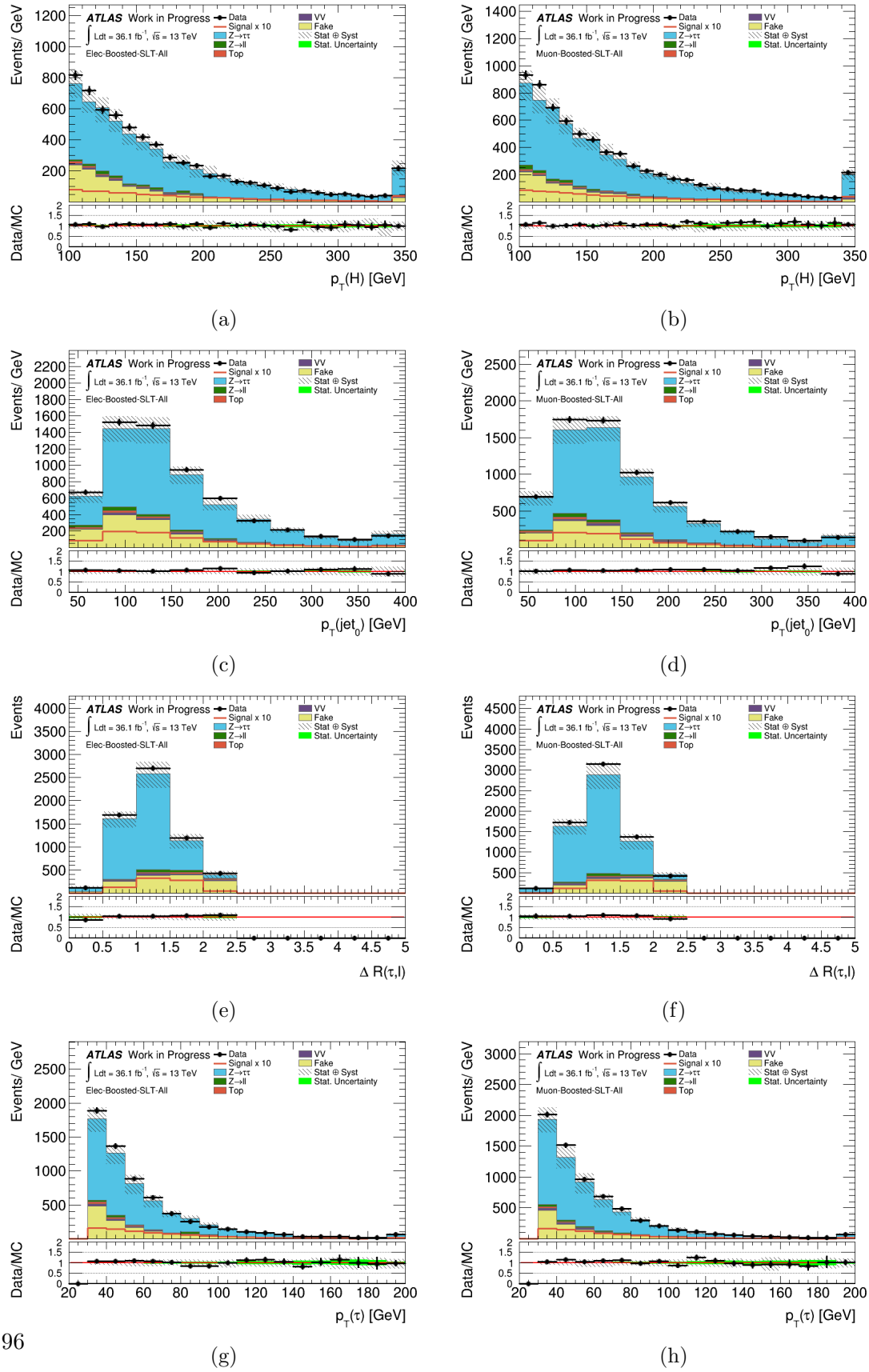


Figure 6.21.: Distributions in Boosted signal region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) : *a) (b)* p_T^H *c) (d)* lead jet p_T, *e) (f)* $\Delta R(lep, \tau)$, *g) (h)* τ p_T.

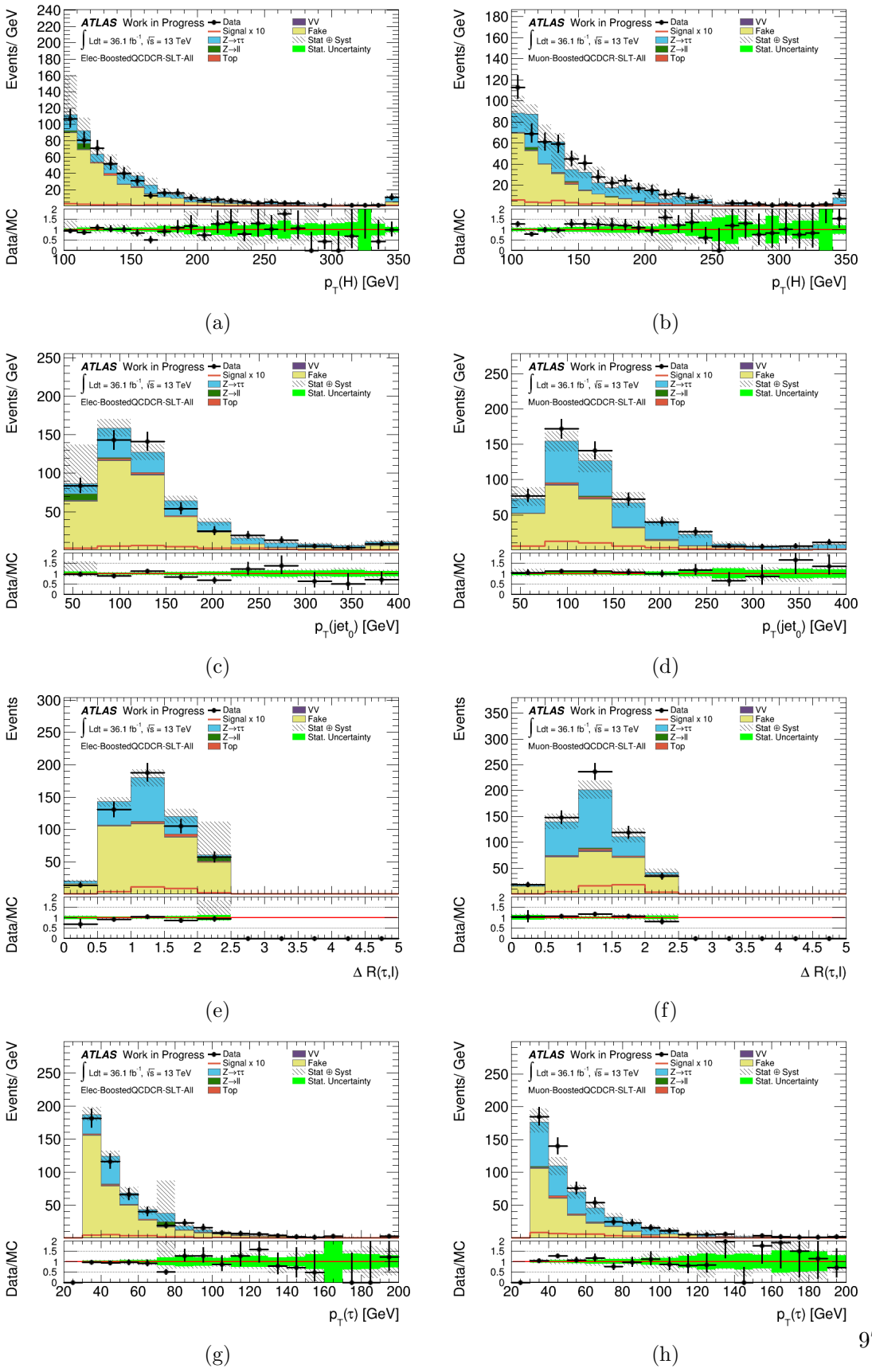


Figure 6.22.: Distributions in QCD CR in Boosted region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* p_T^H, *c) (d)* lead jet p_T, *e) (f)* $\Delta R(lep, \tau)$, *g) (h)* τ p_T. The large uncertainty observed in some bins in the $\tau_e\tau_{had}$ channel are due to large weights coming from Sherpa MC generator for the $Z \to ll$ process.

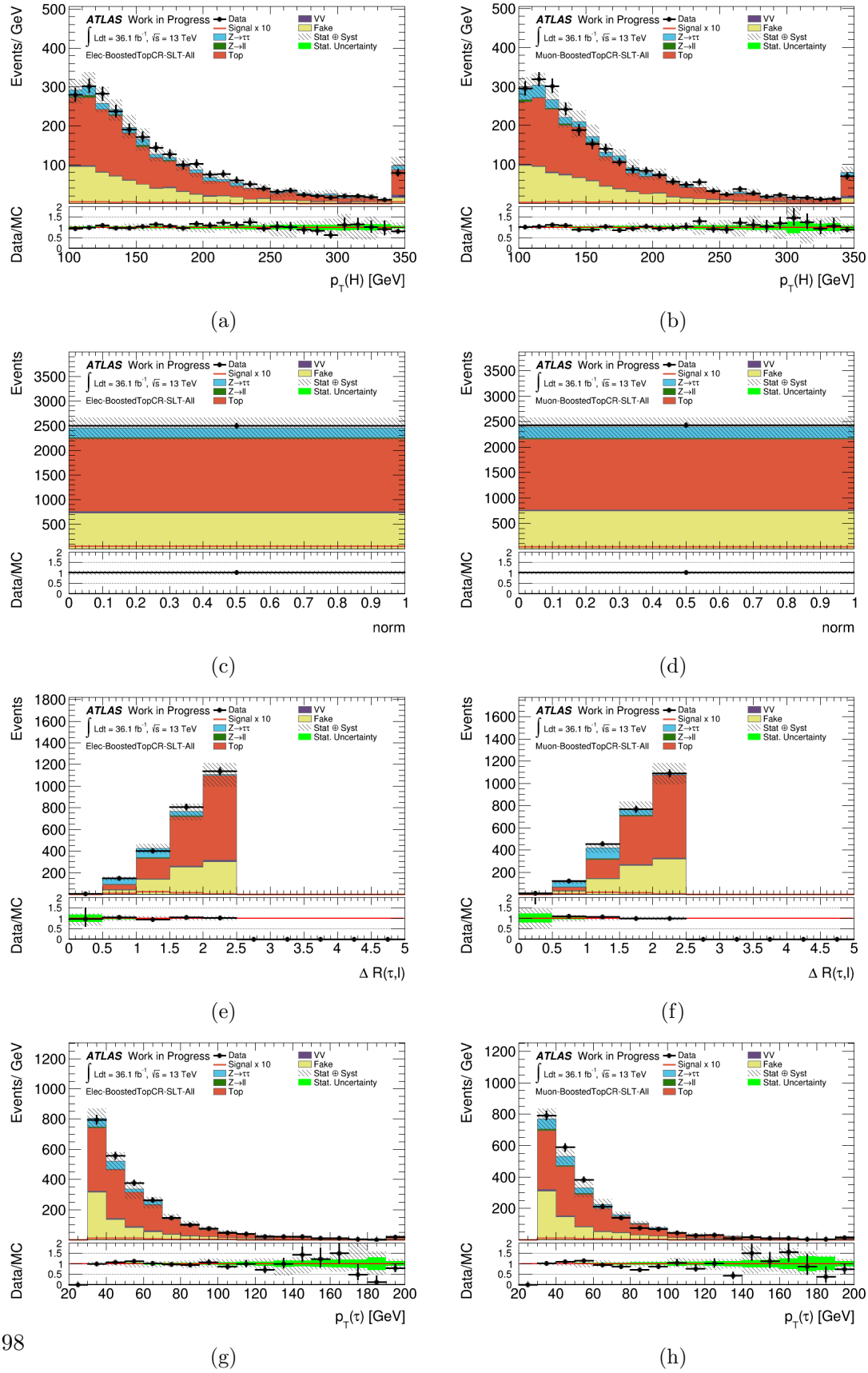


Figure 6.23.: Distributions in Top CR in Boosted region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* p_T^H, *c) (d)* norm, *e) (f)* $\Delta R(lep, \tau)$, *g) (h)* τ p_T.

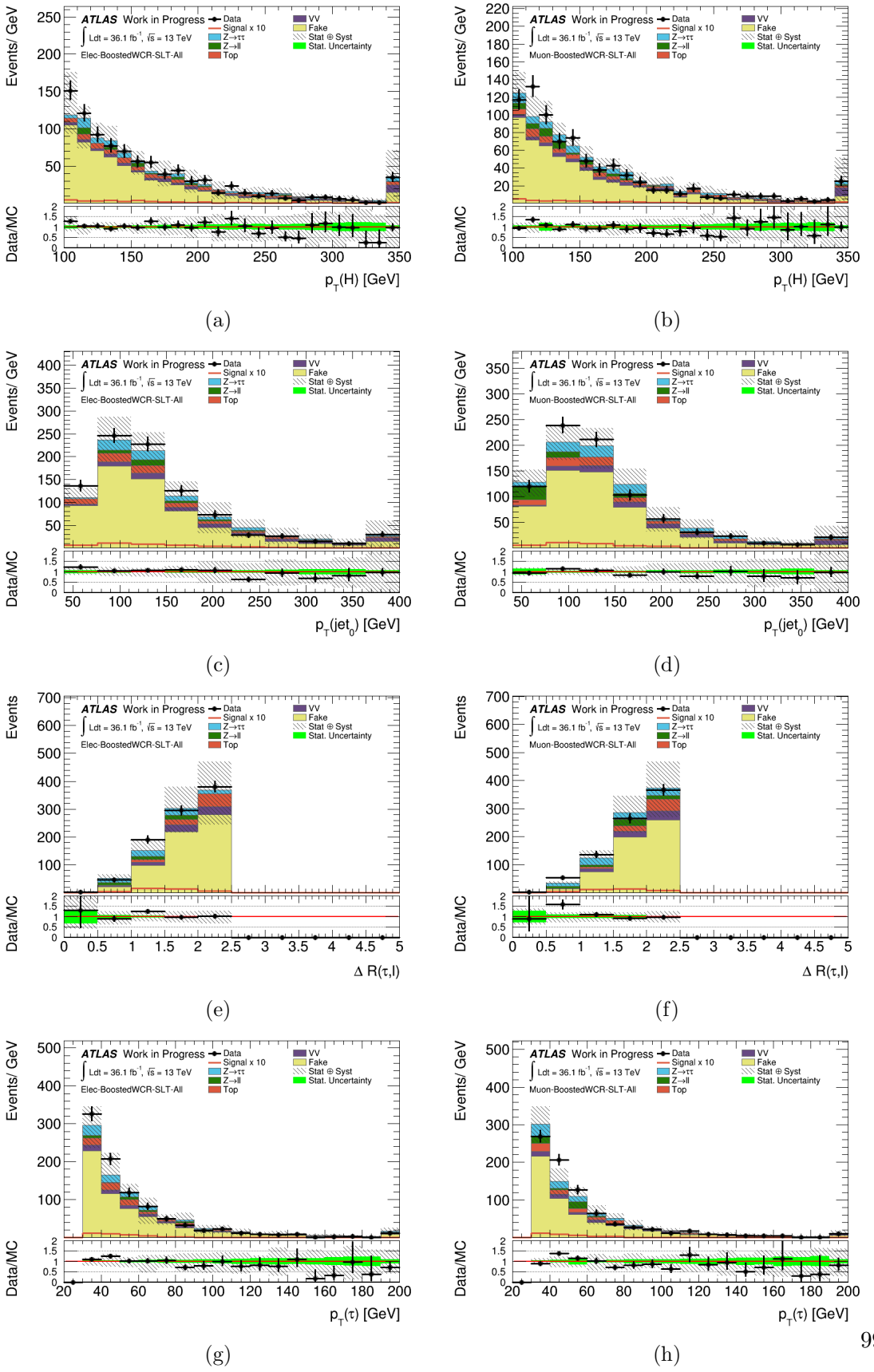


Figure 6.24.: Distributions in W CR in Boosted region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* p_T^H, *c) (d)* lead jet p_T, *e) (f)* $\Delta R(lep, \tau)$, *g) (h)* τ p_T.

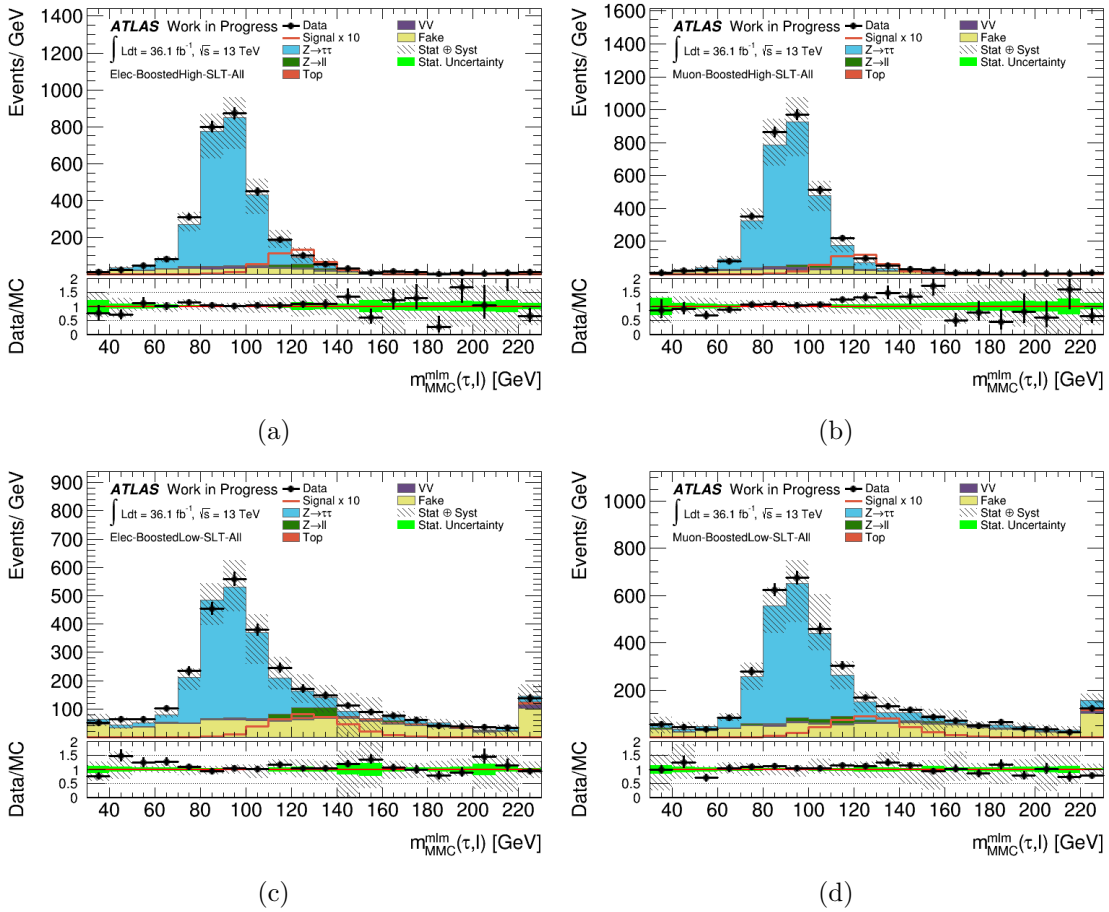


Figure 6.25.: MMC distributions in the final signal regions for the $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel: *a) (b)* Boosted Tight, *c) (d)* Boosted Loose. For the signal, $m_H = 125$ GeV is chosen.

6.6. Systematic Uncertainties

Systematic uncertainties are related to both theoretical and experimental sources and are described in this section.

6.6.1. Experimental Uncertainties

The experimental uncertainties are estimated using the Moriond 2017 recommendations provided by the ATLAS Combined Performance groups. The objects related uncertainties are:

- Electron/Muon uncertainties [92]:
  - *Resolution* : variations related to ID (electron/muon) and MS tracks (muon only);

  - *Scale* : variation of the momentum scale;
  - *Efficiency* : statistical and systematic uncertainties for trigger efficiency, identification efficiency and isolation efficiency; reconstruction efficiency is considered only for electron.

- Tau uncertainties:
  - *Scale* : variation of the energy scale due to the modelling of the detector geometry, measurement performed in tag-and-probe analysis and Geant4 physics list;
  - *Efficiency* : efficiency uncertainties due to the τ identification, reconstruction, electron overlap removal performed independently for true electrons and hadronically decaying taus.

- Jet uncertainties [93]:
  - *Resolution* : the jet resolution uncertainty is parametrised using an 11 nuisance parameters scheme; this scheme expects also data smearing in order the derive the final uncertainty which should be applied on the Monte Carlo samples;
  - *Scale* : the jet scale resolution is parametrised using a 21 nuisance parameters scheme;
  - *Efficiency* : efficiency uncertainties for the jet vertex tagger which is applied both in the central and the forward part of the detector.

  Further source of jet related uncertainties is due to the b-tagging which is used to reject (enhance) Top contribution in the signal (control) regions.

- E_T^{miss} uncertainties:
  The uncertainty on the E_T^{miss} calculation using track-soft terms calculated from the total transverse momentum of the hard objects (electrons, muons, jets and taus) reconstructed in the event; this was evaluated by comparing data taken in 2015 and 2016 with Monte Carlo simulations [93].

- Pileup re-weighting:
  Since Monte Carlo simulation was performed using a generalized profile for the distribution of the number of interactions per bunch crossing, the simulated events have to be re-weighted to describe the observed pileup profile for the 2015 and 2016 datasets. To get the best agreement between data and Monte Carlo simulation, a correction factor of 1/1.16 needs to be applied to the simulated number of interactions. The uncertainty at 1 σ level is given by $1/(1.16\pm0.07)$, however it was recommended to use the more conservative value of $1/(1.16^{+0.07}_{-0.16})$ as it was done in this analysis.

- Luminosity uncertainty:
  The uncertainty on the combined 2015 and 2016 luminosity is 2.9 %. This value

was derived from a preliminary calibration of the luminosity scale using x-y beam-separation scans performed in August 2015 and May 2016. This combined uncertainty assumes fully correlated uncertainties between the 2015 and 2016 dataset.

6.6.2. Theoretical Uncertainties

In the following section, the theoretical uncertainties for $Z \to \tau\tau$ and signal are discussed.

Theoretical uncertainties for $Z \to \tau\tau$

The normalisation of the $Z \to \tau\tau$+jets background is left free-floating in the final analysis fit. Two normalisation factors are defined, controlling the overall normalisation of $Z \to \tau\tau$ across channels in the VBF signal regions and Boosted signal regions respectively. Therefore each considered source of systematic uncertainty can be parametrized using the following strategy:

- a set of nuisance parameters (one per each inclusive SR and per channel) to account for the impact of the variation on the discriminant variable as well as the event migration from a signal region to the other one within a given inclusive region. For each channel, it is evaluated after the variation has been constrained to the same normalisation as the nominal prediction in the inclusive signal regions (Boosted inclusive and VBF inclusive regions);
- two nuisance parameters (one for each inclusive SR) to account for the fact that the definition of the inclusive regions are different across the three channels ($\tau_{lep}\tau_{lep}$, $\tau_{lep}\tau_{had}$ and $\tau_{had}\tau_{had}$) and it can therefore be impacted differently by the given systematic uncertainties. It will be evaluated as the impact of the variation on the relative normalisation of each channel with respect to the total expectation regardless of the decay channel. It is worth to notice that these nuisance parameters are not considered in the case of a standalone fit performed using only one decay channel.

The following sources of uncertainty have been considered:

- *PDF*: evaluated using event-weights provided by the *Sherpa* MC generator;
- *renormalisation and factorisation scales* (μ_R/μ_F) : evaluated using event-weights provided by the *Sherpa* MC generator;
- *CKKW* : jet-to-parton matching uncertainty, evaluated using truth-level parametrisation of the jet multiplicity and p_T(Z);
- *resummation scale (qsf)* : evaluated using truth-level parametrisation as a function of the jet multiplicity and p_T(Z);
- *underlying-event* : evaluating using sample with different setup of multiple interactions. It has been found that this uncertainty has negligible impact and it was not used in the final fit setup;

- parton-shower: due to a lack of recommendations for this source of uncertainty, this has been covered comparing *Sherpa* and *Madgraph* $Z \to \tau\tau$+jets MC generators.

Theoretical uncertainties for Signal

The relevant theoretical uncertainties on the SM predictions used in the analysis arise from three main sources:

- QCD scale uncertainties due to missing higher orders in perturbation theory;
- non-perturbative parts of the calculation (underlying event and hadronisation);
- uncertainty on experimental input parameters such as parton density functions (PDFs) and the value of the strong interaction coupling constant, α_S.

In the final fit model, the uncertainty is based on the variations in the number of events predicted in each MMC bin :

$$n_{bin} = \sigma_{tot} \times \mathcal{L} \times A \times f_{bin} \tag{6.8}$$

where σ_{tot} is the total cross section, $\mathcal{L}$ is the luminosity, A is the acceptance of the total sample in a signal region and f_{bin} is the fraction of events in a certain bin of the MMC distribution. The theory uncertainties are factorised into uncertainties on the total cross section, uncertainties on acceptances into the signal regions and uncertainties of the MMC shape distribution.
The total cross section uncertainties are provided by the *LHC Higgs cross section Working Group (LHCHXSWG)* and are shown in Tables 6.11 and 6.12, respectively for QCD, PDF and α_S uncertainties.
The QCD scale uncertainties are evaluated varying the renormalisation and factorisation scales μ_R, μ_F by factors 2 and 1/2 around the central value with the constraint $1/2 \leqslant \mu_F/\mu_R \leqslant 2$. For the VBF and VH process, the samples are generated with Powheg+Pythia8 at NLO and the internal weights of Powheg are used for the scale variations; the envelope of these variations is then kept as final uncertainty.
For the gluon fusion Higgs production a simple variation of the factorisation and renormalisation scales is not sufficient. Due to accidental cancellations in the perturbative calculation, the variation of the renormalisation and factorisation scales underestimates the uncertainties. For this reason, QCD scale uncertainties for gluon fusion need a specific treatment and 9 sources of uncertainties related to the truncation of the perturbative series are recommended:

- 4 scale variations uncertainties:
  - $\Delta\mu$: factorisation and renormalisation scale variations;
  - $\Delta\phi$: resummation scale variation;
  - $\Delta^{0/1}_{cut}(\Delta^{1/2}_{cut})$: $0 \iff 1(1 \iff 2)$ jet bin migration.
- 2 VBF topology uncertainties:

  - variation of the VBF phase space;
  - 3rd jet veto.

- 2 Higgs p_T-shape uncertainties:
  - Higgs p_T 0-60/60-∞ GeV;
  - Higgs p_T 0-120/120-∞ GeV.

- 1 top mass dependent uncertainty.

The first four variations are established using the method described in the Yellow Report 4 [94], the two VBF topology uncertainties are derived using Yellow Report 3 [29] methods, the Higgs p_T shape variations are taken from the QCD scale variations of Powheg NNLOPS and the last uncertainty, related to the top mass, is derived from differences between LO and NLO rescaling.
Parton shower uncertainties are evaluated comparing two parton shower algorithms, Pythia8 and Herwig7; to do this, two samples for each channel and production mode were produced using ATLAS Fast simulation samples. It is worth to mention that these samples suffer from a lack of statistics and probably should be replaced with larger statistics samples.
PDF uncertainties are evaluated according to the most recent recommendations of the PDF4LHC collaboration: the used PDF set consists of 30 eigenvector variations and 2 variations of α_S which are then statistically combined as independent nuisance parameters in the fit.

Production process	+ QCD scale	- QCD scale
ggH	+3.9 %	-3.9 %
VBF	+0.4 %	-0.3 %
WH	+0.5 %	-0.7 %
ZH	+3.8 %	-3.1 %

Table 6.11.: Total cross section uncertainties due to missing higher orders (QCD scale variations) from YR4 of the LHCHXSWG.

Production process	PDF	α_s	(PDF + α_s)
ggH	±1.8 %	±2.5 %	±3.1 %
VBF	±2.1 %	±0.5 %	±2.1 %
WH	±1.7 %	±0.9 %	±1.9 %
ZH	±1.3 %	±0.9 %	±1.6 %

Table 6.12.: Total cross section uncertainties due to PDFs and α_s from YR4 of the LHCHXSWG.

6.7. Fit model description

The statistical analysis used to determine the final results uses a binned likelihood function constructed as a product of Poisson probability terms. Considering all the three different final states allowed by the different τ decay modes, the global fit model is illustrated in Fig. 6.26. While in the signal region the binned MMC distribution is used to infer information on the parameter of interest (the signal normalisation with respect to the SM prediction), control regions are used to constrain the event yield of a specific background with a single-bin histogram containing the number of events in the control region.

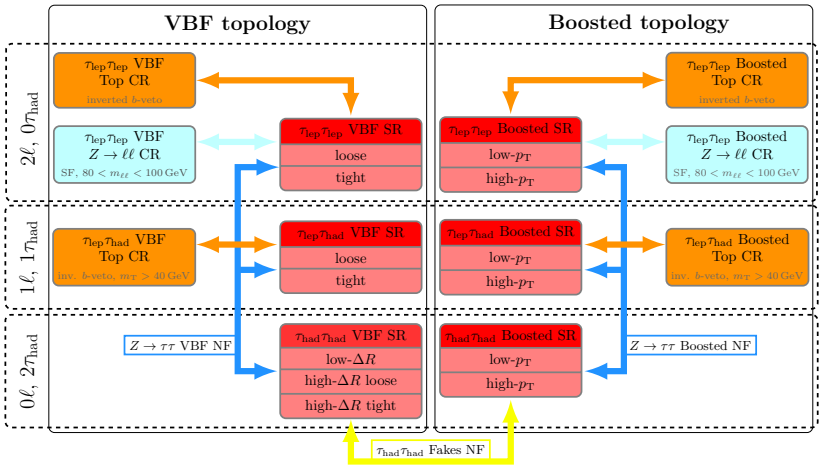


Figure 6.26.: Schematic summary of the analysis fit model. All regions which are used directly in the combined fit are indicated. They are grouped by topology (boosted and VBF) and the number of leptons and hadronic taus. Arrows indicate the free floating normalisation factors which are acting on various regions.

In all the channels the various signal samples and the most relevant sources of backgrounds are considered and their related systematic uncertainties are used in the fit as separate nuisance parameters of the model. For the $\tau_{lep}\tau_{had}$ final state, the Z+jets, Top and Diboson backgrounds are considered as additional separate backgrounds. Processes which contribute with an event yield less than 0.01 events in any given region are removed from the fit for that particular region; this increases fit stability and speed.
As figures of merit to quantify fit results, two important quantities are considered:

- the *significance* (σ), which quantifies the rejection of the background only hypothesis; this will be further described in Section 6.7.1;
- the *signal strength* (μ), which is defined as

$$\mu = \frac{\sigma \times BR}{(\sigma \times BR)_{SM}}$$

and represents the signal normalisation with respect to the SM prediction. The value $\mu = 0$ corresponds to the absence of signal, while the value $\mu = 1$ suggests the signal presence as predicted by the SM.

6.7.1. Statistical analysis

For purposes of discovering a new signal process (like $H \to \tau\tau$), one can define the null hypothesis, H_0, as describing only known processes which are considered as background [95]. This has to be tested against the alternative hypothesis H_1, which includes both background as well as signal processes. To summarise the result of such a search, one quantifies the level of agreement of the observed data with a given hypothesis H by computing a *p-value*, that is a probability, assuming H, of finding data of equal or greater incompatibility with the predictions of H. One can regard the hypothesis as excluded if its *p-value* is observed below some threshold.
It is possible to convert the *p-value* into an equivalent significance, called Z, defined such that a Gaussian distributed variable found Z standard deviations above its mean has an upper-tail probability equal to p. That is

$$Z = \Phi^{-1}(1 - p) \tag{6.9}$$

where Φ^{-1} is the inverse of the cumulative distribution of the standard Gaussian (referred as *quantile*). For a signal process, the particle physics community tends to regard rejection of the background hypothesis H_0 with a significance of at least Z = 5 as an appropriate level to claim a discovery. This corresponds to $p = 2.87 \times 10^{-7}$.
A widely used procedure to establish discovery is based on the frequentist significance test using a *likelihood ratio* as test statistics. In addition to the parameter of interest such as the rate of the signal process, the signal and the background models will contain in general nuisance parameters whose values are not taken as known *a priori* but rather must be fitted from the data. The flexibility introduced to parametrise systematic effects results in a loss of sensitivity.

6.7.2. Likelihood construction

The parametrised probability density function (likelihood) constructed by *HistFactory* is of the form:

$$P(n_c, x_e, a_p|\phi_p, \alpha_p, \gamma_b) = \prod_{c \,\epsilon\, chan} \Big[Pois(n_c|\nu_c) \prod_{e=1}^{n_c} f_c(x_e|\alpha)\Big] G(L_0|\lambda, \Delta_L) \prod_{p \,\epsilon\, S+\Gamma} f_p(a_p|\alpha_p) \tag{6.10}$$

where $f_p(a_p|\alpha_p)$ is a constraint term describing an auxiliary measurement a_p that constraints the nuisance parameter α_p and $G(L_0|\lambda, \Delta_L)$ is a Gaussian constraint term which depends on the luminosity. Usually data are represented using histograms, so that the expected (mean) number of events in a given bin is given by:

$$\nu_{cb}(\phi_p, \alpha_p, \gamma_b) = \lambda_{cs}\gamma_{cb}\phi_{cs}(\alpha)\eta_{cs}(\alpha)\sigma_{sb}(\alpha) \tag{6.11}$$

where

- λ_{cs} is a luminosity parameter for a given channel and sample;
- γ_{cb} is a bin-by-bin scale factor used for statistical uncertainties, bin-by-bin shape systematics and data-driven shape extrapolation;
- $\phi_{cs}(\alpha)$ is the product of the unconstrained normalisation factors for a given sample within a given channel. These typically includes the parameter of interest (signal cross-section or branching ration);
- $\eta_{cs}(\alpha)$ are the parametrized normalisation uncertainties for a given samples within a given channel;
- $\sigma_{sb}(\alpha)$ is the parametrized histogram (nominal histograms and variations) for a given sample within a given channel.

The mean number of events in each bin implies the following probability density function:

$$f_c(x_e|\phi_p, \alpha_p, \gamma_b) = \frac{\nu_{cb_e}}{\nu_c} \qquad \nu_c = \sum_{b\ \epsilon\ bins\ of\ chan\ c} \nu_{cb} \tag{6.12}$$

so then the likelihood can be also expressed as a product over bins

$$P(n_{cb}, a_p|\phi_p, \alpha_p, \gamma_b) = \prod_{c\ \epsilon\ chan} \prod_{b\ \epsilon\ bins} Pois(n_{cb}|\nu_{cb}) G(L_0|\lambda, \Delta_L) \prod_{p\ \epsilon\ S+\Gamma} f_p(a_p|\alpha_p) \tag{6.13}$$

6.7.3. Test Statistics

To test a hypothesized value of μ, we consider the profile likelihood (L) ratio:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})} \tag{6.14}$$

where θ is the set of the nuisance parameters in a given model, $\hat{\hat{\theta}}$ in the numerator denotes the value of θ that maximizes L for the specified μ, so it is the conditional maximum-likelihood (ML) estimator of θ. The denominator is the maximised (unconditional) likelihood function, so $\hat{\mu}$ and $\hat{\theta}$ are the likelihood estimators. The presence of the nuisance parameters broadens the profile likelihood as a function of μ relative to what one would have if their values were fixed. This reflects to the loss of information about μ due to systematic uncertainties.
By definition, λ can take values $0 \leq \lambda \leq 1$, with λ near 1 implying good agreement between the data and the hypothesized value of μ. Equivalently it is possible to use the quantity:

$$t_\mu = -2\ln(\lambda(\mu)) \tag{6.15}$$

as the basis of a statistical test. Higher values of t_μ will correspond to increasing incompatibility between the data and μ.
To quantify the level of disagreement, the *p-value* is used:

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu)dt_\mu \tag{6.16}$$

where $t_{\mu,obs}$ is the value of the test statistics t_μ observed from data and $f(t_\mu|\mu)$ denotes the probability density function of t_μ under the assumption of the signal strength μ. If it is assumed that the signal process necessarily has $\mu > 0$, it is possible to define an alternative test statistics called $\tilde{t_\mu}$ based on the fact that if one finds data such that $\tilde{\mu} \leq 0$ then the best level of agreement between the data and any physical value of μ occurs for $\mu = 0$. Thus

$$\tilde{t_\mu} = -2\ln\tilde{\lambda}(\mu) = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \tilde{\mu} \leq 0 \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & \tilde{\mu} > 0 \end{cases} \tag{6.17}$$

An important case of the statistics $\tilde{t_\mu}$ is used to test $\mu = 0$ in a class of models where it is assumed $\mu > 0$, like it was assumed in the analysis described in this thesis. Rejecting the $\mu = 0$ hypothesis effectively leads to the discovery of a new signal. Thus in this case the previous equation becomes:

$$\tilde{t_0} = \begin{cases} 0 & \tilde{\mu} \leq 0 \\ -2\ln\lambda(0) & \tilde{\mu} > 0 \end{cases} \tag{6.18}$$

where $\lambda(0)$ is the profile likelihood ratio for $\mu = 0$. Using $\tilde{t_0}$, it is considered the data to show lack of agreement with the background only hypothesis only if $\tilde{\mu} > 0$. A value of $\tilde{\mu}$ much below zero could be an evidence for the background-only model, but this type of discrepancy does not show that the data contain signal events but rather points to some other systematic error. As the event yield increases above the expected background, one finds increasingly large values of $\tilde{t_0}$, corresponding to an increasing level of incompatibility between the data and the $\mu = 0$ hypothesis.

6.7.4. Pruning strategy

All systematic uncertainties enter in the fit as nuisance parameters (NPs) on the normalisation, provided that their effect is larger than 0.5 %; additionally, some of the NPs also act as shape NPs in the fit. An important issue in dealing with the shape NPs in the fit model is the fact that many of the samples have relatively low number of events after all selections have been applied; in this case small systematic variations can correspond to upwards or downwards varied shapes which are dominated by statistical *noise*. This noise can cause fit instabilities and it will bring to incorrect or anomalous variation of

the NPs. In order to suppress the noise without accidentally removing genuine and significant shape variations, a strategy has been developed to prune and smooth the shape of the systematics before they enter in the fit. The following smoothing and pruning criteria are applied:

- *Symmetrisation*: some systematic variations have *one-sided bins*, that is cases in which in one or more bins, the +1 (-1) σ systematic variation lie in the same direction (up or down) with respect to the nominal value. This can cause bad behaviours in the fit, so for this reason a *local symmetrisation* procedure is applied: for any one-sided bin, the greatest variation (+1 σ or -1 σ) is kept, while the other one is set to a variation with respect to the nominal as large as the greatest variation, but in the opposite direction with respect to the nominal. The modified yield is set to a tiny value (10^{-6}) if a suggested symmetrised yield is negative;

- *Pruning 0*: shape variation of a nominal histogram with a large statistical uncertainty tends to make falsely large shape variation due to statistical fluctuation (e.g. bin-to-bin migration). Therefore, shape systematics are pruned away when the statistical uncertainty of the integral (total yield) is greater than 0.1;

- *Pruning 1*: a χ^2 test between the upwards and downwards fluctuated shape with respect to the nominal is performed, for each potential shape systematic NP and for each sample. The shape systematic is retained if the result of the reduced χ^2 test is greater than 0.1 for either of the upwards or downwards fluctuated shape. This pruning criterion is not applied to those systematics whose variation only occurs through the variation of the weights of MC events;

- *Smoothing*: the ratio of the variation with respect to nominal is smoothed using ROOT method *TH1::Smooth*. The smoothed varied shape is then obtained by multiplying the nominal with the smoothed ratio. This smoothing procedure is only applied to all the bins except overflow and underflow bins;

- *Pruning 2*: for a potential shape systematic to be considered significant and thus used in the fit, it must also match the following additional pruning criterion: the maximum bin-by-bin variation significance, $max_i S_i$, should be at least 0.1, where the significance S_i is defined as $S_i = |u_i - d_i|/\sigma_i^{tot}$, with u_i (d_i) being the upwards (downwards) variation in bin i for a given sample, while σ_i^{tot} is the statistical uncertainty for the total estimation in bin i. So if for a given sample, for a given systematic uncertainty, $S_i \leq 0.1$ for all MMC bins, then the shape variation is considered not significant and not considered further.

In addition, normalisation degrees of freedom are only considered if either the up or the downward variation differs from the nominal yield by more than 0.5%, and provided they are larger than the statistical uncertainty of the sample to which they are applied.

6.8. Binning optimisation strategy

In this section, the strategy followed to choose the MMC binning in order to maximise the significance in the final fit will be explained. Optimisation has been made in the MMC range 30 GeV $\leq$ MMC $\leq$ 230 GeV and fitting Boosted (VBF) Tight/Loose categories simultaneously; in this way, possible correlation effects between Tight and Loose categories are taken into account. It is worth to mention that the optimisation has been done doing *Asimov* fit, that is replacing data in signal and control regions with the corresponding MC estimation before performing the fit at each iteration. The optimisation algorithm is divided in three steps:

- scan to choose first bin high edge: in this first step, the first bin high edge has been varied and the significance has been recalculated according to the new binning. Fig. 6.27 shows the values of the significance for each choice of the first bin high edge. From the values reported in the plot, first bin boundaries have been chosen to be [30, 60] GeV for both Boosted and VBF regions; this is a good compromise both to avoid empty bins and to leave the rest of the MMC range large enough for the last optimisation step;

- scan to choose last bin low edge: in this second step, after having fixed the first bin according to results from the previous step, the last bin low edge has been varied and the significance has been recalculated according to the new binning. Fig. 6.28 shows the values of the significance for each choice of the last bin low edge. From the values reported in the plot, last bin boundaries have been chosen to be [180, 230] GeV for both Boosted and VBF regions; this is a good compromise both to avoid empty bins and to leave the rest of the MMC range large enough for the last optimisation step;

- scan to choose the best bin size in the middle MMC range: in this third step, after having fixed first and last bin according to the previous steps, the bin size in the MMC range 60 GeV $\leq$ MMC $\leq$ 180 GeV has been varied and the significance have been recalculated according to the new binning. It is worth to notice that the bin size will automatically determine the number of bins. Fig. 6.29 shows the values of the significance for each choice of the middle bins size. From the values reported in the plot, a middle bin size of 27 (25) GeV has been chosen for Boosted (VBF) region.

Using this procedure, the final MMC binning has been determined. An extension of this procedure will consist of optimising simultaneously VBF and Boosted regions together; this has been tried and the final results will not improve what has been done optimising VBF and Boosted region MMC binning separately.

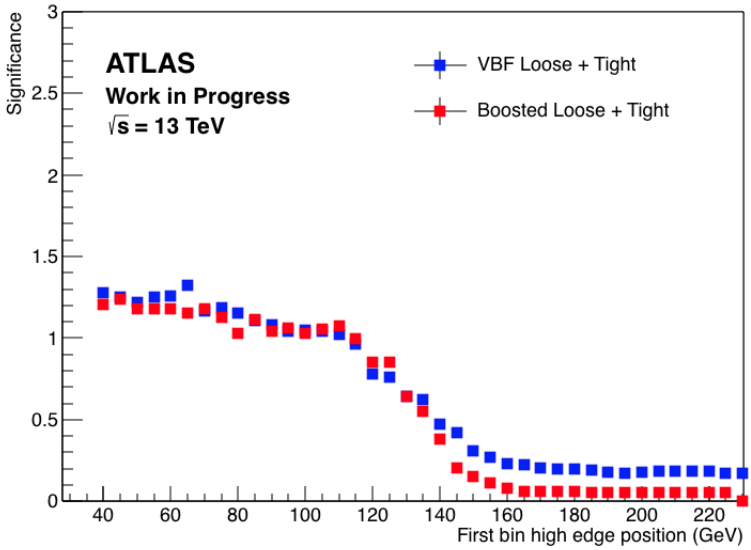


Figure 6.27.: First step: optimisation of the first bin high edge.

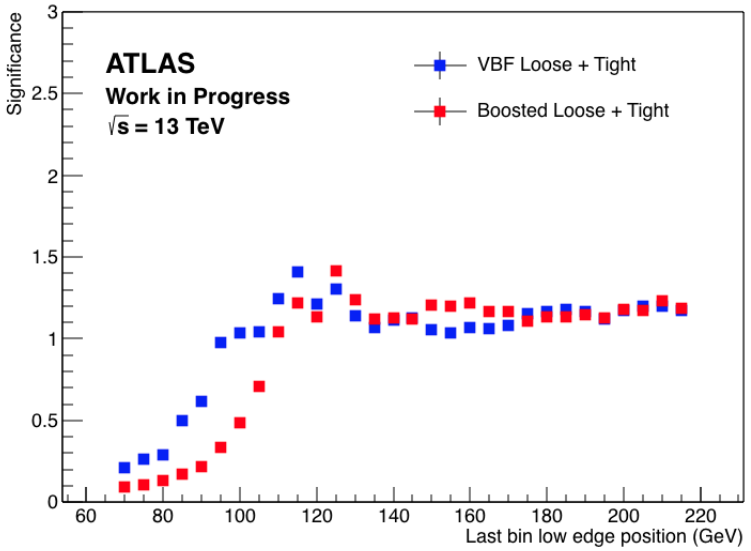


Figure 6.28.: Second step: optimisation of the last bin low edge.

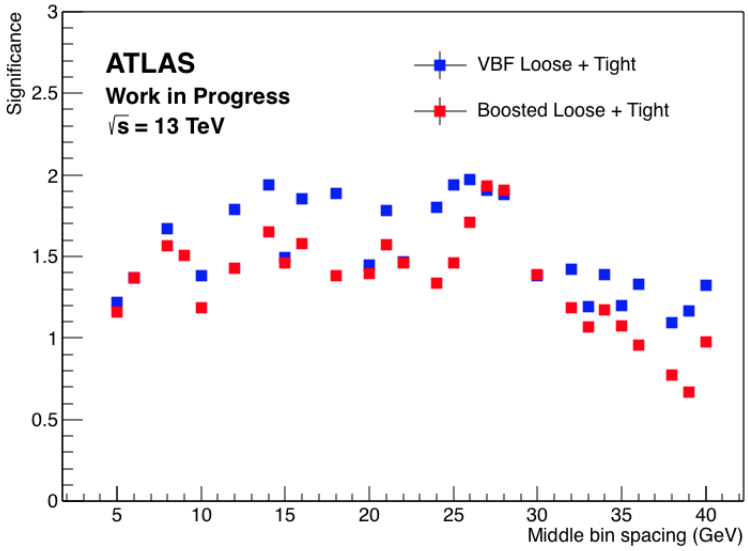


Figure 6.29.: Third step: optimisation of the middle bins size.

6.9. Results for standalone $\tau_{lep}\tau_{had}$ fit

As mentioned at the beginning of this section, the two most important figures of merit which are used to quantify fit results are the significance (σ) and the signal strength (μ). In Table 6.13 the final fit results are reported for both Asimov and Unblinded fit, while in Fig. 6.30 the post-fit MMC distributions for Unblinded fit are shown. As explained before, the Asimov fit is performed substituting total MC predictions to data, while the Unblinded fit is performed using data both in the signal and control regions. Both the observed and the expected significances are reported, where the expected significance has been calculated performing an Asimov fit after fixing the nuisance parameters and the normalisation factors to their post-fit values derived from Unblinded fit. Looking at the values reported in the table, the value of μ compatible with 1 suggests the presence of the signal as predicted by the SM; however, the significance is below the 5 standand deviations, so the background only hypothesis can not be excluded yet. Finally, Tables 6.14 and 6.15 are reporting the post-fit yields in Boosted and VBF signal regions, respectively.

	Fit type / Categ.	Combined	Boost	VBF
Exp. (Obs.) Significance	Asimov	2.813 (2.813)	1.912 (1.912)	1.941 (1.941)
	Unblinded	2.425 (2.627)	1.887 (1.840)	1.648 (1.702)
Signal strength	Asimov	$1^{+0.45}_{-0.38}$	$1^{+0.64}_{-0.53}$	$1^{+0.64}_{-0.53}$
	Unblinded	$0.9^{+0.43}_{-0.38}$	$1.01^{+0.68}_{-0.55}$	$0.96^{+0.70}_{-0.60}$

Table 6.13.: Comparison of the significance/signal strength for Asimov/Unblinded standalone $\tau_{lep}\tau_{had}$ fit.

Sample/Region	Boost High	Boost Low	Boost Top CR
ggH125	60.59 ± 28.27	56.39 ± 23.76	6.64 ± 3.18
VBFH125	12.36 ± 5.64	10.80 ± 5.04	1.23 ± 0.56
WH125	4.39 ± 1.94	3.50 ± 1.522	0.96 ± 0.45
ZH125	2.38 ± 1.04	1.62 ± 0.71	0.64 ± 0.29
Ztt	5371.65 ± 85.65	4211.30 ± 97.61	440.93 ± 54.53
Zll	111.46 ± 23.60	128.41 ± 39.64	34.89 ± 25.40
Top	49.80 ± 10.68	138.76 ± 28.52	2916.22 ± 162.26
Fake	606.63 ± 28.64	1878.86 ± 85.01	1441.81 ± 126.56
Otherslh	136.79 ± 12.07	121.80 ± 11.85	5.5189 ± 7.61
Total	6356.19 ± 79.72	6551.82 ± 80.37	4932.98 ± 70.55
Data	6347 ± 79.67	6556 ± 80.97	4934 ± 70.24

Table 6.14.: Postfit yields for Boost region.

Sample/Region	VBF Tight	VBF Loose	VBF Top CR
ggH125	6.74 ± 3.06	3.53 ± 1.66	1.16 ± 0.64
VBFH125	21.23 ± 9.56	5.68 ± 2.59	1.03 ± 0.46
WH125	0.12 ± 0.07	0.11 ± 0.08	0.03 ± 0.02
ZH125	0.03 ± 0.02	0.03 ± 0.02	0.01 ± 0.01
Ztt	336.34 ± 25.25	182.90 ± 20.21	26.88 ± 9.25
Zll	11.20 ± 3.37	12.19 ± 4.36	17.61 ± 16.51
Top	16.59 ± 5.68	8.18 ± 3.39	201.91 ± 41.64
Fake	96.67 ± 15.89	98.97 ± 17.14	67.87 ± 29.64
Otherslh	9.45 ± 2.124	3.55 ± 1.68	6.083 ± 1.27
Total	498.41± 21.97	315.15± 17.19	323.50± 18.14
Data	496 ± 22.27	318 ± 17.83	324 ± 18

Table 6.15.: Postfit yields for VBF region.

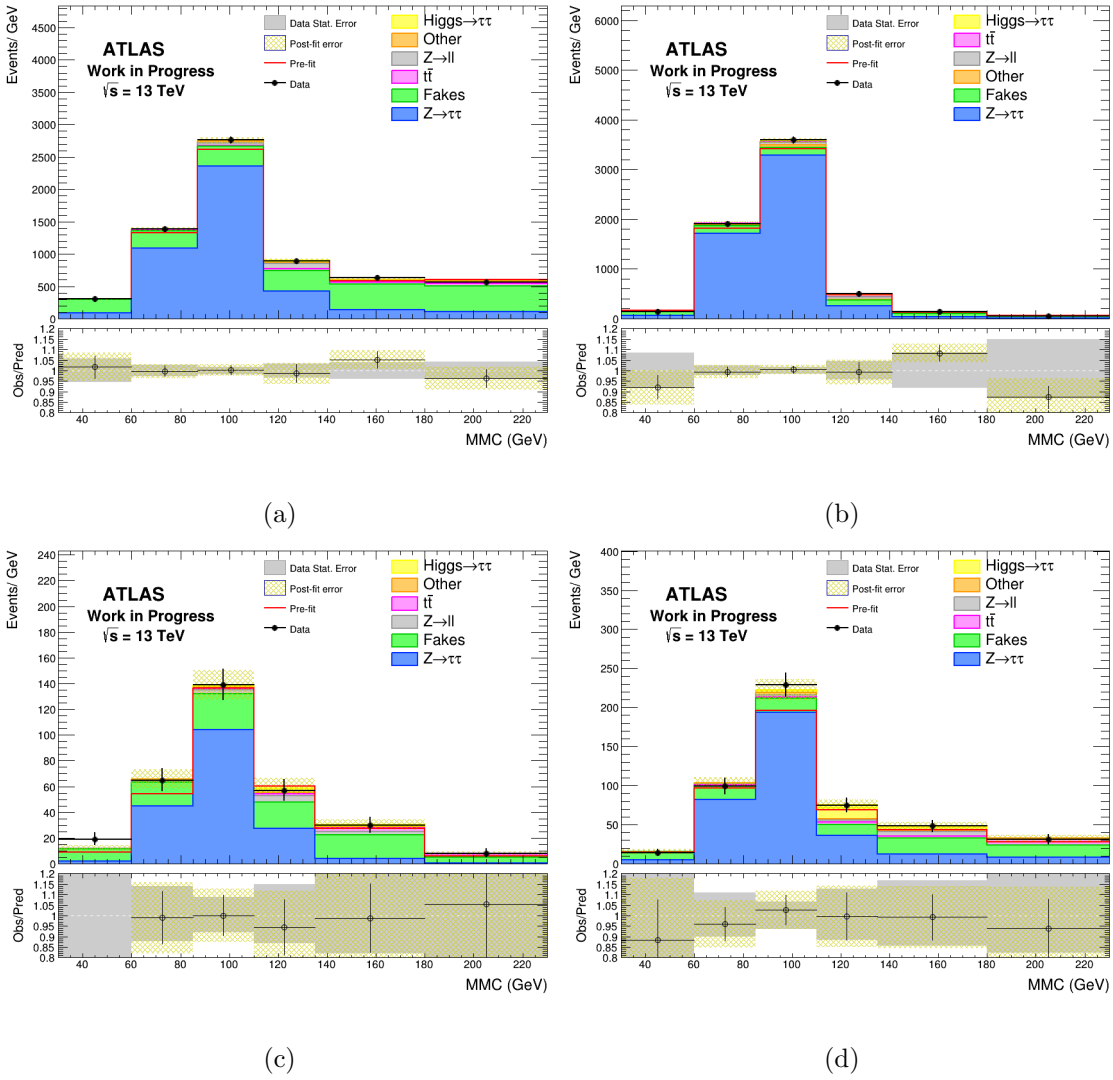


Figure 6.30.: MMC Postfit distributions for $\tau_{lep}\tau_{had}$ channel in: *a)* Boost Loose, *b)* Boost Tight, *c)* VBF Loose *d)* VBF Tight regions.

The error quoted in Table 6.13 for the signal strength can be decomposed to check which source of systematic uncertainty is having larger impact on the global error. Tables 6.16-6.18 show the uncertainty breakdown on μ for combined Boosted+VBF and Boosted/VBF only fit categories, respectively. Looking at the results in the tables, it is possible to conclude that:

- systematic uncertainties have larger impact than statistical uncertainty; this is indicating that the addition of the data recorded by ATLAS during 2017 without improving systematic uncertainty measurements will not lead to a substantial improvement of the final results;

- among the systematic uncertainties, the theoretical uncertainties are having similar impact as the experimental uncertainties in the Boosted region, while in the VBF region the experimental systematics are dominating. This could be expected given the conservative uncertainties which are provided for the gluon-fusion Higgs production mode, which is the main source of signal in the Boosted region;

- among the experimental uncertainties, the Jet and E_T^{miss} related systematics have bigger impact with respect to other objects related uncertainties. This could be also expected since jets and E_T^{miss} have poor reconstruction quality with respect to leptons, so the uncertainties are usually larger; moreover, both Boosted and VBF region selections are based on jet selection, so the related uncertainties have large impact both in terms of events acceptance and also on the MMC shape.

- the MC statistical uncertainty is having large impact especially in VBF region; this is suggesting that the analysis would benefit from larger MC samples with better event filtering to get more statistics in this particular phase space.

POI Value	0.896
Total error	+0.440 -0.382
Statistical	+0.267 -0.261
Systematic	+0.350 -0.279
Theo. Systematics	+0.194 -0.094
Exp Systematics	+0.296 -0.245
Jet/E_T^{miss} Systematics	+0.223 -0.183
Tau Systematics	+0.095 -0.038
MC Statistics	+0.150 -0.140

Table 6.16.: Uncertainty breakdown on μ for Standalone $\tau_{lep}\tau_{had}$ Boost+VBF categories fit.

POI Value	1.015
Total error	+0.684 -0.554
Statistical	+0.359 -0.353
Systematic	+0.582 -0.427
Theo. Systematics	+0.417 -0.182
Exp Systematics	+0.441 -0.349
Jet/E_T^{miss} Systematics	+0.384 -0.318
Tau Systematics	+0.122 -0.042
MC Statistics	+0.051 -0.040

Table 6.17.: Uncertainty breakdown on μ for Standalone $\tau_{lep}\tau_{had}$ Boost category fit.

POI Value	0.960
Total error	+0.698 -0.593
Statistical	+0.427 -0.405
Systematic	+0.552 -0.433
Theo. Systematics	+0.175 -0.045
Exp. Systematics	+0.528 -0.427
Jet/E_T^{miss} Systematics	+0.412 -0.280
Tau Systematics	+0.108 -0.030
MC Statistics	+0.283 -0.294

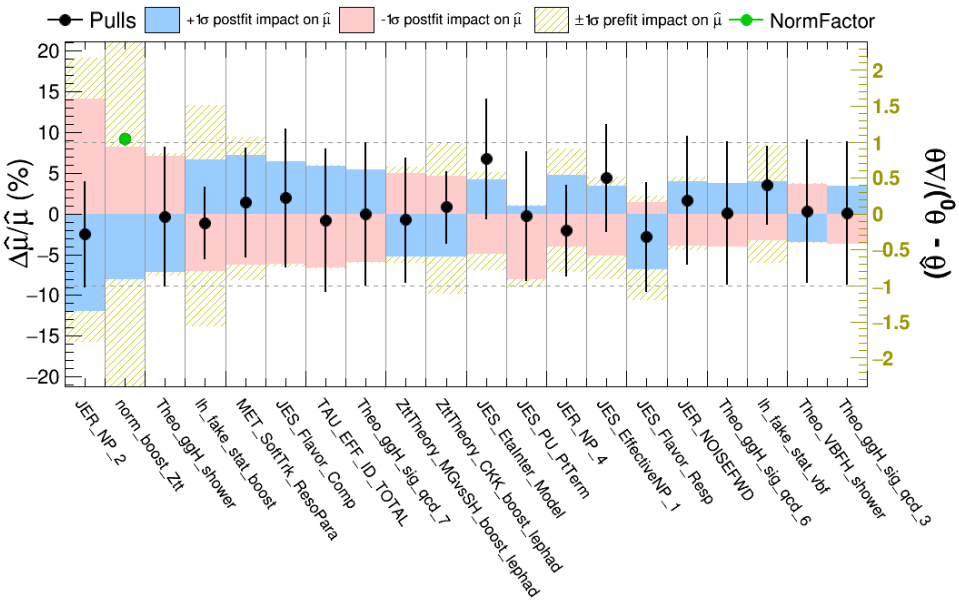
Table 6.18.: Uncertainty breakdown on μ for Standalone $\tau_{lep}\tau_{had}$ VBF category fit.

As further cross-check on the fit results comparison between Asimov and Unblinded fit, it is possible to check if in both cases the same set of nuisance parameters have a similar impact on the quoted signal strength. The nuisance parameter rankings with respect to μ are shown in Fig. 6.31-6.33 for combined Boost+VBF and Boosted/VBF only fit categories, respectively. In the plots, only the first 20 highly ranked nuisance parameters are reported, since these have the largest impact on μ; the full list of the nuisance parameters which are used in the fit is reported in Table B.1. Looking at the plots, it is possible to conclude that:

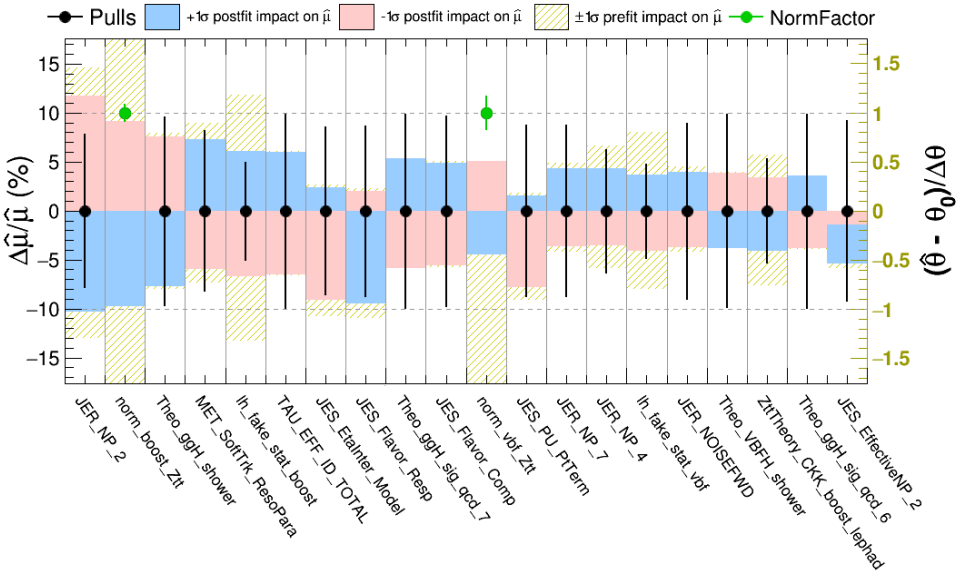
- rankings are similar between Asimov and Unblinded fit, confirming that the analysis is pretty solid against possible side effects which are not well modelled in MC;

- systematics uncertainties related to the Jet and E_T^{miss} as well as theoretical uncertainties have a large impact on μ; this confirms what was previously observed looking at the uncertainty breakdown on μ.

Finally, another important cross-check to spot fit anomalies is to look at the shape of the negative log likelihood (NLL) profiles for the nuisance parameters which have the largest impact on μ. Due to Gaussian constraints which are used in the global profile likelihood, the NLL shape is expected to have a parabolic shape with the minimum center at 0 or at least not pulled beyond the 1 σ band shown on the right axis of the ranking plots. A large deviation from 0 will indicate a strong model dependency on a given nuisance parameter which should be investigated later. Looking at Fig. 6.31, there are no nuisance parameters which are pulled beyond 1 σ level. However, it is worth to check if the NLL shapes are really parabolic or they present some anomalies like double minima structures or kink which can bias the error estimation. The NLL profile scans for the highly ranked nuisance parameters are shown in Fig. 6.34; looking at these plots, no anomalies have been observed.

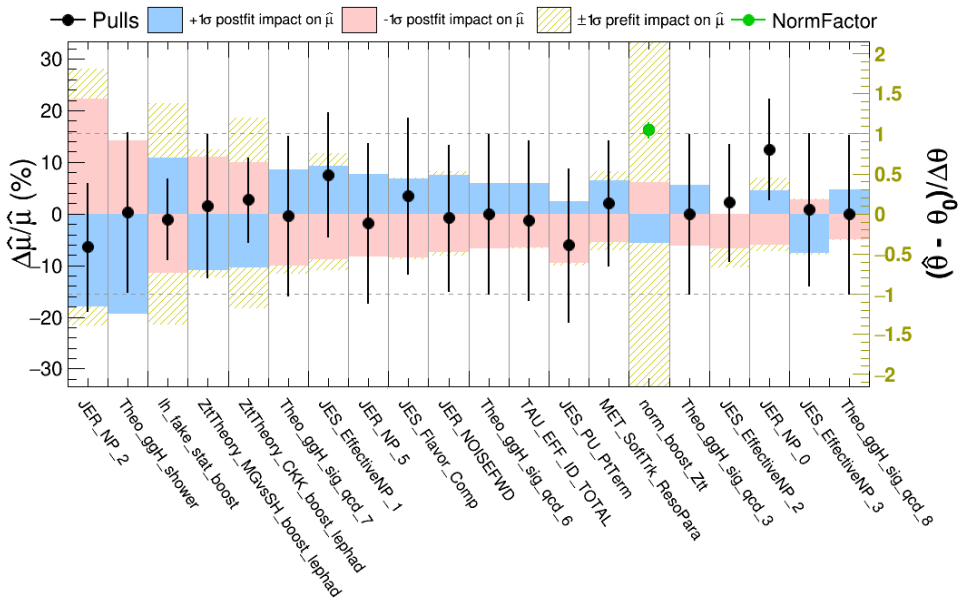


(a)

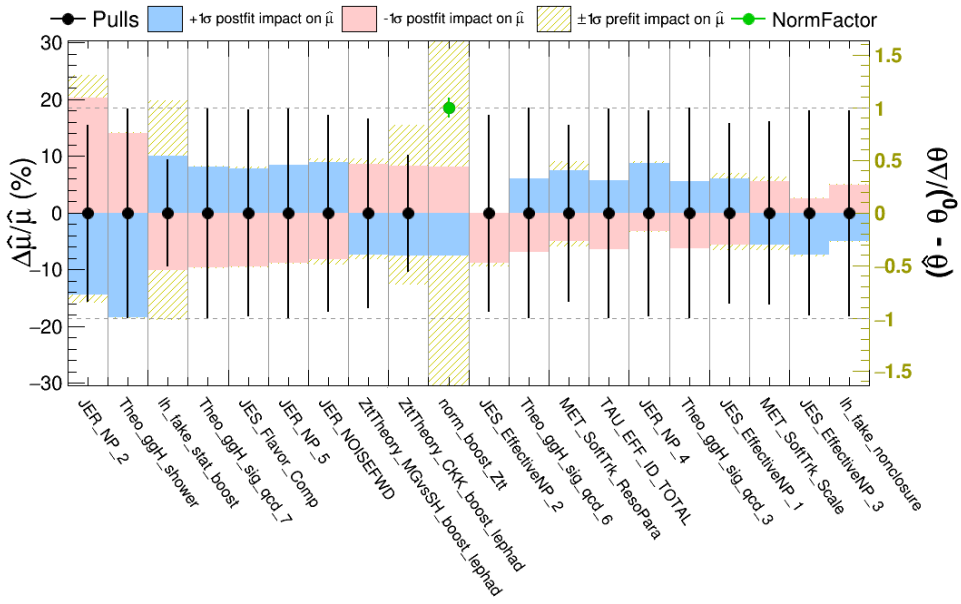


(b)

Figure 6.31.: Comparison of nuisance parameter ranking for Standalone $\tau_{lep}\tau_{had}$ Boost+VBF categories fit: *a)* Unblinded Fit, *b)* Asimov Fit.

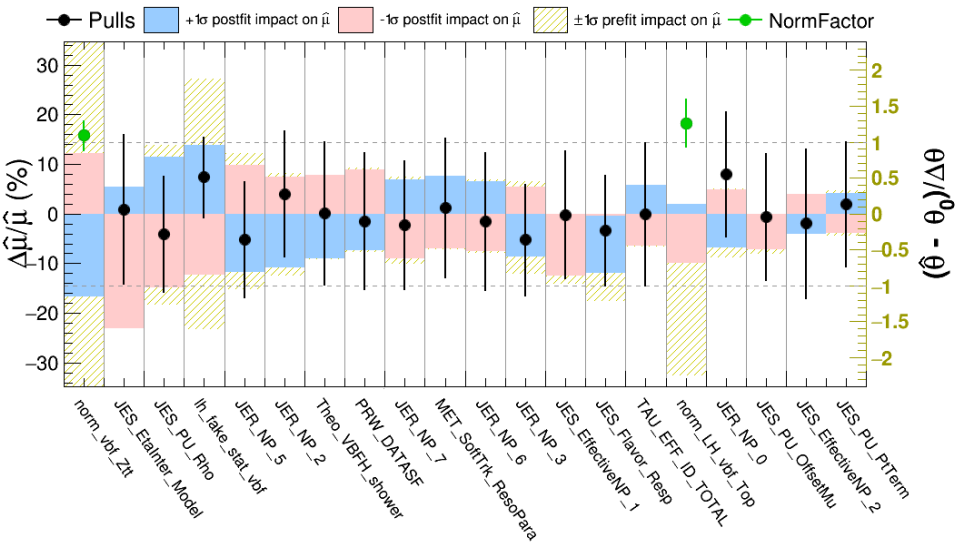


(a)

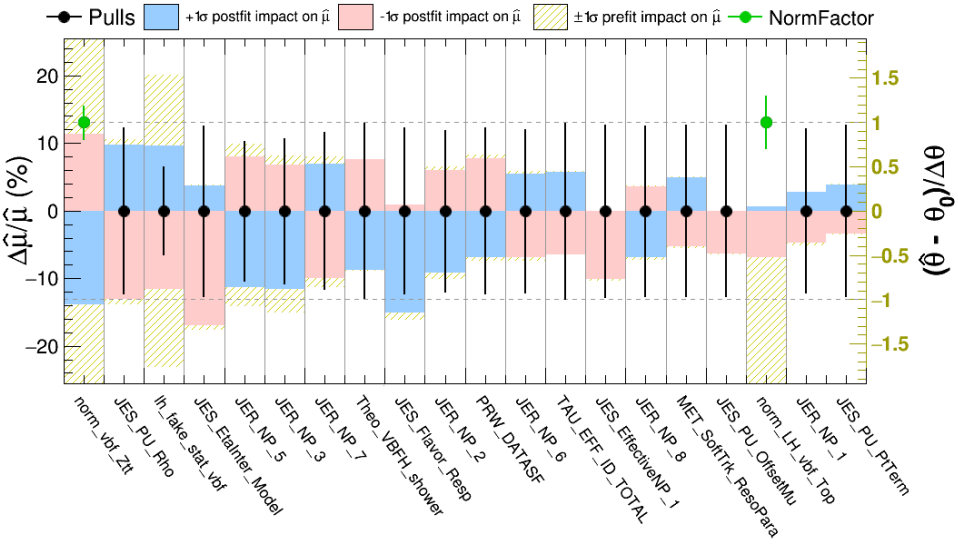


(b)

Figure 6.32.: Comparison of nuisance parameter ranking for Standalone $\tau_{lep}\tau_{had}$ Boost category fit: *a)* Unblinded Fit, *b)* Asimov Fit.



(a)



(b)

Figure 6.33.: Comparison of nuisance parameter ranking for Standalone $\tau_{lep}\tau_{had}$ VBF category fit: *a)* Unblinded Fit, *b)* Asimov Fit.

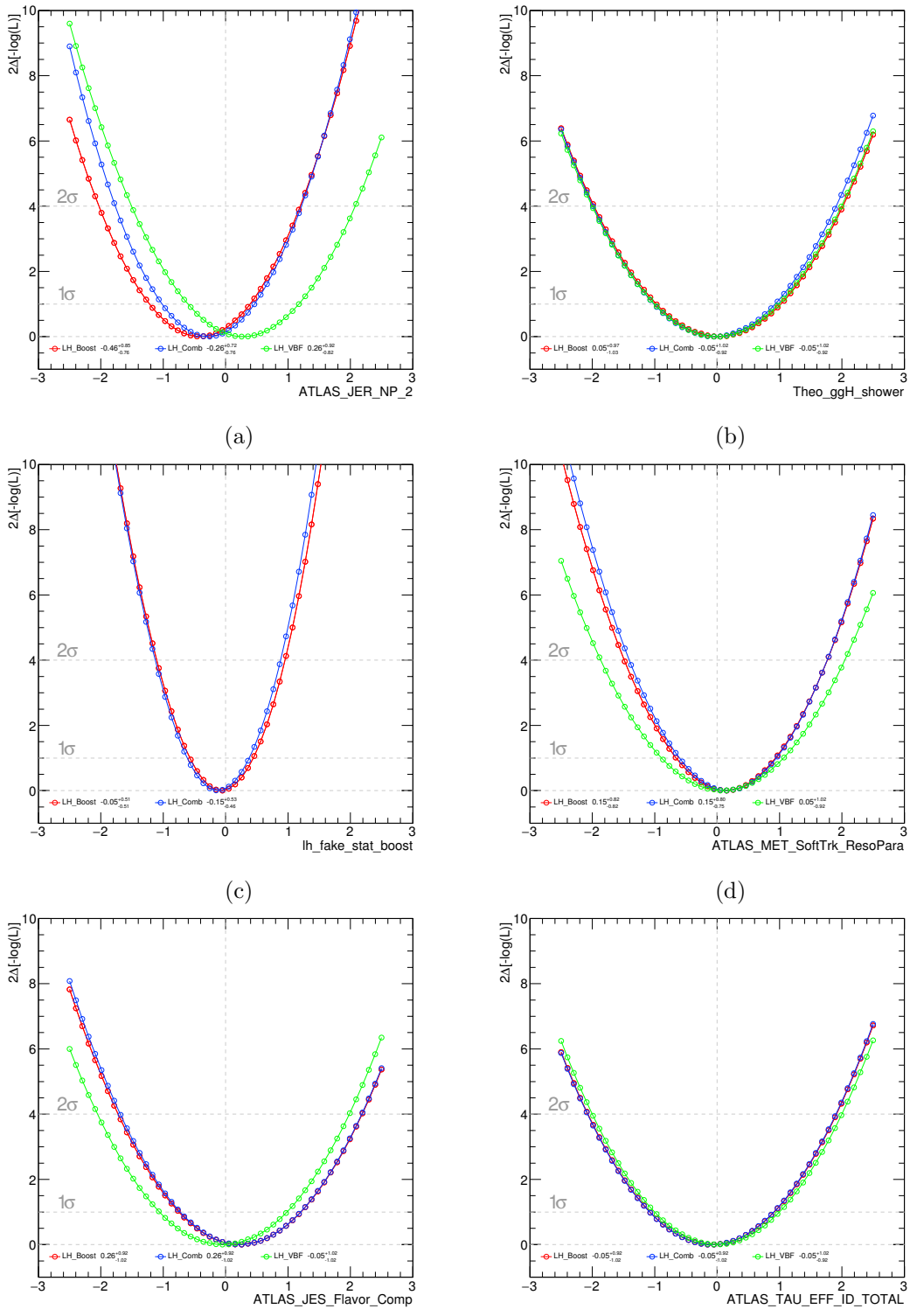


Figure 6.34.: NLL profiles scans for the highly ranked NPs in Fig. 6.31 (a).

6.9.1. Cross section extraction

After having performed all the cross-checks and verified the healthiness of the fit, it is possible to extract also the cross-sections for the ggH and the VBF production processes. This is done performing a similar fit to the one described above with two main differences:

- the fit is done considering ggH and VBF cross-sections as independent parameters in the fit, while in the fit described before the μ parameter is correlated across the different signal production processes;
- the theory total cross section uncertainties described in Table 6.11 are not considered as part of the nuisance parameters of the fit model.

In Tables 6.19-6.20 the values of the cross-sections time the $H \to \tau\tau$ branching ratio extracted from the unblinded fit and the error breakdown are reported, respectively. The expected values have been derived from the Yellow Report tables.

	Observed value (pb)	Expected value (pb)
$\sigma_{H\tau\tau}$(VBF)	$0.174^{+0.207}_{-0.193}$	$0.237^{+0.001}_{-0.001}$
$\sigma_{H\tau\tau}$(ggH)	$3.303^{+2.943}_{-2.286}$	$3.047^{+0.140}_{-0.204}$

Table 6.19.: Values of cross-section times $H \to \tau\tau$ branching ratio from standalone $\tau_{lep}\tau_{had}$ fit.

	$\sigma_{H\tau\tau}$(VBF) (pb)	$\sigma_{H\tau\tau}$(ggH) (pb)
Value	0.174	3.303
Total error	+0.207 -0.193	+2.943 -2.287
Statistical	+0.144 -0.137	+1.551 -1.527
Systematic	+0.149 -0.136	+2.501 -1.702
Theo. Systematics	+0.058 -0.040	+1.905 -0.714
Exp. Systematics	+0.139 -0.130	+1.863 1.395
Jet/E_T^{miss} Systematics	+0.090 -0.075	+1.555 -1.222
Tau Systematics	+0.026 -0.006	+0.409 -0.174
MC Statistics	+0.092 -0.097	+0.355 -0.292

Table 6.20.: Uncertainty breakdown on the VBF and ggH cross sections from standalone $\tau_{lep}\tau_{had}$ fit.

From the results reported in Table 6.19, it is possible to conclude that the measured values in agreement with expected values from SM predictions, as visible also in Fig.

6.35 which shows the two dimensional 68 % and 95 % confidence level (CL) contours in the plane ($\sigma_{H\tau\tau}$(ggH), $\sigma_{H\tau\tau}$(VBF)). However, the large associated error suggests to repeat the same measurement using the combination of the $\tau_{lep}\tau_{had}$ final state and the other di-tau decay final state, as it will be described in sec. 6.10.1.

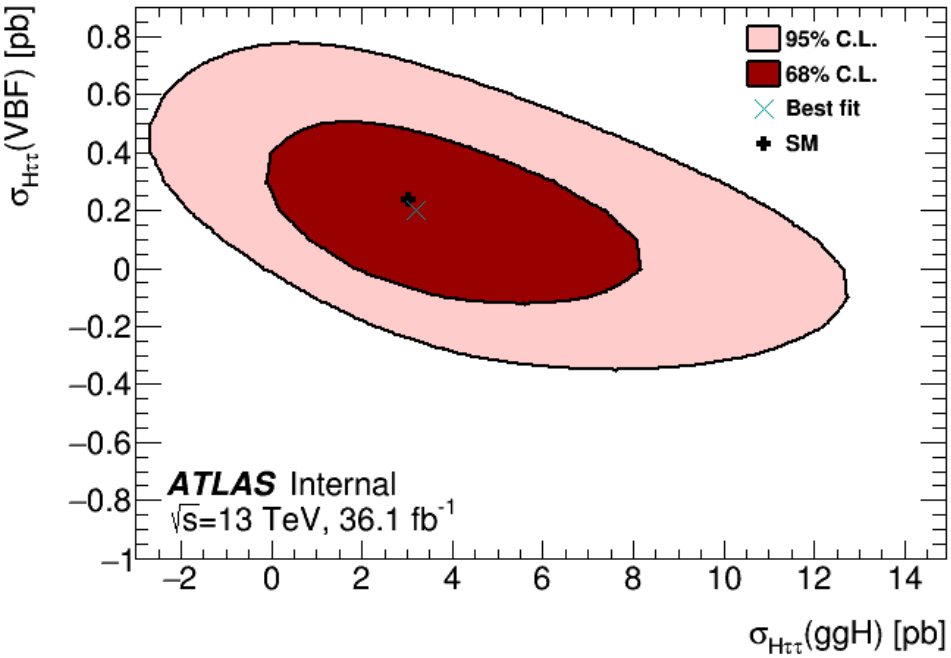


Figure 6.35.: Two dimensional 68 % and 95 % confidence level contours in the plane ($\sigma_{H\tau\tau}$(ggH), $\sigma_{H\tau\tau}$(VBF)). Results from standalone $\tau_{lep}\tau_{had}$ fit.

6.10. Results for Combined Fit

Similarly to what has been presented in the previous section, a fit was also performed combining all the different $H \to \tau\tau$ decay channels. In Table 6.21, fit results are reported for Asimov/Unblinded fit on Boosted+VBF and Boosted/VBF only fit categories including all the the decay channels. Results for the single decay channels are reported in Table 6.22, while a comparison between observed Run 2 analysis significance and observed Run 1 significance analysis is shown in Table 6.23; looking at the table, it should be mentioned that in Run 1 two types of analyses have been performed: one similar to the one which was described in this thesis, *cut-based* (CBA), and another one which was relying on multivariate analysis technique (*MVA*) to get the best possible value for the significance. These results show that Run 2 CBA analysis is performing much better than Run 1 CBA analysis and it is comparable with Run 1 MVA.

	Fit type / Categ.	Combined	Boost	VBF
Exp. (Obs.) Significance	Asimov	4.482 (4.482)	2.934 (2.934)	2.945 (2.945)
	Unblinded	4.359 (4.127)	3.168 (2.839)	2.602 (2.731)
Signal strength	Asimov	$1^{+0.28}_{-0.25}$	$1^{+0.42}_{-0.36}$	$1^{+0.42}_{-0.37}$
	Unblinded	$1.08^{+0.32}_{-0.28}$	$1.14^{+0.47}_{-0.39}$	$0.98^{+0.47}_{-0.41}$

Table 6.21.: Comparison of the significance/signal strength for Asimov/Unblinded Combined Boost+VBF categories fit.

	Fit type / Categ.	$\tau_{lep}\tau_{lep}$	$\tau_{lep}\tau_{had}$	$\tau_{had}\tau_{had}$
Exp. (Obs.) Significance	Asimov	1.198 (1.198)	2.813 (2.813)	2.839 (2.839)
	Unblinded	2.202 (1.196)	2.425 (2.627)	2.206 (2.851)
Signal strength	Asimov	$1^{+0.83}_{-0.83}$	$1^{+0.45}_{-0.38}$	$1^{+0.44}_{-0.37}$
	Unblinded	$1.97^{+0.83}_{-0.81}$	$0.90^{+0.44}_{-0.38}$	$0.86^{+0.57}_{-0.44}$

Table 6.22.: Breakdown of the significance for each $\tau\tau$ final state.

	VBF			Boost		
	$\tau_{lep}\tau_{lep}$	$\tau_{lep}\tau_{had}$	$\tau_{had}\tau_{had}$	$\tau_{lep}\tau_{lep}$	$\tau_{lep}\tau_{had}$	$\tau_{had}\tau_{had}$
Run 2 CBA	1.29	1.65	1.39	1.49	1.89	1.26
Run 1 CBA	0.66	1.59	1.41	0.39	0.72	0.87
Run 1 MVA	1.15	2.11	1.70	0.57	1.11	0.82

Table 6.23.: Significance comparison between Run 1 and Run 2 analyses.

Similarly to what has been done for the standalone $\tau_{lep}\tau_{had}$ fit, the error quoted in Table 6.22 for the signal strength can be decomposed to check which source of systematic uncertainty is having more impact on the global error. Tables 6.24-6.26 show the uncertainty breakdown on μ for combined Boost+VBF and Boosted/VBF only fit categories, respectively. Results in these tables confirm what has been previously discussed in Section 6.9 about the source of uncertainty which have the largest impact on the signal strenght.

POI Value	1.085
Total error	+0.317 -0.281
Statistical	+0.175 -0.171
Systematic	+0.265 -0.222
Theo. Systematics	+0.169 -0.113
Exp. Systematics	+0.204 -0.191
Jet/E_T^{miss} Systematics	+0.125 -0.102
Tau Systematics	+0.050 -0.034
MC Statistics	+0.121 -0.110

Table 6.24.: Uncertainty breakdown on μ for Combined Boost+VBF categories fit.

POI Value	1.145
Total error	+0.467 -0.393
Statistical	+0.229 -0.226
Systematic	+0.407 -0.321
Theo. Systematics	+0.300 -0.179
Exp. Systematics	+0.288 -0.239
Jet/E_T^{miss} Systematics	+0.185 -0.153
Tau Systematics	+0.076 -0.061
MC Statistics	+0.112 -0.099

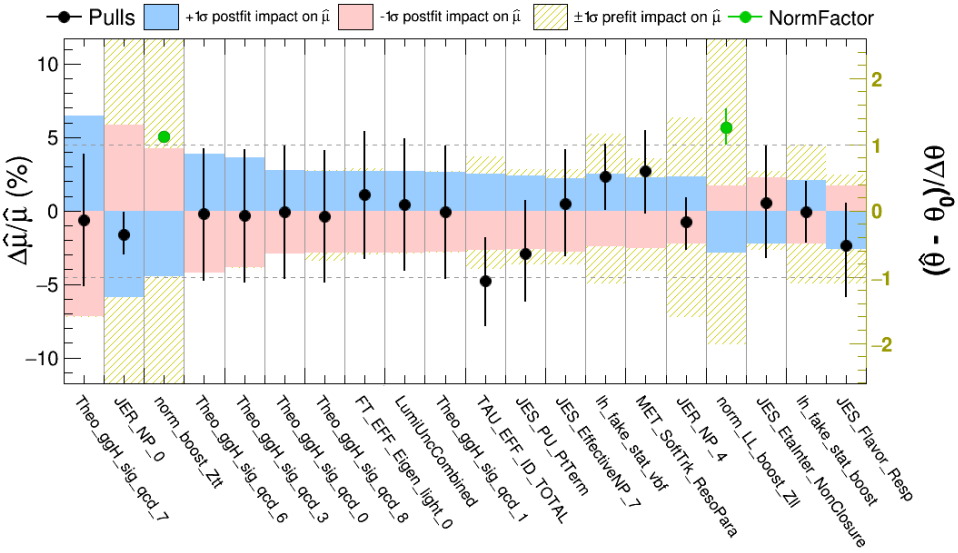
Table 6.25.: Uncertainty breakdown on μ for Combined Boost category fit.

POI Value	0.983
Total error	+0.475 -0.414
Statistical	+0.261 -0.249
Systematics	+0.396 -0.33
Theo. Systematics	+0.140 -0.095
Exp. Systematics	+0.351 -0.305
Jet/E_T^{miss} Systematics	+0.284 -023
Tau Systematics	+0.108 -0.053
MC Statistics	+0.226 -0.207

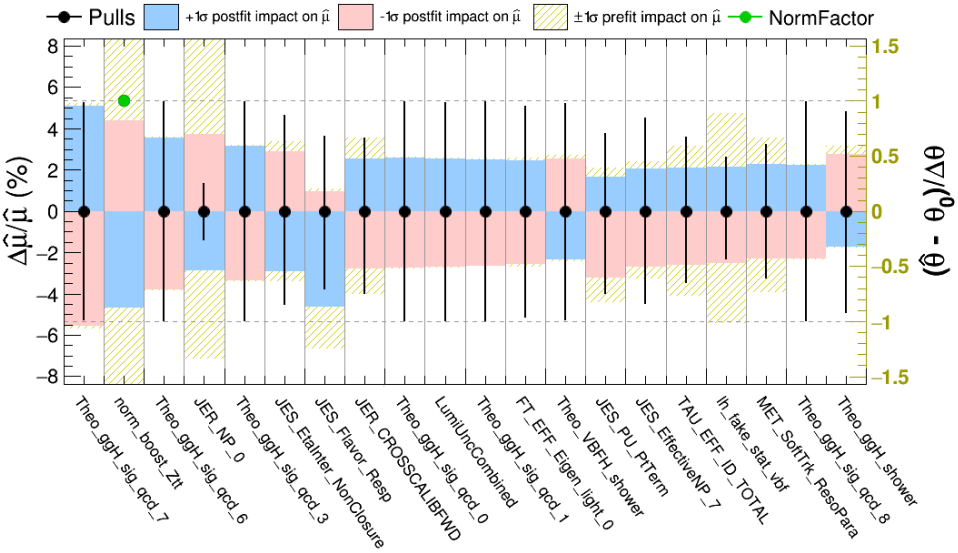
Table 6.26.: Uncertainty breakdown on μ for Combined VBF category fit.

The nuisance parameter rankings with respect to μ are shown in Fig. 6.37-6.38 for the combined Boost+VBF and Boosted/VBF only fit categories, respectively. In the plots, only the first 20 highly ranked nuisance parameters are reported, since these have the

largest impact on μ The plots further confirm what has been discussed looking at the breakdown of the uncertainties on μ.

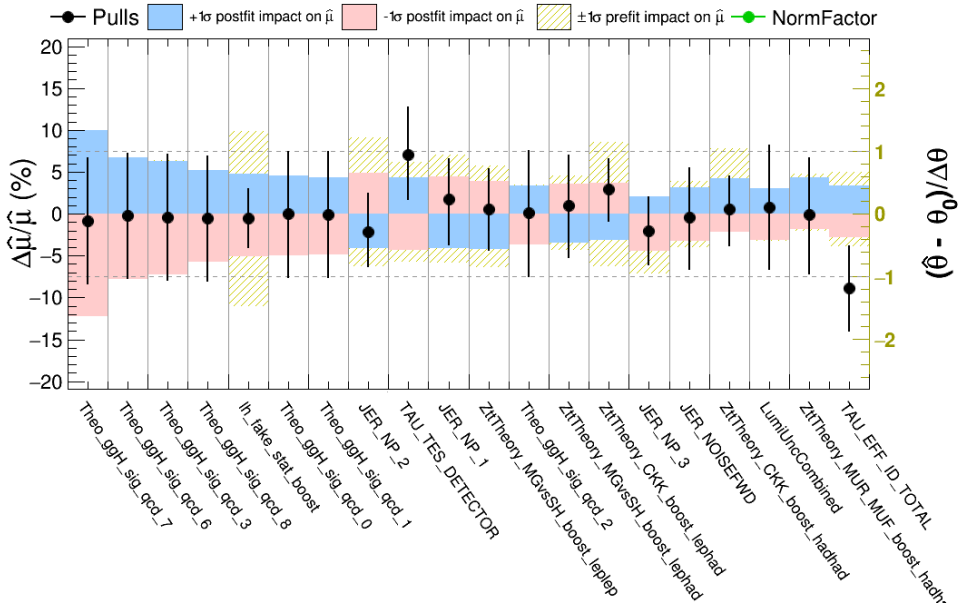


(a)

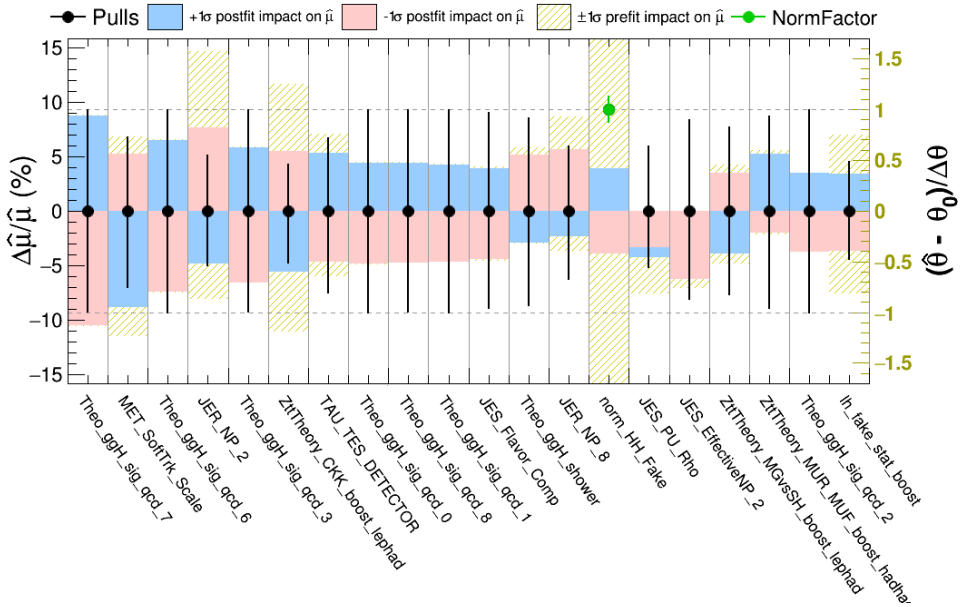


(b)

Figure 6.36.: Comparison of nuisance parameter ranking for Combined Boost+VBF categories fit: *a)* Unblinded Fit, *b)* Asimov Fit.

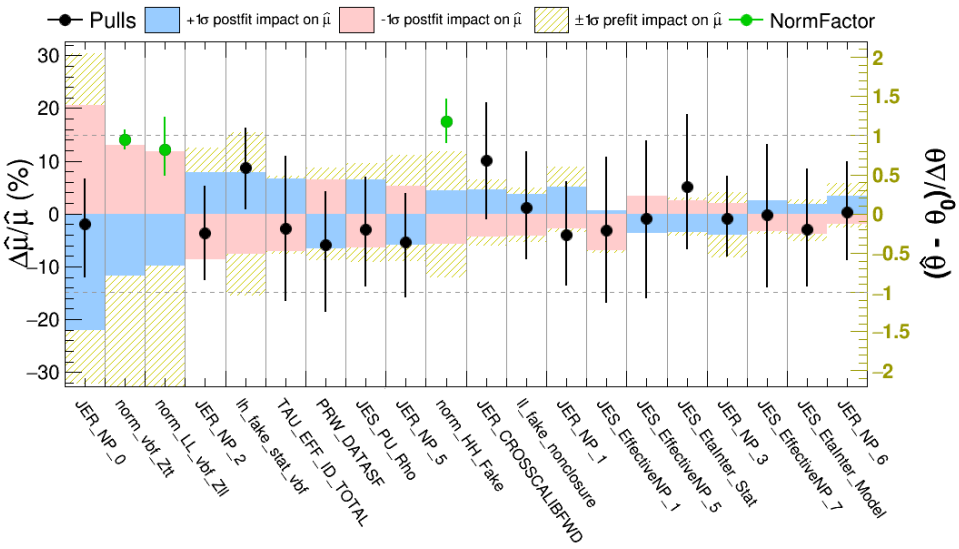


(a)

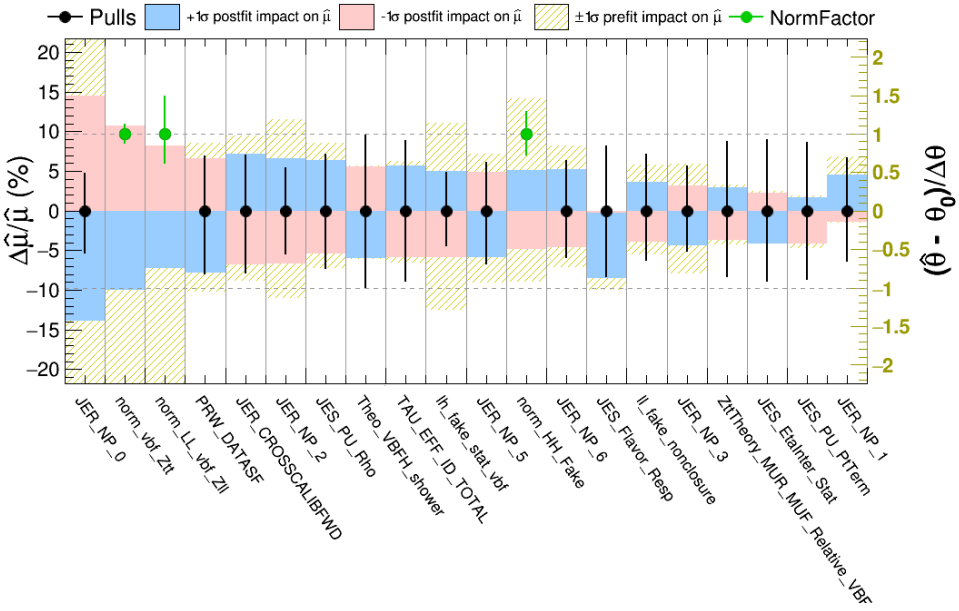


(b)

Figure 6.37.: Comparison of nuisance parameter ranking for Combined Boost category fit: *a)* Unblinded Fit, *b)* Asimov Fit.



(a)



(b)

Figure 6.38.: Comparison of nuisance parameter ranking for Combined VBF category fit: *a)* Unblinded Fit, *b)* Asimov Fit.

6.10.1. Cross section extraction

Using the same fit procedure described in Section 6.9.1, the values of the ggH and VBF cross sections have been extracted from the combined Boost+VBF fit categories. In Tables 6.27-6.28 the values of the cross-sections time the $H \to \tau\tau$ branching ratio extracted from the unblinded fit and the error breakdown are reported, respectively.

	Value (pb)	Expected value (pb)
$\sigma_{H\tau\tau}$(VBF)	$0.285^{+0.138}_{-0.127}$	$0.237^{+0.001}_{-0.001}$
$\sigma_{H\tau\tau}$(ggH)	$2.969^{+1.959}_{-1.596}$	$3.047^{+0.140}_{-0.204}$

Table 6.27.: Values of cross-section times $H \to \tau\tau$ branching ratio for Combined Boost+VBF categories fit.

	$\sigma_{H\tau\tau}$(VBF) (pb)	$\sigma_{H\tau\tau}$(ggH) (pb)
Value	0.285	2.969
Total error	+0.138 -0.127	+1.959 -1.596
Statistical	+0.092 -0.088	+1.027 -1.015
Systematic	+0.102 -0.091	+1.668 -1.232
Theo. Systematics	+0.047 -0.033	+1.233 -0.573
Exp. Systematics	+0.090 -0.085	+1.123 -1.091
Jet/E_T^{miss} Systematics	+0.063 -0.053	+0.775 -0.604
Tau Systematics	+0.024 -0.016	+0.271 -0.266
MC Statistics	+0.064 -0.060	+0.574 -0.517

Table 6.28.: Uncertainty breakdown on the VBF and ggH cross sections for Combined Boost+VBF categories fit.

From the results reported in Table 6.27, it is possible to conclude that the measured values are in agreement with the expected values from SM predictions, as also visible from the contour plot in Fig. 6.39; while the measured uncertainties are quite larger with respect to the theoretical uncertainties for the VBF production process, in the case of the ggH production process the two uncertainties have similar values. Thus the analysis results can contribute to reduce the theoretical uncertainty associated to the ggH cross section.

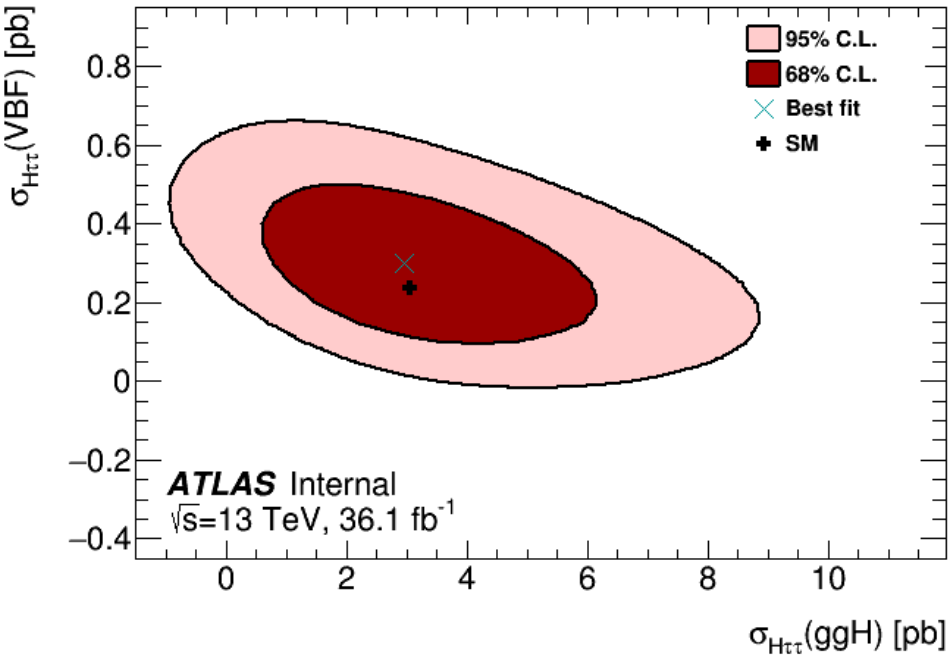


Figure 6.39.: Two dimensional 68 % and 95 % confidence level contours in the plane ($\sigma_{H\tau\tau}$(ggH), $\sigma_{H\tau\tau}$(VBF)). Results from Combined Boost+VBF categories fit.

6.10.2. Combination with Run 1 analysis

The significance and the signal strength values reported in Table 6.21 strongly suggest the observation of the $H \to \tau\tau$ decay and it confirms that the Higgs boson behaves as it is predicted by the SM. However, the observed significance is still below the 5 standard deviations which is the threshold used in the scientific community to claim the observation of a process. To confirm the observation, the current analysis has been combined with the Run 1 MVA analysis; the combination has been done:

- updating the signal cross section values for 7 TeV and 8 TeV center of mass energy to their most updated values;
- keeping all the nuisance parameters in the fit model not correlated between the two analyses.

The choice to not correlate the nuisance parameters between the two analyses is due to two main reasons:

- the object (jet, muon, E_T^{miss}) related nuisance parameter scheme provided by each Combined Performance group is different between Run 1 and Run 2, so it is difficult to trace back the same source of uncertainty and correlate it;

- the estimate of the main irreducible background, $Z \to \tau\tau$, is based on MC estimation in Run 2 analysis, while in Run 1 it was estimated through a data-driven technique called *Embedding* [96].

The significance values for the combination of the Run 1 and Run 2 analyses are reported in Table 6.29; the combination gives an observed (expected) significance value of 6.37 (5.43) standard deviation, from which it is possible to claim the observation of the $H \to \tau\tau$ decay with the ATLAS detector.

Significance	Run 1	Run 2	Combination
Expected	3.43	4.13	5.43
Observed	4.54	4.36	6.37

Table 6.29.: Significance values for the Run 1 and Run 2 analyses combination.

CHAPTER 7

Analysis improvements

In this chapter two possible analysis improvements will be discussed: the usage of the E_T^{miss} significance to improve the rejection of the $Z \to ll$ background and the addition of a low electron/muon p_T region through the usage of the TLT trigger.

7.1. Usage of E_T^{miss} Significance to reject $Z \to ll$ background

An important background in the $H \to \tau\tau$ analysis in the $\tau_{lep}\tau_{had}$ final state is given by $Z \to ll$+jets ($l = e, \mu$) events. These events can pass analysis selection criteria if an electron/muon is wrongly reconstructed as a τ; dedicated rejection methods have been developed by the Tau performance group and their usage has already strongly reduced this background. However, it would be good if the analysis would be capable of further reducing this source of background, since the MMC calculation pushes it in the region around the $H \to \tau\tau$ mass peak, as can be seen in Fig. 6.25 and 7.7.
A good handle which can improve the $Z \to ll$ rejection is given by the E_T^{miss} significance; this is a significance that is based on the uncertainties of all objects that enter the calculation of the E_T^{miss}, from hard objects to the track-based soft term, and it should provide a variable that is more powerful at rejecting backgrounds (like $Z \to ll$) where fake E_T^{miss} comes from mismeasured objects. Unfortunately, the E_T^{miss} significance calculation was developed by the ATLAS collaboration after the production of the final samples which were used to derive all the results described in chapter 6; so before inserting it in the analysis selection criteria, a preliminary study has been done to quantify the potential impact of the E_T^{miss} significance. In Fig. 7.1, correlation plots between MMC/E_T^{miss} and E_T^{miss} significance are shown; looking at these plots, it is possible to conclude that:

- the E_T^{miss} significance is linearly correlated to E_T^{miss}, as it is expected because events with real E_T^{miss} should have higher values of the E_T^{miss} significance;

- $Z \to ll$ events which are located in the 100 GeV $\leq$ MMC $\leq$ 150 GeV region have lower values of the E_T^{miss} significance with respect to events with real E_T^{miss} in the same MMC region.

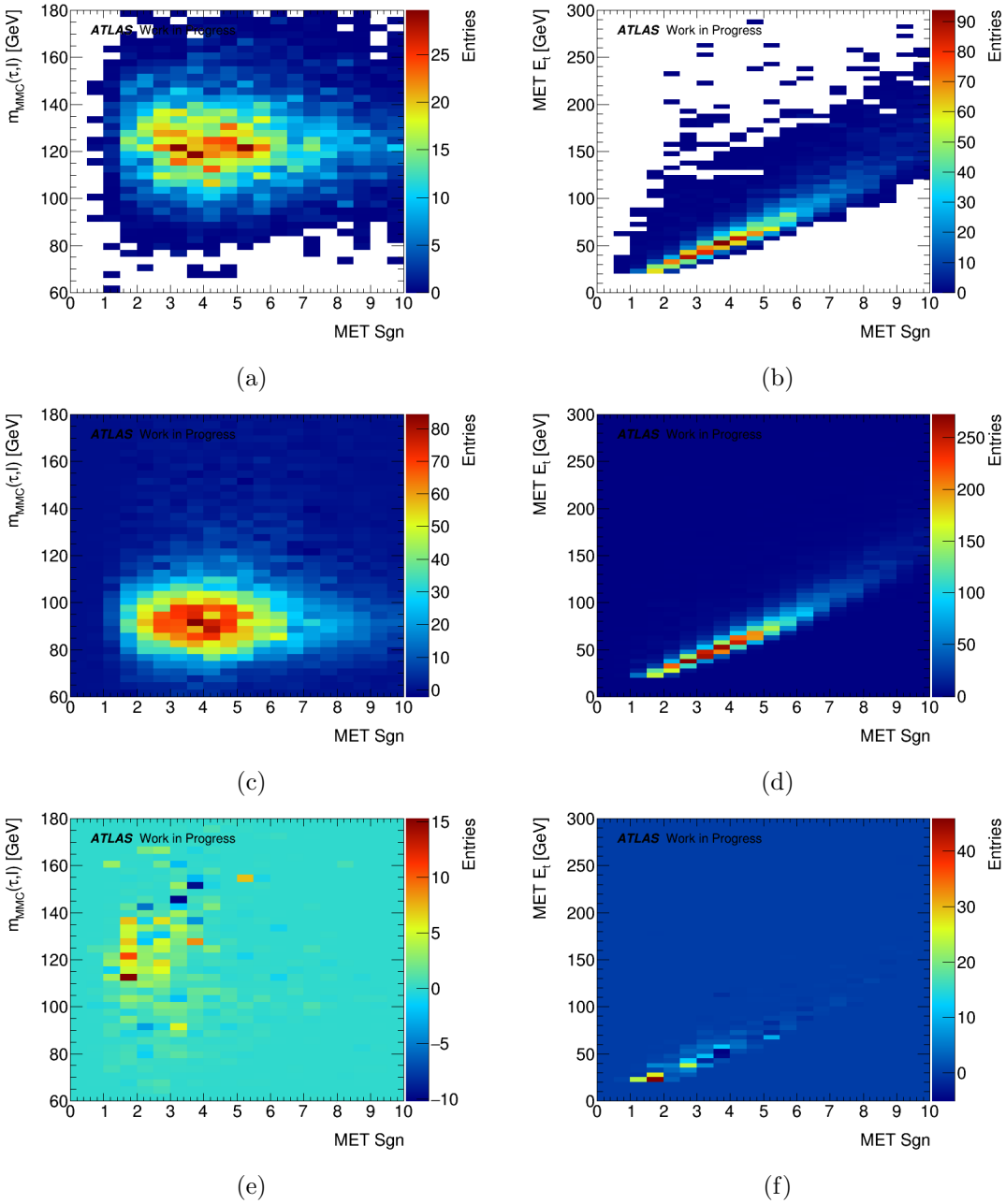
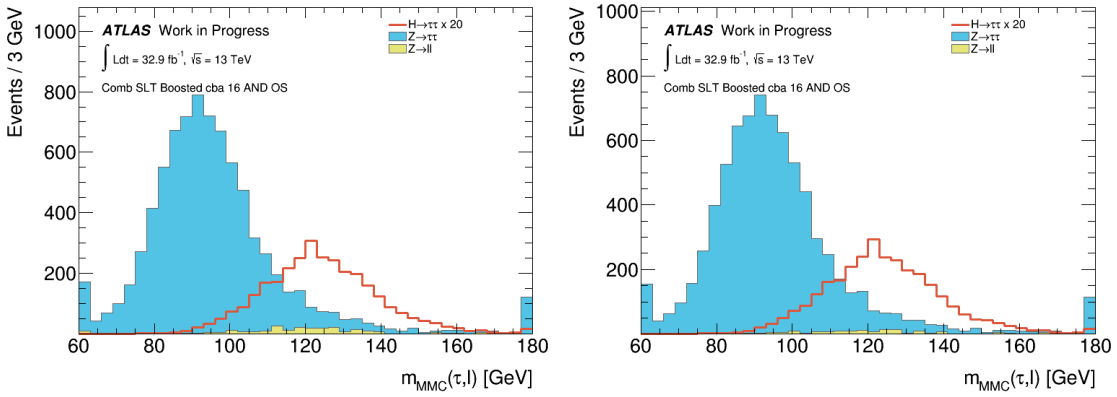


Figure 7.1.: MMC vs E_T^{miss} significance correlation plots for: *a)* $H \to \tau\tau$, *c)* $Z \to \tau\tau$, *e)* $Z \to ll$. E_T^{miss} vs E_T^{miss} significance correlation plots for: *b)* $H \to \tau\tau$, *d)* $Z \to \tau\tau$, *f)* $Z \to ll$.

To check the impact of cutting on E_T^{miss} significance in the analysis selection, a cut E_T^{miss} significance ≥ 2 has been inserted on top of the Boosted region definition which has been defined in section 6.5.2. In Table 7.1 and in Fig. 7.2, the yields and the MMC distribution for $H/Z \to \tau\tau$ and $Z \to ll$ before/after the E_T^{miss} significance cut are reported, respectively.
These results show that the usage of the E_T^{miss} significance can lead to region reduction of 40 % of $Z \to ll$ background in the 100 GeV $\leq$ MMC $\leq$ 150 GeV, while it has marginal impact on the signal. Same studies have also been performed in the VBF region, but no strong $Z \to ll$ reduction has been observed; however, it is worth to mention that the $Z \to ll$ background is already a negligible background in the VBF region with respect to the Boosted region.

Sample	No E_T^{miss} Sgn cut	E_T^{miss} Sgn ≥ 2	Eff. (%)
Zll	244.08 $\pm$ 35.10	138.77 $\pm$ 31.67	57
Ztt	7861.29 $\pm$ 71.91	7453.08 $\pm$ 70.02	95
VBF	23.45 $\pm$ 0.23	22.21 $\pm$ 0.23	95
ggH	131.97 $\pm$ 2.14	125.63 $\pm$ 2.07	95

Table 7.1.: Impact of the E_T^{miss} significance in the Boosted region.



(a) (b)

Figure 7.2.: MMC distribution in the Boosted region: *a)* no E_T^{miss} Sgn cut, *b)* E_T^{miss} Sgn ≥ 2.

7.2. Addition of a low electron/muon p_T region using TLT trigger

As mentioned in section 6.3, the $H \to \tau\tau$ analysis in the $\tau_{lep}\tau_{had}$ final state uses only single lepton triggers (SLT) to select events. However, as shown in Fig. 6.2, it is possible

to use τ+lepton triggers (TLT) to select events in which electron/muon p_T is below the thresholds imposed by SLT. In this section, a study about analysis feasibility using only TLT selected events and 2016 dataset will be presented; to ensure the orthogonality with the SLT region, a cut on muon (electron) $p_T < 27.3$ (27) GeV has been imposed. The analysis uses the same background estimation method, the *Fake-Factor* method, as well as similar event categorisation in Boosted and VBF region as used in the SLT analysis presented in shapter 6; also background control regions are defined in a similar way.
Before entering in the selection criteria for the signal regions, it is worth to mention what are the advantages/drawbacks of performing a study in this region; these are summarised in Table 7.2.

Advantages	Low contamination from W+jets, since electron/muon p_T well below Jacobian peak good rejection of QCD due to not balanced lepton p_T and τ p_T
Drawbacks	Additional L1 jet in the trigger, which imposes leading jet $p_T \geq 70$ GeV and leading jet $\|\eta\| \leq 3.2$ trigger has already an online τ identification requirement, which reduces statistics in the anti-τ regions for Fake Factor calculation

Table 7.2.: Advantages/Limitations of a TLT based analysis.

In Table 7.3 the triggers used for TLT analysis are reported.

Trigger menu	Data period	Chain name (in the menu)
τ+electron	2016	HLT_e17_lhmedium_nod0_ivarloose_tau25_medium1_tracktwo
τ+muon	2016	HLT_mu14_ivarloose_tau25_medium1_tracktwo

Table 7.3.: Triggers used for TLT analysis.

7.2.1. VBF selection

The VBF selection is designed to enrich signal events from the Higgs vector boson fusion production process. The requirements are :

- at least 2 jets in the event with $p_T \geq 30$ GeV;
- the pseudo-rapidity difference between the two leading jets must be $|\Delta\eta|_{jj} \geq 2.0$;
- the invariant mass of the two leading jets must be $M_{jj} > 250$ GeV;
- the product of pseudo-rapidities of the two jets must be negative $\eta_{j1} \times \eta_{j2} < 0$;
- both leptonic and hadronic tau candidates must lie between the two leading jets in pseudorapidity (*centrality*).

Comparing these selection criteria with the one reported in section 6.5.1, this study is using a looser selection in order to get more statistics in this region and there is no further event categorisation in the *Tight/Loose* regions. In Tables 7.4 and 7.5, event yields in VBF signal region and relative control regions are reported, while in Fig. 7.3-7.6 the modelling of the main variables used to define this signal region are shown. In general, good Data/MC agreement and good modelling has been observed for all the considered variables.

Sample/Region	$\tau_e\tau_{had}$ SR	$\tau_\mu\tau_{had}$ SR
Fake	207.6 ± 11.0	464.1 ± 15.7
VV	6.2 ± 0.8	10.3 ± 0.9
Top	4.8 ± 1.0	8.7 ± 1.4
Zll	16.0 ± 4.3	4.5 ± 3.1
Ztt	242.0 ± 15.4	463.6 ± 20.1
Total Bkg	476.6 ± 19.4	951.3 ± 25.8
ggH	3.0 ± 0.2	5.3 ± 0.3
VBF	5.3 ± 0.1	9.0 ± 0.1
VH	0.1 ± 0.1	0.2 ± 0.1
Total Sgn	8.4 ± 0.3	14.5 ± 0.3
Data	502.0 ± 22.4	1040.0 ± 32.2

Table 7.4.: Yields in VBF signal region.

Sample/Region	$\tau_e\tau_{had}$ QCD CR	$\tau_\mu\tau_{had}$ QCD CR	$\tau_e\tau_{had}$ Top CR	$\tau_\mu\tau_{had}$ Top CR	$\tau_e\tau_{had}$ W CR	$\tau_\mu\tau_{had}$ W CR
Fake	123.8 ± 8.1	318.5 ± 13.6	26.4 ± 4.0	48.2 ± 5.3	68.3 ± 5.8	115.5 ± 7.7
VV	0.3 ± 0.1	0.2 ± 0.1	0.7 ± 0.1	0.9 ± 0.1	5.7 ± 0.7	6.9 ± 0.8
Top	1.1 ± 0.5	1.3 ± 0.5	92.1 ± 6.7	98.6 ± 5.8	11.5 ± 2.1	14.6 ± 2.0
Zll	1.0 ± 0.4	-0.6 ± 1.1	-0.2 ± 0.2	0.6 ± 0.3	2.5 ± 1.0	2.9 ± 1.6
Ztt	16.0 ± 3.1	33.8 ± 5.3	1.3 ± 0.6	3.9 ± 2.3	19.4 ± 11.2	40.2 ± 6.4
Total Bkg	142.3 ± 8.7	353.2 ± 14.6	120.4 ± 7.9	152.3 ± 8.2	107.5 ± 12.8	180.2 ± 10.4
ggH	0.4 ± 0.1	0.2 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.3 ± 0.1	0.6 ± 0.1
VBF	0.2 ± 0.1	0.5 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.7 ± 0.1	1.1 ± 0.1
VH	0.2 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
Total Sgn	0.6 ± 0.1	0.7 ± 0.2	0.0 ± 0.0	0.1 ± 0.1	1.1 ± 0.2	1.8 ± 0.2
Data	159.0 ± 12.6	321.0 ± 17.9	121.0 ± 11.0	184.0 ± 13.6	150.0 ± 12.2	194.0 ± 13.9

Table 7.5.: Yields in the *QCD/Top* and *W* CRs at VBF selection level.

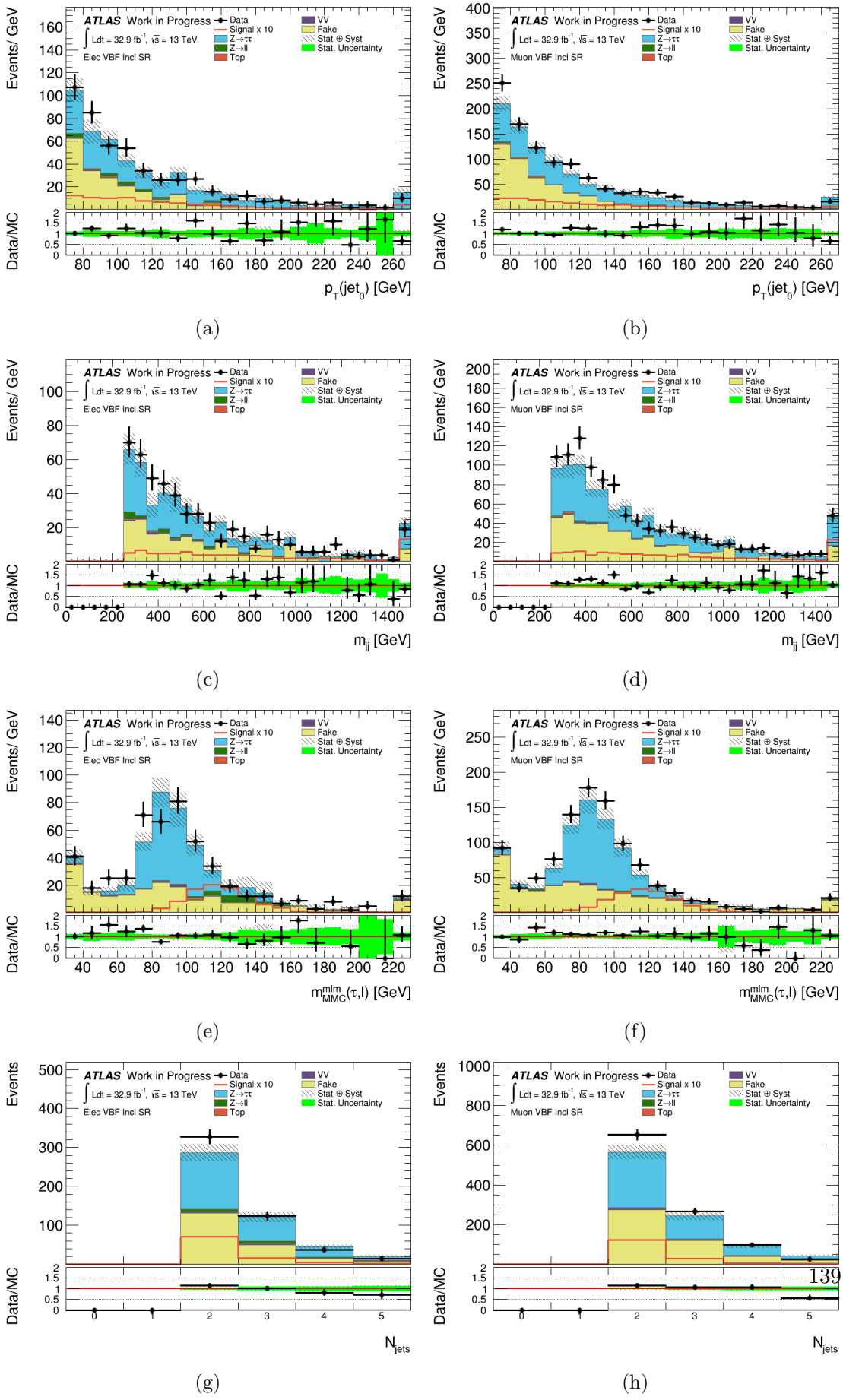


Figure 7.3.: Distributions in VBF signal region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* lead jet p_T, *c) (d)* M_{jj}, *e) (f)* MMC, *g) (h)* jet multiplicity.

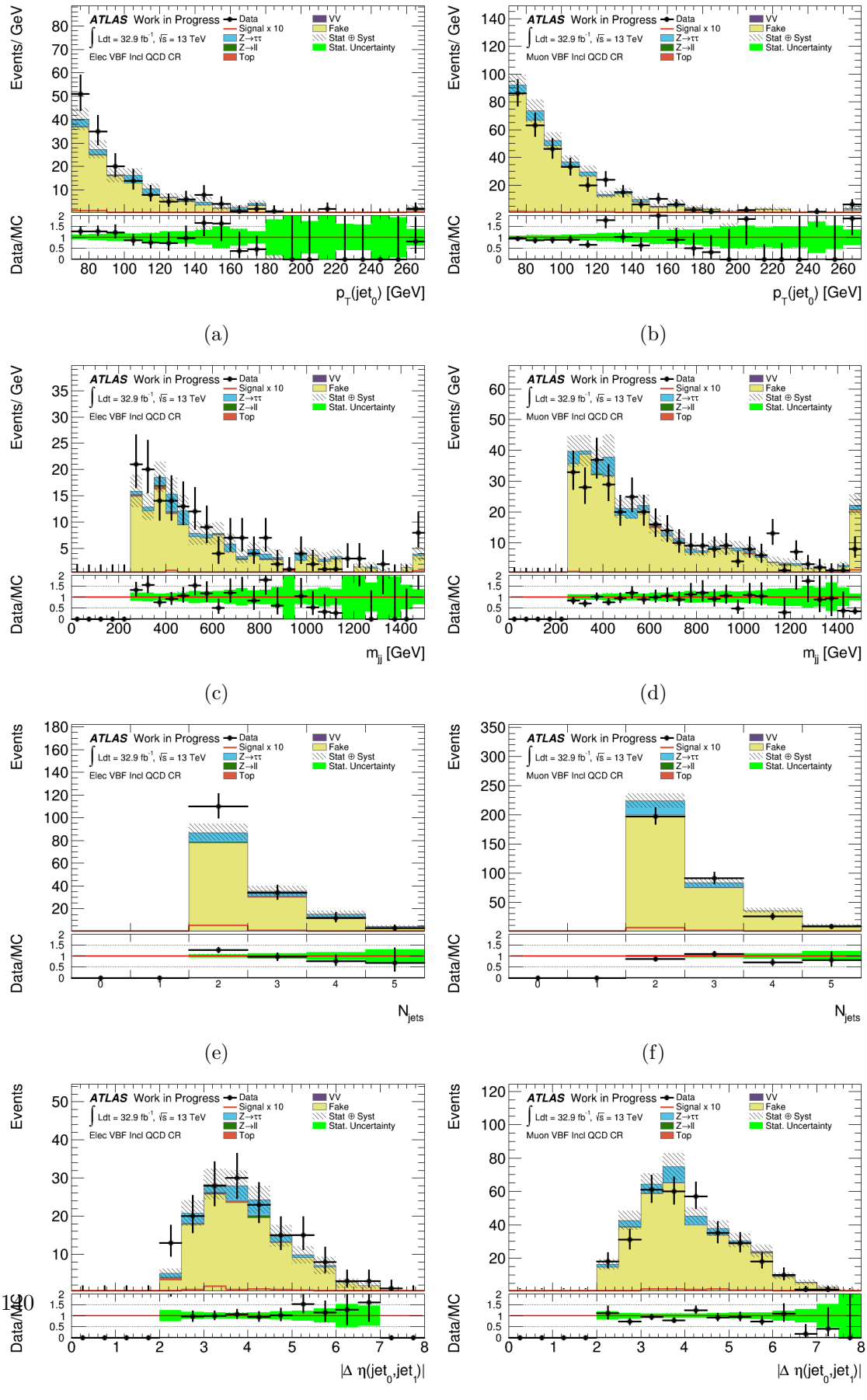


Figure 7.4.: Distributions for QCD CR for VBF region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* lead jet p_T, *c) (d)* M_{jj}, *e) (f)* jet multiplicity, *g) (h)* $\Delta\eta(jj)$.

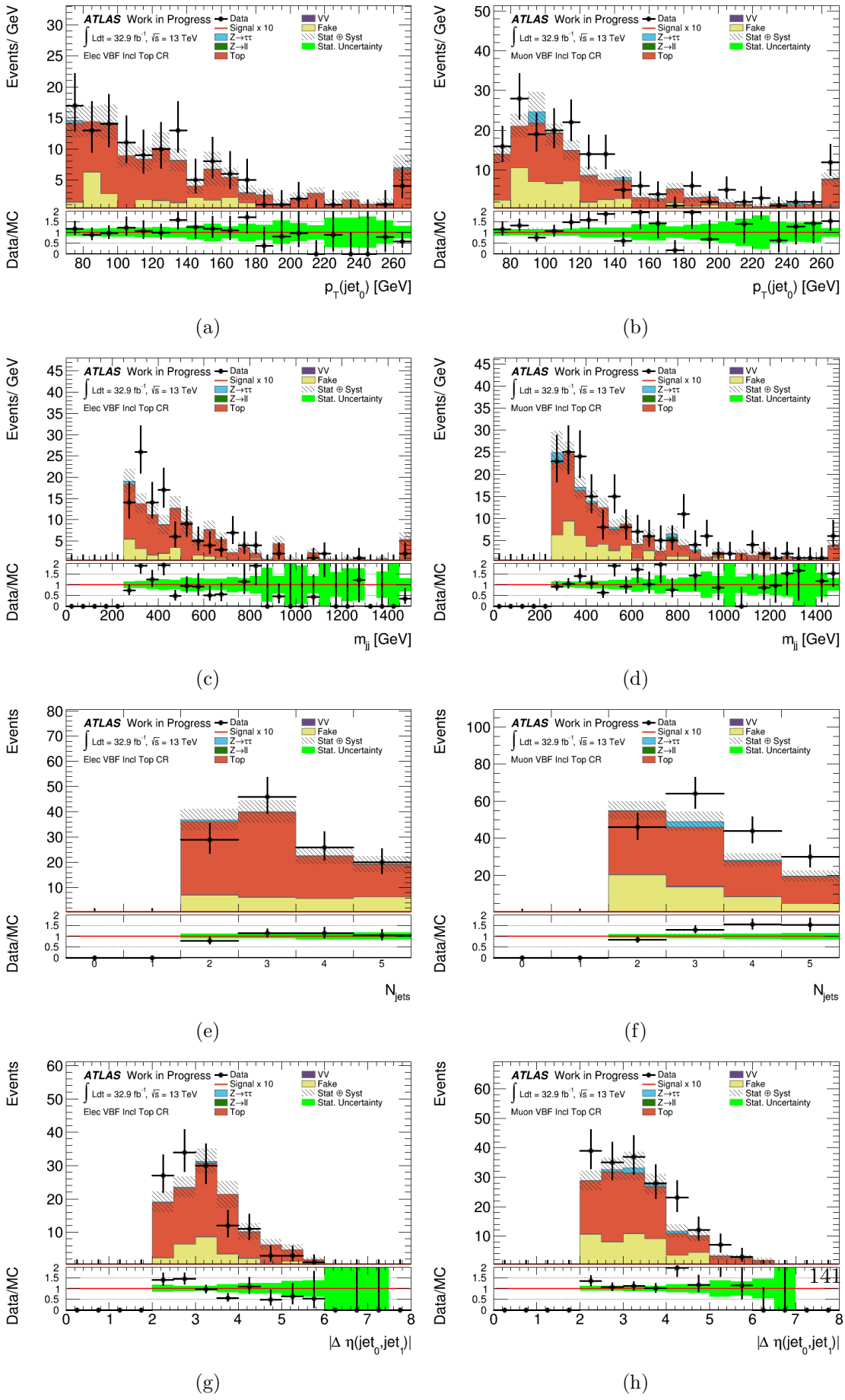


Figure 7.5.: Top CR distributions in VBF region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* lead jet p_T, *c) (d)* M_{jj}, *e) (f)* jet multiplicity, *g) (h)* $\Delta\eta(jj)$.

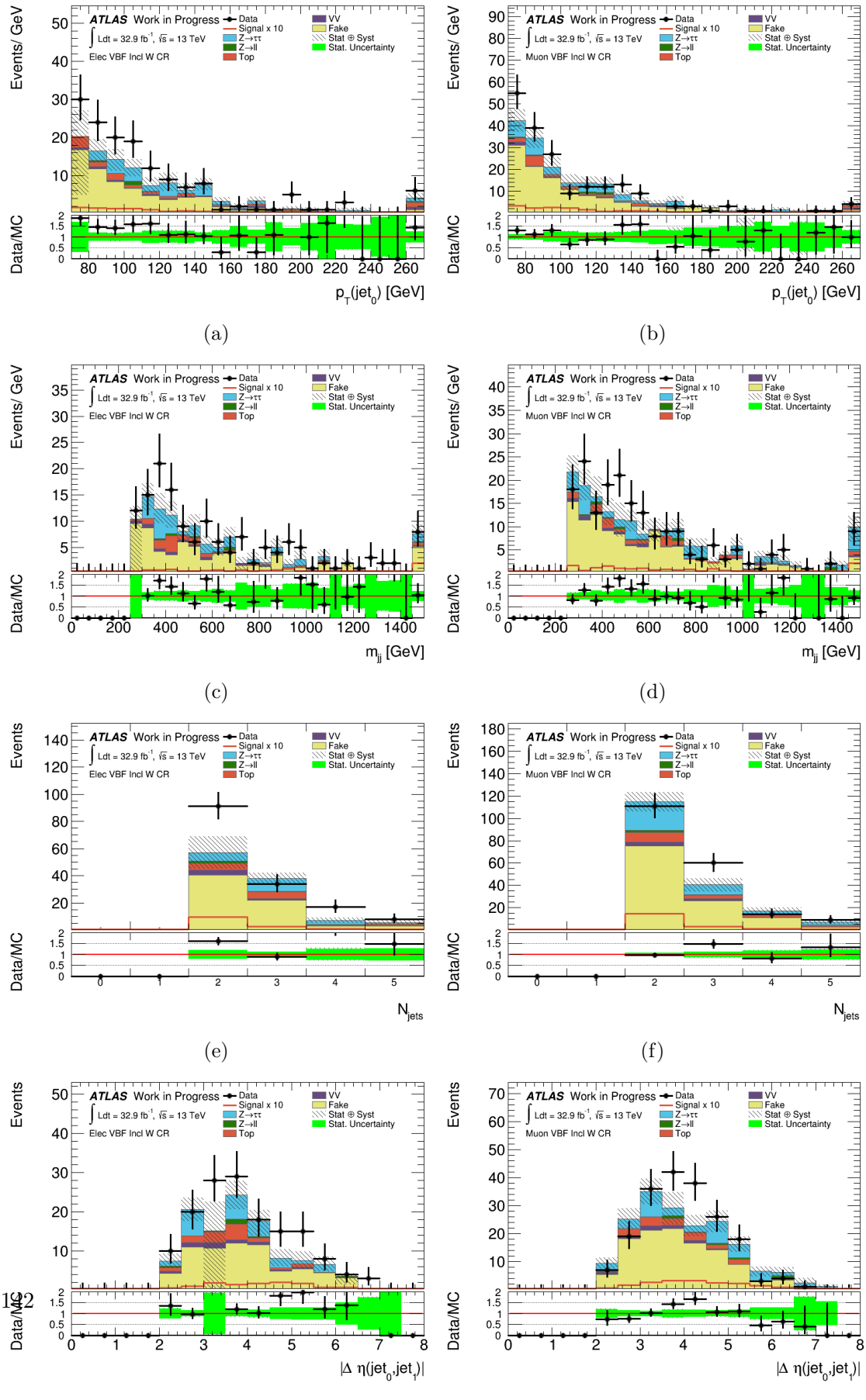


Figure 7.6.: Distributions for WCR in VBF region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* lead jet p_T, *c) (d)* M_{jj}, *e) (f)* jet multiplicity, *g) (h)* $\Delta\eta(jj)$.

7.2.2. Boosted selection

The *Boosted* selection enriches signal from Higgs boson production with large p_T, typically in the gluon-gluon fusion process if a hard jet is produced together with the Higgs boson. The Boosted region is defined by the following requirements:

- pass all the requirements of the preselection, but not VBF selection;
- the transverse momentum of the Higgs boson $p_T^H \geq 100$ GeV.

Comparing these selection criteria with the one reported in section 6.5.2, this study uses a looser selection in order to get more statistics in this region and there is no further event categorisation in *Tight/Loose* regions. In Tables 7.6 and 7.7, event yields in VBF signal region and relative control regions are reported, while in Fig. 7.7-7.10 the modelling of the main variables used to define this signal region are shown. In general, good Data/MC agreement and good modelling has been observed for all the considered variables.

Sample/Region	$\tau_e\tau_{had}$ SR	$\tau_\mu\tau_{had}$ SR
Fake	1603.6 ± 33.4	3249.9 ± 43.7
VV	51.0 ± 2.0	91.3 ± 3.3
Top	45.7 ± 4.1	67.0 ± 4.9
Zll	279.4 ± 36.3	88.1 ± 13.3
Ztt	2673.4 ± 51.0	5034.4 ± 71.0
Total Bkg	4653.1 ± 71.1	8530.7 ± 84.6
ggH	22.5 ± 0.6	41.3 ± 0.8
VBF	4.3 ± 0.1	7.6 ± 0.1
VH	2.7 ± 0.1	4.3 ± 0.1
Total Sgn	29.4 ± 0.6	53.2 ± 0.9
Data	4507.0 ± 67.1	8904.0 ± 94.4

Table 7.6.: Yields in Boosted region.

Sample/Region	$\tau_e\tau_{had}$ QCD CR	$\tau_\mu\tau_{had}$ QCD CR	$\tau_e\tau_{had}$ Top CR	$\tau_\mu\tau_{had}$ Top CR	$\tau_e\tau_{had}$ W CR	$\tau_\mu\tau_{had}$ W CR
Fake	996.8 ± 23.4	1891.3 ± 32.0	207.3 ± 12.4	318.6 ± 15.1	661.7 ± 18.6	1184.1 ± 24.3
VV	3.2 ± 0.5	4.9 ± 0.6	7.7 ± 0.8	14.9 ± 1.5	55.6 ± 2.5	83.4 ± 3.1
Top	5.9 ± 1.3	6.9 ± 1.8	660.8 ± 17.1	956.9 ± 19.7	71.4 ± 4.7	104.0 ± 5.8
Zll	3.2 ± 9.4	3.9 ± 3.0	5.2 ± 3.9	6.8 ± 2.7	60.3 ± 14.9	53.4 ± 11.3
Ztt	134.8 ± 10.7	298.5 ± 19.3	9.9 ± 3.3	16.2 ± 4.9	250.0 ± 18.3	421.9 ± 28.3
Total Bkg	1143.8 ± 27.4	2205.5 ± 37.6	891.0 ± 21.8	1313.5 ± 25.5	1098.9 ± 30.5	1846.8 ± 39.6
ggH	1.3 ± 0.1	2.2 ± 0.2	0.1 ± 0.1	0.3 ± 0.1	3.9 ± 0.3	6.1 ± 0.3
VBF	0.2 ± 0.1	0.4 ± 0.1	0.0 ± 0.0	0.1 ± 0.1	0.7 ± 0.1	1.1 ± 0.1
VH	0.2 ± 0.1	0.2 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.4 ± 0.1	0.9 ± 0.1
Total Sgn	1.7 ± 0.2	2.8 ± 0.2	0.2 ± 0.2	0.5 ± 0.2	5.0 ± 0.3	8.1 ± 0.3
Data	1074.0 ± 32.8	2308.0 ± 48.0	860.0 ± 29.3	1422.0 ± 37.7	1069.0 ± 32.7	1842.0 ± 42.9

Table 7.7.: Yields in the *QCD*/*Top* and *W* CRs at Boosted selection level.

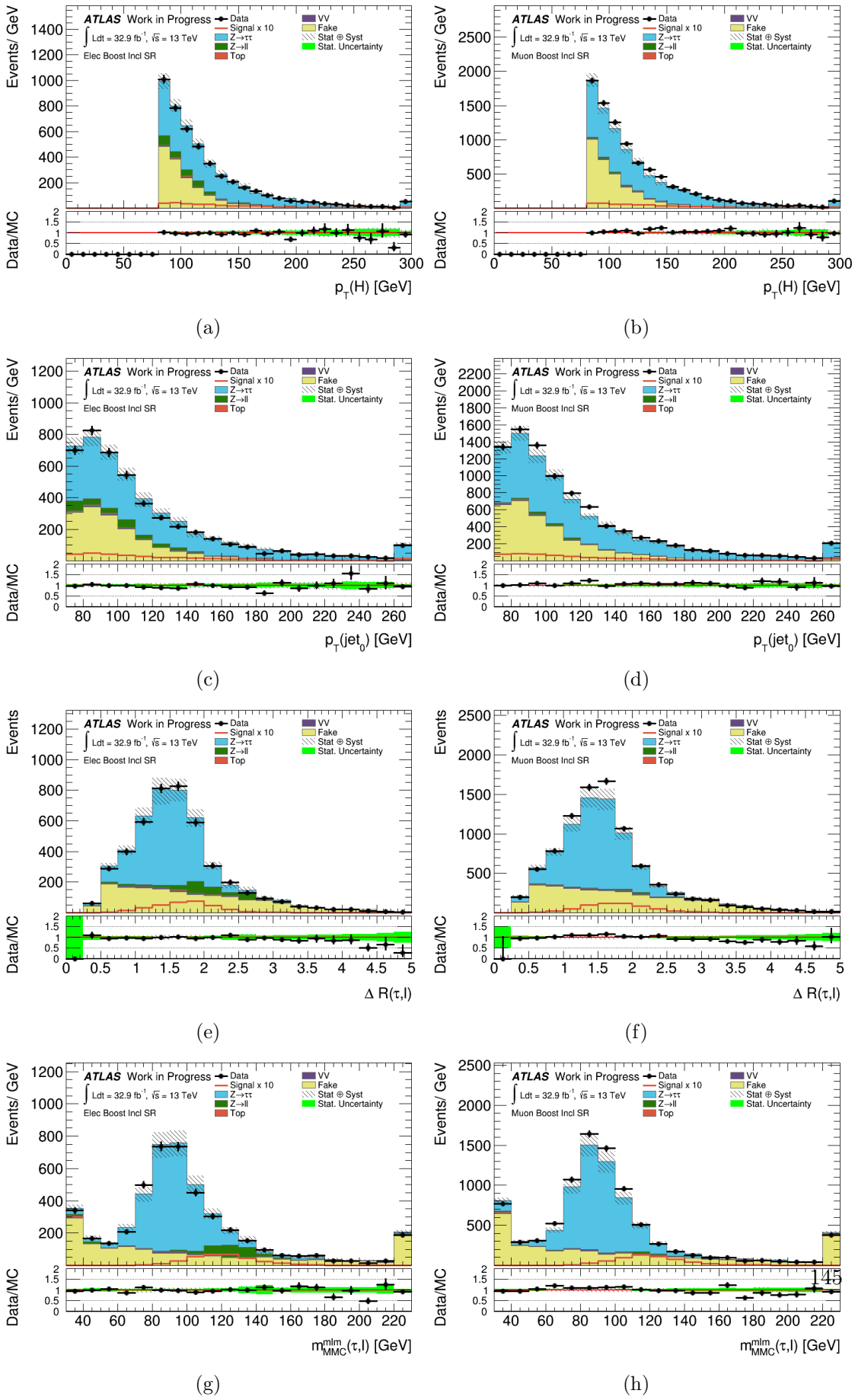


Figure 7.7.: Distributions in Boosted signal region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* Higgs p_T, *c) (d)* lead jet p_T, *e) (f)* $\Delta R(lep, \tau)$, *g) (h)* MMC.

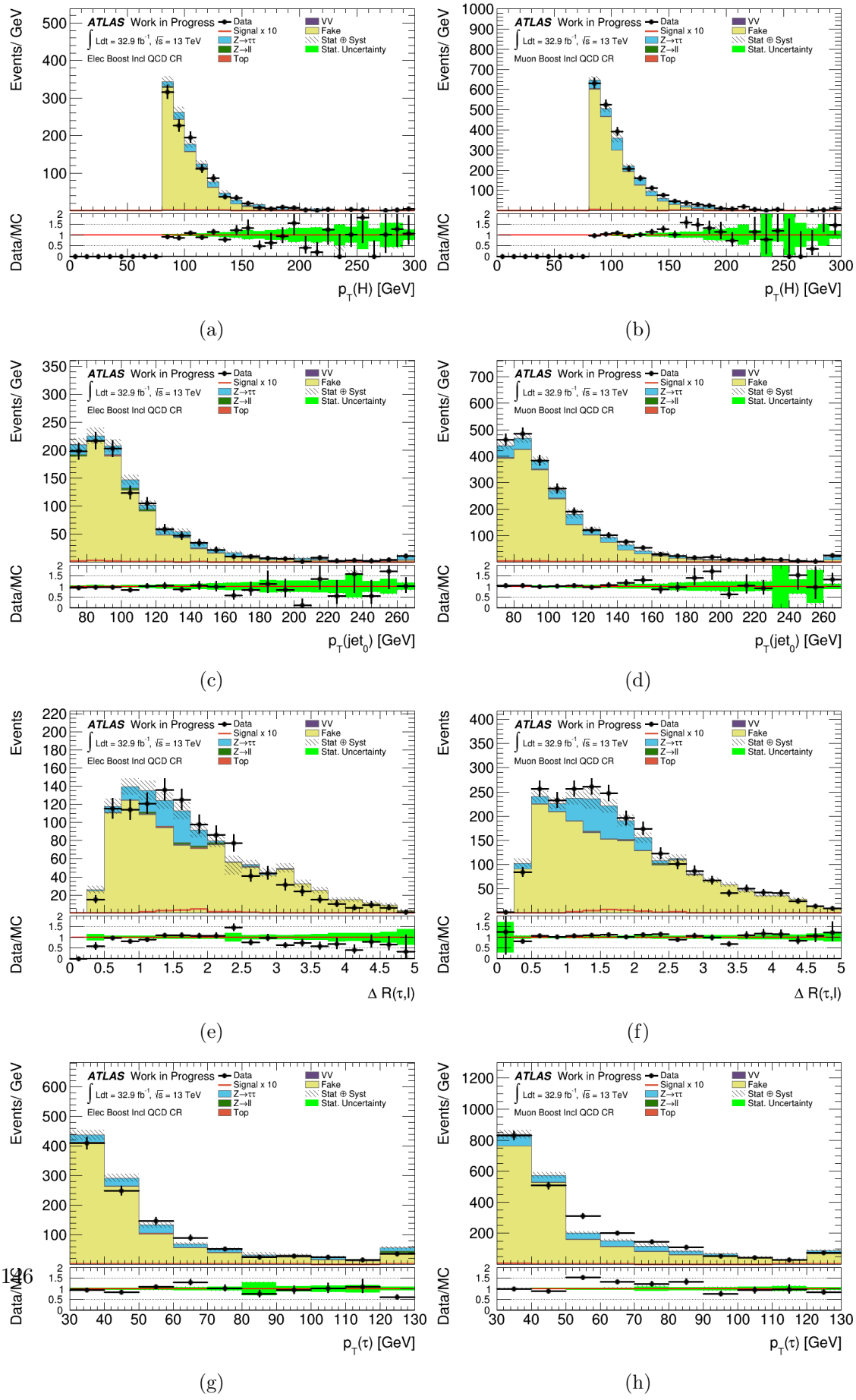


Figure 7.8.: Sistributions in QCD CR in Boosted region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* Higgs p_T, *c) (d)* lead jet p_T, *e) (f)* $\Delta R(lep, \tau)$, *g) (h)* τ p_T.

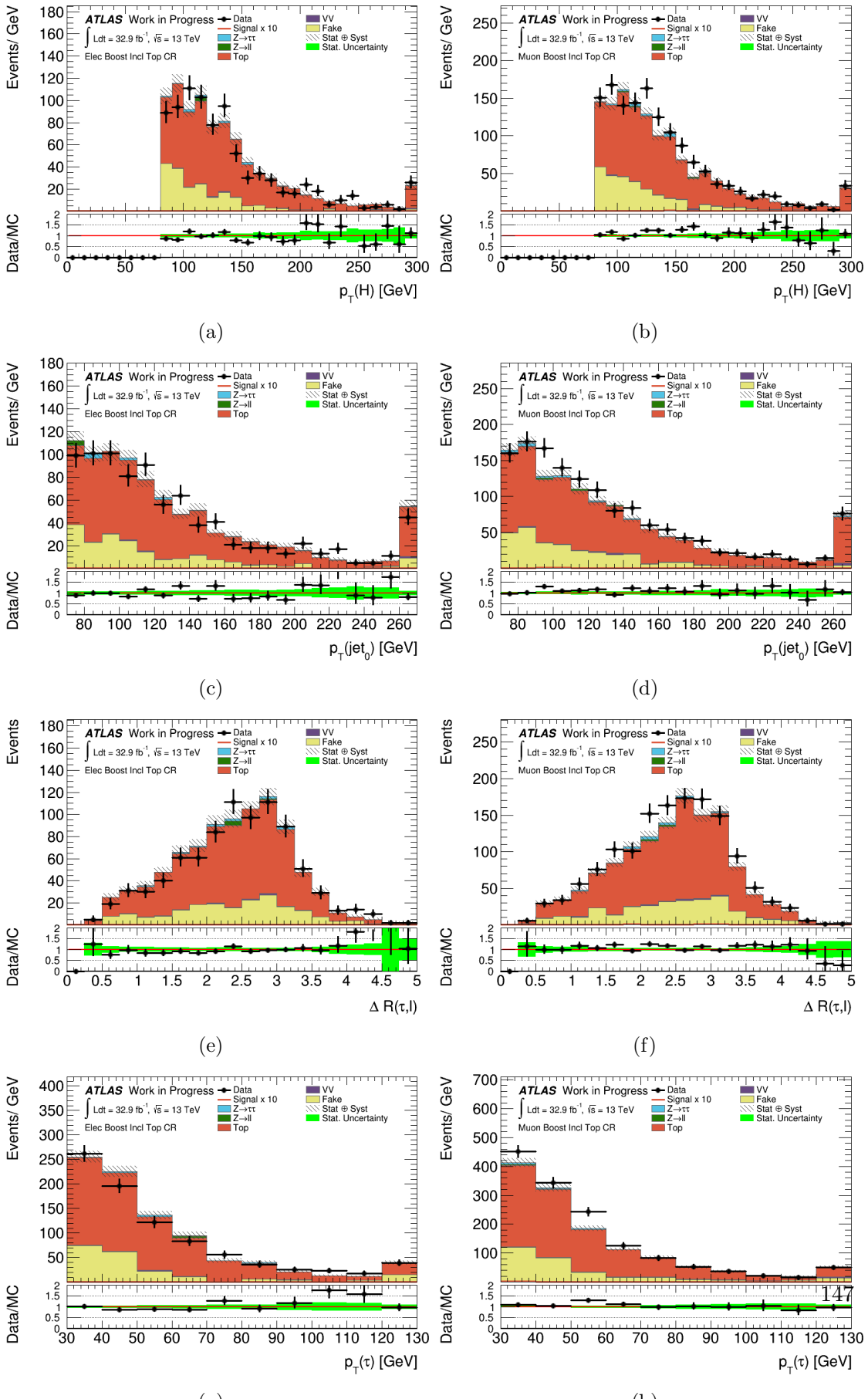


Figure 7.9.: Distributions in Top CR in Boosted region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* Higgs p_T, *c) (d)* lead jet p_T, *e) (f)* $\Delta R(lep, \tau)$, *g) (h)* τ p_T.

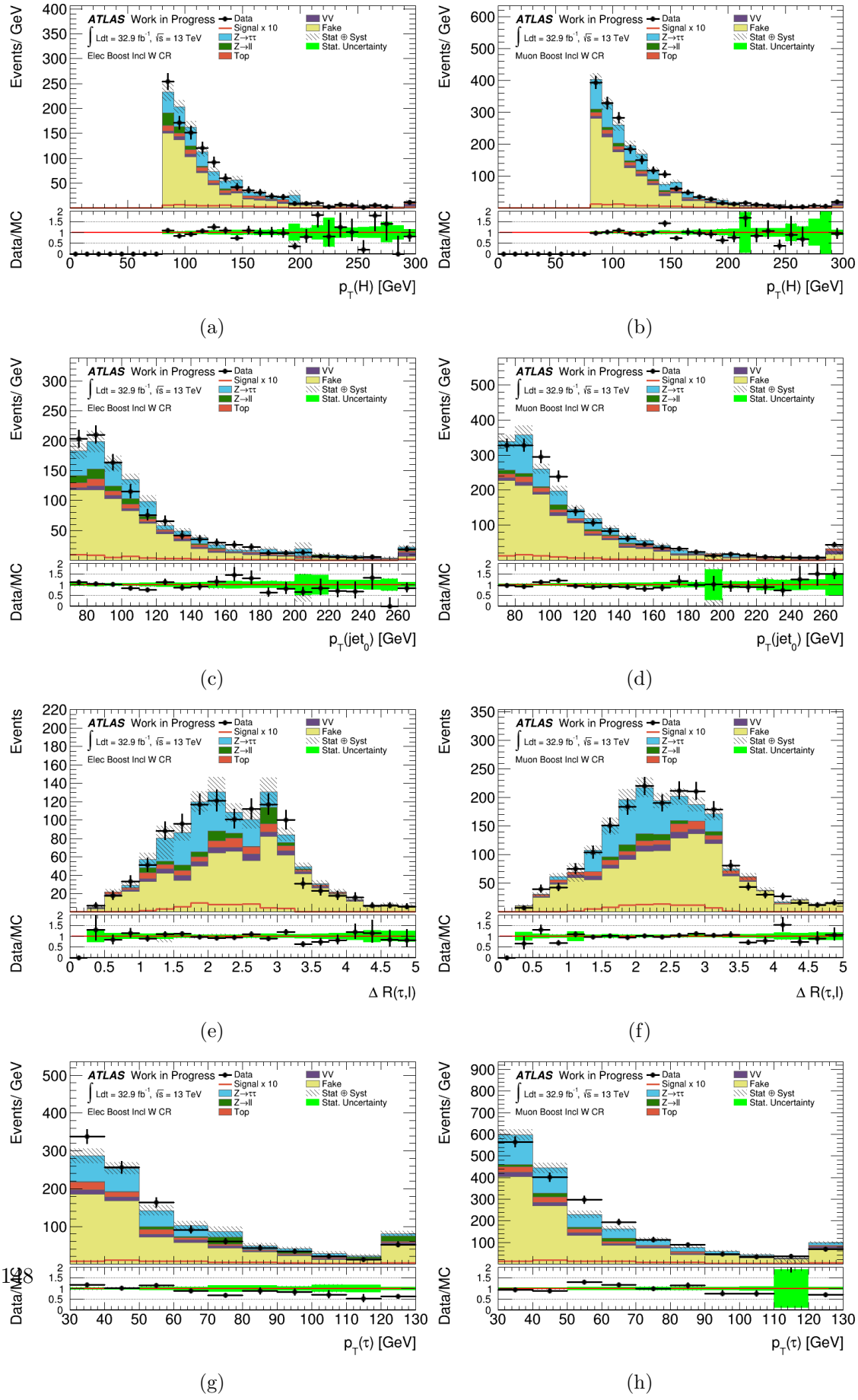


Figure 7.10.: Distributions in WCR in Boosted region for $\tau_e\tau_{had}$ ($\tau_\mu\tau_{had}$) channel : *a) (b)* Higgs p_T, *c) (d)* lead jet p_T, *e) (f)* $\Delta R(lep, \tau)$, *g) (h)* τ p_T.

7.2.3. Expected sensitivity estimation

After having categorised the events in Boosted and VBF region, an expected sensitivity estimation study has been done instead of performing a full fit like as it was described in section 6.7. This approach has been chosen since the analysis was not fully optimised for signal search, but it was mainly a proof of concept that this TLT trigger region can be used as complementary analysis region to the one already explored by SLT trigger. The sensitivity has been estimated counting signal (S) and background (B) events and using two figures of merit, the signal significance defined as $S/\sqrt{S+B}$, and the Asimov significance defined as Asimov Sgn $= \sqrt{(2\times(S+B)*ln(1+S/B)-S}$. The Asimov significance gives similar results to another common figure of merit, $S/\sqrt{B}$, in the limit in which S << B. Results are reported in Table 7.8; from these results it is possible to conclude that adding the TLT region could improve the final results for the analysis. Moreover, the TLT phase space will become larger in the next years when the SLT thresholds will necessarily increase due to the improvement of the LHC performance. As future improvements for this analysis, two main directions can be pursued:

- split the Boosted and VBF regions in *Loose* and *Tight* categories to increase the final sensitivity;
- use multivariate techniques to improve the signal selection.

Region	$S/\sqrt{S+B}$	Asimov Sgn
Elec VBF Incl	0.556	0.565
Muon VBF Incl	0.721	0.733
Comb VBF Incl	0.910	0.925
Elec Boosted Incl	0.646	0.650
Muon Boosted Incl	0.917	0.923
Comb Boosted Incl	1.119	1.126

Table 7.8.: Expected sensitivity for TLT based analysis.

CHAPTER 8

Conclusion

This thesis presents a measurement of the Standard Model Higgs boson production in the $H \to \tau\tau$ decay focusing mostly on the final state in which one τ decays hadronically and the other one decays leptonically ($H \to \tau\tau \to \tau_{lep}\tau_{had}$ channel). The analysis uses a dataset corresponding to 36.1 fb^{-1} collected by the ATLAS detector in the 2015 and 2016 data taking period at a centre of mass energy $\sqrt{s} = 13$ TeV.
The $H \to \tau\tau$ decay is important because at the LHC running condition is the only possible way to measure the Yukawa coupling of the Higgs boson to leptons. Among the different final states, the $\tau_{lep}\tau_{had}$ decay channel has an important role due to the high branching ratio and the moderate background. The analysis has been optimised to be sensitive to the major production processes of a SM Higgs boson, the production via gluon fusion and vector-boson fusion; the selection criteria have been harmonised across the different final states of a di-tau decay in order to select a similar phase space region. This reduces the tensions between the channels when the different signal regions are combined in a profile likelihood fit performed on the mass of the di-tau system, which is reconstructed using the Missing Mass Calculator technique. The signal strength and the values of the observed (expected) significance for each decay channel as well as for the combination are reported in tab. 8.1.

	Combined	$\tau_{lep}\tau_{lep}$	$\tau_{lep}\tau_{had}$	$\tau_{had}\tau_{had}$
Signal strength	$1.08^{+0.32}_{-0.28}$	$1.97^{+0.83}_{-0.81}$	$0.90^{+0.44}_{-0.38}$	$0.86^{+0.57}_{-0.44}$
Exp. (Obs.) Significance	4.359 (4.127)	2.202 (1.196)	2.425 (2.627)	2.206 (2.851)

Table 8.1.: Results of the search for $H \to \tau\tau$ Run 2 analysis.

The results are in good agreement with the Standard Model expectation. The breakdown

of the error on the signal strength shows that the analysis is limited by the systematic uncertainties; in particular, the jet and missing transverse energy related systematics have the largest impact both on the events acceptance and on the di-tau mass shape.
Besides the signal strength, also the values of the cross section times the branching ratio for the gluon-fusion and the vector boson fusion processes have been measured:

$$\sigma_{H\tau\tau}(VBF) = 0.28^{+0.09}_{-0.09}(\text{stat.})^{+0.10}_{-0.09}(\text{syst.}) \quad \text{pb}$$

$$\sigma_{H\tau\tau}(ggH) = 2.97^{+1.03}_{-1.01}(\text{stat.})^{+1.67}_{-1.23}(\text{syst.}) \quad \text{pb}$$

The results are in good agreement with the expected values, confirming the robustness of the analysis. The combination of this analysis with the multivariate analysis performed during the Run 1 data taking period gives an observed (expected) significance of 6.37 (5.43), from which it is possible to claim the observation of $H \to \tau\tau$ decay process with the ATLAS detector. After this measurement, a detailed program of the Higgs boson property measurements in this decay channel can start; investigations can be done to achieve better results for the Higgs boson production processes measurements both in the context of the *Simplified Template Cross Section* (STXS) framework and of the differential cross-section. The STXS framework's primary goal is to maximize the sensitivity of all the different Higgs boson measurements while at the same time minimizing their theory dependence. Additionally, an investigation of the CP nature of the Higgs boson in fermionic decays can be done since this is a not well explored sector so far.
However, to perform these challenging tasks, the analysis would benefit from reduced experimental/theoretical systematic uncertainties and it can also be improved adding new signal regions and using new multivariate τ reconstruction/identification algorithms which are now in development/deployment phase.

Acknowledgement

I would like to thank Prof. Dr. Arnulf Quadt and Prof. Cavasinni for having been my supervisors during my joint PhD between Goettingen and Pisa; thanks for all tips, the advices and the fruitful discussions we had during the PhD period. Moreover, I would like to thank both of you for the effort in setting up to co-tutelle and I hope that such a unique opportunity I had would be offered to other students.
Thanks also to Prof. Lai for his availability and the all the hints.
Another huge thanks goes to Ulla and Zinonas for their supervision and for being a constant example to look at. Your dedication and joy in doing physics all day was and it still is a huge influence in my daily *job*.
A big thanks goes of course to all the SM $H \to \tau\tau$ analysis team for all the support, the endless discussions and the stress we suffered all together towards the paper publication. In particular, I would like to thank Luca, Pier-Olivier, Quentin, Michel, Christian and Daniel for all their patience and the time they spent with me.
Last but not least I would like to thank my family and my friends for their constant support in the all the PhD period; despite being far, your constant presence and encouragement were always there to push me to give me my best.

Appendices

APPENDIX A

OS-SS background estimation correction factors

The Tables A.1 and A.2 show the values of r_{QCD} and $k_{OS}^{W(Top)}$ ($k_{SS}^{W(Top)}$) for the different analysis regions for 1/3 prong taus. All these coefficients are used in the OS-SS background estimation method described in Section 4.3.2 .

p_T	ID	RQCD	kW-OS	kW-SS	kTop OS	kTop SS
[20,25]	! lse	1.091 ± 0.037	0.921 ± 0.010	0.878 ± 0.011	1.294 ± 0.035	1.409 ± 0.046
	lse ! med	1.059 ± 0.080	1.073 ± 0.032	1.038 ± 0.044	1.269 ± 0.078	1.431 ± 0.190
	med ! tig	1.078 ± 0.103	1.056 ± 0.039	1.234 ± 0.061	1.318 ± 0.065	1.544 ± 0.180
	tig	1.113 ± 0.086	1.283 ± 0.047	1.328 ± 0.071	1.287 ± 0.037	2.079 ± 0.152
[25,30]	! lse	1.102 ± 0.038	0.938 ± 0.015	0.964 ± 0.020	1.216 ± 0.048	1.393 ± 0.096
	lse ! med	1.201 ± 0.100	1.077 ± 0.046	1.194 ± 0.073	1.254 ± 0.057	1.467 ± 0.150
	med ! tig	0.966 ± 0.101	1.207 ± 0.061	1.178 ± 0.110	1.302 ± 0.068	1.929 ± 0.248
	tig	0.973 ± 0.118	1.241 ± 0.060	1.371 ± 0.111	1.174 ± 0.042	1.910 ± 0.146
[30,40]	! lse	1.122 ± 0.033	0.983 ± 0.018	1.058 ± 0.026	1.236 ± 0.034	1.476 ± 0.051
	lse ! med	1.109 ± 0.096	1.096 ± 0.048	1.092 ± 0.085	1.247 ± 0.071	1.744 ± 0.115
	med ! tig	1.165 ± 0.123	1.174 ± 0.064	1.385 ± 0.130	1.261 ± 0.051	1.878 ± 1.877
	tig	0.727 ± 0.137	1.190 ± 0.056	1.271 ± 0.133	1.149 ± 0.028	1.896 ± 0.137
[40,-]	! lse	1.133 ± 0.043	1.116 ± 0.027	1.434 ± 0.061	1.262 ± 0.022	1.523 ± 0.035
	lse ! med	1.220 ± 0.063	1.341 ± 0.079	1.812 ± 0.223	1.207 ± 0.041	1.537 ± 0.099
	med ! tig	1.286 ± 0.085	1.109 ± 0.073	1.973 ± 0.308	1.210 ± 0.040	1.403 ± 0.144
	tig	1.181 ± 0.087	1.382 ± 0.083	1.712 ± 0.235	1.114 ± 0.019	1.825 ± 0.129

Table A.1.: Correction factors from OS-SS bkg method for 1 prong channel.

p_T	ID	RQCD	kW-OS	kW-SS	kTop OS	kTop SS
[20,25]	! lse	1.070 ± 0.026	0.944 ± 0.009	0.934 ± 0.010	1.377 ± 0.030	1.396 ± 0.036
	lse ! med	1.177 ± 0.091	1.221 ± 0.061	1.165 ± 0.076	1.567 ± 0.130	1.804 ± 0.168
	med ! tig	1.002 ± 0.109	1.514 ± 0.117	1.590 ± 0.190	1.530 ± 0.114	2.569 ± 0.244
	tig	1.248 ± 0.273	1.414 ± 0.154	1.252 ± 0.1501	1.685 ± 0.116	2.552 ± 0.401
[25,30]	! lse	1.084 ± 0.027	0.986 ± 0.012	0.991 ± .015	1.313 ± 0.029	1.417 ± 0.030
	lse ! med	1.300 ± 0.191	1.296 ± 0.078	1.518 ± 0.130	1.834 ± 0.096	1.706 ± 0.179
	med ! tig	1.332 ± 0.121	1.276 ± 0.105	1.753 ± 0.225	1.624 ± 0.181	1.985 ± 0.223
	tig	1.139 ± 0.177	1.596 ± 0.162	1.445 ± 0.243	1.467 ± 0.097	2.116 ± 0.306
[30,40]	! lse	1.133 ± 0.017	1.057 ± 0.013	1.064 ± 0.017	1.312 ± 0.023	1.336 ± 0.025
	lse ! med	1.170 ± 0.085	1.299 ± 0.068	1.301 ± 0.110	1.531 ± 0.072	1.915 ± 0.139
	med ! tig	1.394 ± 0.136	1.286 ± 0.105	1.459 ± 0.095	2.037 ± 0.196	1.459 ± 0.095
	tig	1.124 ± 0.189	1.236 ± 0.155	1.511 ± 0.244	1.427 ± 0.064	1.691 ± 0.205
[40,-]	! lse	1.200 ± 0.018	1.171 ± 0.018	1.308 ± 0.032	1.299 ± 0.014	1.354 ± 0.019
	lse ! med	1.230 ± 0.072	1.393 ± 0.102	1.493 ± 0.165	1.331 ± 0.050	1.765 ± 0.114
	med ! tig	1.387 ± 0.151	1.645 ± 0.191	1.352 ± 0.260	1.219 ± 0.061	1.576 ± 0.293
	tig	1.876 ± 0.202	1.416 ± 0.210	1.349 ± 0.291	1.188 ± 0.034	1.912 ± 0.207

Table A.2.: Correction factors from OS-SS bkg method for 3 prong channel.

APPENDIX B

Nuisance parameters list

The Table B.1 shows the list of nuisance parameters of the fit model described in Section 6.7

Systematic	Description
EG_RESOLUTION_ALL	
EG_SCALE_ALLCORR	
EG_SCALE_E4SCINTILLATOR	
EG_SCALE_LARCALIB_2015PRE	Electron nuisance parameters:
EG_SCALE_LARTEMP_2015PRE	Momentum Resolution and Scale, Trigger,
EG_SCALE_LARTEMP_2016PRE	Identification, Isolation,
EL_EFF_ID_TOTAL	Reconstruction
EL_EFF_ISO_TOTAL	
EL_EFF_RECO_TOTAL	
EL_EFF_TRIG2015_TOTAL	
EL_EFF_TRIG2016_TOTAL	
FT_EFF_Eigen_b_0	
FT_EFF_Eigen_b_1	
FT_EFF_Eigen_b_2	
FT_EFF_Eigen_c_0	BTagging nuisance parameters
FT_EFF_Eigen_c_1	
FT_EFF_Eigen_c_2	
FT_EFF_Eigen_light_0	
FT_EFF_Eigen_light_1	
FT_EFF_Eigen_light_2	
FT_EFF_Eigen_light_3	
FT_EFF_Eigen_light_4	
FT_EFF_extrapolation	
FT_EFF_extrapolation_from_charm	
JVT	(forward)Jet Vertex Tagger

Forward_JVT	nuisance parameters
JER_CROSSCALIBFWD	
JER_NOISEFWD	
JER_NP_0	
JER_NP_1	
JER_NP_2	Jet Energy Resolution
JER_NP_3	11 nuisance parameters
JER_NP_4	scheme
JER_NP_5	
JER_NP_6	
JER_NP_7	
JER_NP_8	
JES_BJES	
JES_EffectiveNP_1	
JES_EffectiveNP_2	
JES_EffectiveNP_3	
JES_EffectiveNP_4	
JES_EffectiveNP_5	
JES_EffectiveNP_6	
JES_EffectiveNP_7	
JES_EffectiveNP_8	
JES_EtaInter_Model	
JES_EtaInter_NonClosure	Jet Energy Scale
JES_EtaInter_Stat	21 nuisance parameters
JES_Flavor_Comp	scheme
JES_Flavor_Resp	
JES_HighPt	
JES_PU_OffsetMu	
JES_PU_OffsetNPV	
JES_PU_PtTerm	
JES_PU_Rho	
JES_PunchThrough	
JET_TileCorr	
MET_SoftTrk_ResoPara	
MET_SoftTrk_ResoPerp	E_T^{miss} nuisance parameters:
MET_SoftTrk_Scale	Scale and Resolutions
MUONS_ID	
MUONS_MS	
MUONS_SAGITTA_RESBIAS	
MUONS_SAGITTA_RHO	
MUONS_SCALE	
MUON_EFF_LOWPT_STAT	Muon nuisance parameters:
MUON_EFF_LOWPT_SYST	Momentum Resolution and Scale,
MUON_EFF_STAT	Trigger, Identification, Isolation,
MUON_EFF_SYS	
MUON_EFF_TrigStat2015	
MUON_EFF_TrigStat2016	
MUON_EFF_TrigSyst2015	
MUON_EFF_TrigSyst2016	

MUON_ISO_STAT MUON_ISO_SYS	
PRW_DATASF	Pileup reweighthing nuisance parameter
LumiUncCombined	Luminosity nuisance parameter
Theo_VBFH_MUR_MUF Theo_VH_MUR_MUF	VBF and VH renormalisation/factorisation scale nuisance parameters
Theo_ggH_sig_qcd_0 Theo_ggH_sig_qcd_1 Theo_ggH_sig_qcd_2 Theo_ggH_sig_qcd_3 Theo_ggH_sig_qcd_4 Theo_ggH_sig_qcd_5 Theo_ggH_sig_qcd_6 Theo_ggH_sig_qcd_7 Theo_ggH_sig_qcd_8	 Gluon-fusion QCD scale uncertainties 9 nuisance parameters scheme (Section 6.6.2)
Theo_sig_alphaS	QCD α_s nuisance parameter
Theo_sig_pdf_0 Theo_sig_pdf_1 Theo_sig_pdf_2 Theo_sig_pdf_3 Theo_sig_pdf_4 Theo_sig_pdf_5 Theo_sig_pdf_6 Theo_sig_pdf_7 Theo_sig_pdf_8 Theo_sig_pdf_9 Theo_sig_pdf_10 Theo_sig_pdf_11 Theo_sig_pdf_12 Theo_sig_pdf_13 Theo_sig_pdf_14 Theo_sig_pdf_15 Theo_sig_pdf_16 Theo_sig_pdf_17 Theo_sig_pdf_18 Theo_sig_pdf_19 Theo_sig_pdf_20 Theo_sig_pdf_21 Theo_sig_pdf_22 Theo_sig_pdf_23 Theo_sig_pdf_24 Theo_sig_pdf_25 Theo_sig_pdf_26 Theo_sig_pdf_27 Theo_sig_pdf_28 Theo_sig_pdf_29	 PDF4LHC PDF set 30 nuisance parameters scheme
ZttTheory_CKK_Relative_Bst ZttTheory_MUR_MUF_Relative_Bst ZttTheory_PDF_Relative_Bst	

ZttTheory_QSF_Relative_Bst	Theoretical uncertainties for $Z \to \tau\tau$:
ZttTheory_MGvsSH_Relative_Bst	PDF, renormalisation and factorisation scales,
ZttTheory_CKK_boost_chan_ll (lh and hh)	jet-to-parton matching, resummation scale
ZttTheory_MUR_MUF_chan_ll	(Section 6.6.2)
ZttTheory_PDF_boost_chan_ll (lh and hh)	
ZttTheory_QSF_boost_chan_ll (lh and hh)	
ZttTheory_MGvsSH_boost_chan_ll (lh and hh)	
hh_dPhiWeightStat	
hh_dPhiWeightSys	
hh_dPhiWeightClosSys	
hh_fake_contamination	
lh_fake_stat_vbf	
lh_fake_stat_boost	Nuisance parameters from
lh_fake_stat_vbf_top	Data-Driven background
lh_fake_stat_boost_top	estimate in all the channels
lh_fake_nonclosure	
lh_fake_rvar_boost	
lh_fake_rvar_boost	
lh_fake_rvar_boost_top	
lh_fake_rvar_vbf_top	
ll_fake_nonclosure	
ll_fake_stat	
norm_LH_boost_Top	
norm_LH_vbf_Top	Top normalisation factor
norm_LL_boost_Top	for $\tau_{lep}\tau_{had}$ and $\tau_{lep}\tau_{lep}$
norm_LL_vbf_Top	channels
norm_LL_vbf_Zll	$Z \to ll$ normalisation factor
norm_LL_boost_Zll	for $\tau_{lep}\tau_{lep}$ channel
norm_boost_Ztt	$Z \to \tau\tau$ normalisation factor
norm_vbf_Ztt	for all the channels

Table B.1.: Summary of the nuisance parameters names used in the fit.

Bibliography

[1] ATLAS Collaboration, G. Aad et al., *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, arXiv:1207.7214 [hep-ex].

[2] CMS Collaboration, 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, arXiv:1207.7235 [hep-ex].

[3] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[4] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587.

[5] T. W. B. Kibble, *Symmetry breaking in nonAbelian gauge theories*, Phys. Rev. **155** (1967) 1554–1561.

[6] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[7] ATLAS Collaboration, G. Aad et al., *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) 117, arXiv:1501.04943 [hep-ex].

[8] ATLAS, CMS Collaboration, G. Aad et al., *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, arXiv:1606.02266 [hep-ex].

[9] Particle Data Group Collaboration, C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. **C40** (2016) 100001.

[10] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, http://www.slac.stanford.edu/~mpeskin/QFT.html.

[11] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588.

[12] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[13] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367–377.

[14] C.-N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (1954) 191–195, [,150(1954)].

[15] I. J. R. Aitchison and A. J. G. Hey, *Gauge theories in particle physics: A practical introduction. Vol. 1: From relativistic quantum mechanics to QED*, http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.3.F5A34::2012.

[16] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132–133.

[17] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509, [,160(1964)].

[18] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156–1163.

[19] LHC Higgs Cross Section Working Group, *https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG*,.

[20] H. M. Georgi et al., *Higgs Bosons from Two Gluon Annihilation in Proton Proton Collisions*, Phys. Rev. Lett. **40** (1978) 692.

[21] A. Djouadi and P. Gambino, *Leading electroweak correction to Higgs boson production at proton colliders*, Phys. Rev. Lett. **73** (1994) 2528–2531, arXiv:hep-ph/9406432 [hep-ph].

[22] M. Duhrssen et al., *Extracting Higgs boson couplings from CERN LHC data*, Phys. Rev. **D70** (2004) 113009, arXiv:hep-ph/0406323 [hep-ph].

[23] J. Ohnemus and W. J. Stirling, *Order-s corrections to the differential cross section for the WH intermediate-mass Higgs-boson signal*, Physical review D: Particles and fields **47** (1993) 2722–2729.

[24] A. Stange, W. J. Marciano, and S. Willenbrock, *Higgs bosons at the Fermilab Tevatron*, Phys. Rev. **D49** (1994) 1354–1362, arXiv:hep-ph/9309294 [hep-ph].

[25] A. Stange, W. J. Marciano, and S. Willenbrock, *Associated production of Higgs and weak bosons, with H —¿ b anti-b, at hadron colliders*, Phys. Rev. **D50** (1994) 4491–4498, arXiv:hep-ph/9404247 [hep-ph].

[26] R. Hamberg, W. L. van Neerven, and T. Matsuura, *A complete calculation of the order $\alpha - s^2$ correction to the Drell-Yan K factor*, Nucl. Phys. **B359** (1991) 343–405, [Erratum: Nucl. Phys.B644,403(2002)].

[27] O. Brein et al., *Top-Quark Mediated Effects in Hadronic Higgs-Strahlung*, Eur. Phys. J. **C72** (2012) 1868, arXiv:1111.0761 [hep-ph].

[28] O. Brein, R. V. Harlander, and T. J. E. Zirke, *vh@nnlo - Higgs Strahlung at hadron colliders*, Comput. Phys. Commun. **184** (2013) 998–1003, arXiv:1210.5347 [hep-ph].

[29] LHC Higgs Cross Section Working Group Collaboration, J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, arXiv:1307.1347 [hep-ph].

[30] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001–S08001, https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001.

[31] ATLAS Collaboration, *https://twiki.cern.ch/twiki/bin/view/AtlasPublic/*,.

[32] ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[33] ATLAS Collaboration,, *ATLAS inner detector: Technical design report. Vol. 1*, CERN-LHCC-97-16 (1997).

[34] ATLAS Collaboration,, *ATLAS central solenoid: Technical design report*, CERN-LHCC-97-21 (1997).

[35] ATLAS Collaboration, M. S. Alam et al., *ATLAS pixel detector: Technical design report*, CERN-LHCC-98-13 (1998).

[36] A. Ahmad et al., *The Silicon microstrip sensors of the ATLAS semiconductor tracker*, Nucl. Instrum. Meth. **A578** (2007) 98–118.

[37] ATLAS TRT Collaboration, E. Abat et al., *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: Design and performance*, JINST **3** (2008) P02013.

[38] ATLAS Collaboration, A. Airapetian et al., *ATLAS calorimeter performance Technical Design Report*, CERN-LHCC-96-40 (1996).

[39] ATLAS Collaboration,, *ATLAS muon spectrometer: Technical design report*, CERN-LHCC-97-22 (1997).

[40] ATLAS Collaboration,, *ATLAS first level trigger: Technical design report*, CERN-LHCC-98-14 (1998).

[41] C. Cuenca Almenar et al., *ATLAS online data quality monitoring*, Nucl. Phys. Proc. Suppl. **215** (2011) 304–306.

[42] ATLAS Collaboration, T. A. collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ TeV, ATLAS-CONF-2017-029 (2017).

[43] T. A. collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, https://cds.cern.ch/record/2064383.

[44] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[45] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B641** (2006) 57–61, arXiv:hep-ph/0512210 [hep-ph].

[46] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, ATL-LARG-PUB-2008-002 (2008).

[47] ATLAS Collaboration, T. Barillari et al., *Local hadronic calibration*, ATL-LARG-PUB-2009-001-2 (2009).

[48] A. Elagin et al., *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. **A654** (2011) 481–489, arXiv:1012.4686 [hep-ex].

[49] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].

[50] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv:0709.2092 [hep-ph].

[51] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv:1002.2581 [hep-ph].

[52] E. Bagnaschi et al., *Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM*, JHEP **02** (2012) 088, arXiv:1111.2854 [hep-ph].

[53] K. Hamilton, P. Nason, and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05** (2015) 140, arXiv:1501.04637 [hep-ph].

[54] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[55] 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, arXiv:0707.0381 [hep-ph].

[56] 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, arXiv:0710.4749 [hep-ph].

[57] P. Bolzoni et al., *Higgs production via vector-boson fusion at NNLO in QCD*, Phys. Rev. Lett. **105** (2010) 011801, arXiv:1003.4451 [hep-ph].

[58] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys. Lett. **B579** (2004) 149–156, arXiv:hep-ph/0307206 [hep-ph].

[59] L. Altenkamp et al., *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, JHEP **02** (2013) 078, arXiv:1211.5015 [hep-ph].

[60] 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, arXiv:1112.5142 [hep-ph].

[61] K. Hamilton et al., *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222, arXiv:1309.0017 [hep-ph].

[62] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016) 023001, arXiv:1510.03865 [hep-ph].

[63] ATLAS Collaboration, G. Aad et al., *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, arXiv:1406.3660 [hep-ex].

[64] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv:hep-ph/0201195 [hep-ph].

[65] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and Physics Documentation*, Comput. Phys. Commun. **199** (2016) 86–101, arXiv:1011.0937 [hep-ph].

[66] C. Anastasiou et al., *Higgs Boson Gluon-Fusion Production in QCD at Three Loops*, Phys. Rev. Lett. **114** (2015) 212001, arXiv:1503.06056 [hep-ph].

[67] C. Anastasiou et al., *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, JHEP **05** (2016) 058, arXiv:1602.00695 [hep-ph].

[68] S. Actis et al., *NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders*, Phys. Lett. **B670** (2008) 12–17, arXiv:0809.1301 [hep-ph].

[69] C. Anastasiou, R. Boughezal, and F. Petriello, *Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion*, JHEP **04** (2009) 003, arXiv:0811.3458 [hep-ph].

[70] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv:0811.4622 [hep-ph].

[71] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv:0808.3674 [hep-ph].

[72] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv:1111.5206 [hep-ph].

[73] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv:0709.1027 [hep-ph].

[74] S. Hoeche et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv:1207.5030 [hep-ph].

[75] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph].

[76] S. Alioli et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111, arXiv:0907.4076 [hep-ph], [Erratum: JHEP02,011(2010)].

[77] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. **C71** (2011) 1547, arXiv:1009.2450 [hep-ph].

[78] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[79] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018, arXiv:1005.3457 [hep-ph].

[80] ATLAS Collaboration, G. Aad et al., *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823–874, arXiv:1005.4568 [physics.ins-det].

[81] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[82] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[83] ATLAS Collaboration,, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003 (2012).

[84] ATLAS Collaboration, G. Aad et al., *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. **C74** (2014) 2941, arXiv:1404.2240 [hep-ex].

[85] ATLAS Collaboration, G. Aad et al., *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. **C74** (2014) 3071, arXiv:1407.5063 [hep-ex].

[86] ATLAS Collaboration, T. A. collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024 (2016).

[87] ATLAS Collaboration, G. Aad et al., *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, Eur. Phys. J. **C74** (2014) 3130, arXiv:1407.3935 [hep-ex].

[88] ATLAS Collaboration, T. A. collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083 (2013).

[89] T. A. collaboration, *Forward Jet Vertex Tagging: A new technique for the identification and rejection of forward pileup jets*, ATL-PHYS-PUB-2015-034 (2015), https://cds.cern.ch/record/2042098.

[90] A.F.T. Group, *ATLAS Flavour Tagging group recommendation for 2015 data*, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTagCalib2015 .

[91] ATLAS Collaboration, T. A. collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV*, ATLAS-CONF-2013-082 (2013).

[92] ATLAS Collaboration, G. Aad et al., *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. **C76** (2016) 292, arXiv:1603.05598 [hep-ex].

[93] *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, https://cds.cern.ch/record/2037613.

[94] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [hep-ph].

[95] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, arXiv:1007.1727 [physics.data-an], [Erratum: Eur. Phys. J.C73,2501(2013)].

[96] ATLAS Collaboration, G. Aad et al., *Modelling $Z \to \tau\tau$ processes in ATLAS with τ-embedded $Z \to \mu\mu$ data*, JINST **10** (2015) P09018, arXiv:1506.05623 [hep-ex].