UNIVERSITÀ DEGLI STUDI DI MILANO

Department of Physics

PHD SCHOOL IN PHYSICS, ASTROPHYSICS AND APPLIED PHYSICS

Cycle XXXII

Physics of the Higgs boson in the ditau decay channel with the ATLAS detector

Disciplinary Scientific Sector FIS/01 and FIS/04

Director of the School: Prof. Matteo Paris

Supervisor: Prof. Attilio Andreazza

PhD Thesis of:

Alessia Murrone

A. Y. 2019/2020

Commission of the final examination:

External Referee:
Dott. Paolo Francavilla - INFN Pisa
Prof. Elisabetta Gallo - Universitaet Hamburg, DESY

External Member:
Prof. Stan Lai - Universitaet Goettingen
Prof. Marta Calvi - Università di Milano Bicocca

Internal Member:
Prof. Nicola Neri - Università di Milano

Final examination:

January 29th, 2020

Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy

Cover illustration:

Display of an event of Higgs boson decaying in a pair of tau leptons

MIUR subjects:

FIS/01 - FIS/04

PACS:

14.80.Bn, 14.80.Ec

Abstract

This thesis presents two measurements performed in the $H \to \tau\tau$ channel and based on the proton-proton collisions data collected by the ATLAS experiment during 2015 and 2016 and corresponding to an integrated luminosity of 36.1 fb^{-1}. The $H \to \tau\tau$ is a considerably important decay channel because it allows to directly measure the Higgs Yukawa coupling and to probe the Higgs boson properties.
Measurements of the Higgs production cross section in the $H \to \tau\tau$ decay channel are presented. The observed (expected) significance of the $H \to \tau\tau$ signal excess over the expected background amounts to 4.4 (4.1) standard deviations. This result, combined with the data taken at $\sqrt{s} = 7$ and 8 TeV, leads to an observed (expected) significance of 6.4 (5.4) and constitutes the first ATLAS observation of $H \to \tau\tau$. The measured total cross section of $H \to \tau\tau$, using the data collected at $\sqrt{s} = 13$ TeV, is $3.77^{+0.60}_{-0.59}$ (stat.) $^{+0.87}_{-0.74}$ (syst.) pb which is consistent with the Standard Model expectation. In addition, total cross sections for the vector boson fusion and gluon-gluon fusion processes have been measured separately and similar results are reported based on the simplified template cross section framework. All of these measurements are in agreement with the Standard Model predictions.
The second measurement presented in this manuscript concerns a CP violation study in the same decay channel. The coupling structure of the vertex where the Higgs couples to two electroweak gauge bosons in its vector boson fusion production is investigated. The Optimal Observable method is used to directly test CP invariance in this vertex. No sign of CP violation is observed and constraints on the $\tilde{d}$ parameter, which governs the strength of CP violation, have been set. A 68% CL interval is observed for $\tilde{d}$ of [-0.090, 0.035], compared to an expected interval of $\tilde{d} \in [-0.035, 0.033]$. No constraints can be set at 95% CL despite an expected interval $\tilde{d} \in [-0.21, 0.15]$ at 95% CL was determined.

Contents

v

Introduction viii

1 The Standard Model of Particle Physics 1
- 1.1 Electroweak interaction 1
- 1.2 Quantum Chromodynamics 3
- 1.3 Higgs mechanism 5
- 1.4 Higgs boson phenomenology 9
- 1.5 The limitations of the Standard Model 17
- 1.6 CP violation 20

2 The LHC and the ATLAS detector 23
- 2.1 Hadron colliders 23
- 2.2 The Large Hadron Collider 24
- 2.3 The ATLAS experiment 26

3 Data and Monte Carlo samples 39
- 3.1 Data preparation 39
- 3.2 Data format and software framework 41
- 3.3 Monte Carlo generators and simulation 42

4 Physics objects reconstruction in ATLAS 45
- 4.1 Tracks reconstruction in the Inner Detector 45
- 4.2 Electrons and photons 45
- 4.3 Muons 49
- 4.4 Jets 52
- 4.5 Tau 54
- 4.6 Missing transverse momentum 59

5 Cross section measurement of the Higgs decaying into a pair of tau leptons **61**
5.1 Introduction 61
5.2 Data and Monte Carlo samples 62
5.3 Object Reconstruction 64
5.4 Missing Mass Calculator 65
5.5 Event selection 67
5.6 Background estimation 72
5.7 Systematic uncertainties 79
5.8 Statistical analysis 103
5.9 Results 113
5.10 Conclusion 137

6 CP violation measurement in VBF production **139**
6.1 Introduction 139
6.2 Theoretical framework 140
6.3 The Optimal Observable 144
6.4 Data and Monte Carlo samples 146
6.5 Event selection 148
6.6 Background estimation 155
6.7 Systematic uncertainties 160
6.8 Statistical analysis 162
6.9 Results 166
6.10 Comparison with other analysis 190
6.11 Conclusion 191

Appendices **192**

A Pixel detector studies **195**
A.1 Introduction 195
A.2 The ATLAS Pixel detector 195
A.3 Errors studies 199

Bibliography **213**

Acknowledgments **225**

Introduction

On July 4th, 2012, the ATLAS and CMS collaborations announced the discovery of a particle compatible with the Higgs boson predicted by the Standard Model, with a mass around 125 GeV [1, 2]. The Higgs boson is the particle foreseen by the Higgs mechanism, proposed in 1964 by Higgs [3] and, Brout and Englert [4], in an independent way. This mechanism is responsible for the electroweak spontaneous symmetry breaking and then of the consequent acquisition of mass of all the Standard Model particles. The search of the Higgs boson was one of the major quests at accelerators during the previous 20 years since it was the only Standard Model particle still not observed. The discovery made by the ATLAS and CMS experiments, has therefore allowed to explore a new unkown sector. After the Higgs discovery, the studies have been focused on measuring its properties like its coupling strength and spin-parity properties. Many measurements have been performed during the Run 1 data taking at a center of mass energies of 7 and 8 TeV at the LHC and they all have confirmed the Standard Model expectations [5, 6]. From 2015 to 2018, the LHC has been operating at a center of mass energy of 13 TeV and at higher instantaneous luminosity, allowing new high precision measurements in the recent discovered Higgs sector.

Many Run 1 results concerning the Higgs boson rely on the bosonic decay modes of this particle. It is therefore highly important to asses the Higgs couplings to the fermionic sector as well. Among the Higgs fermion couplings, only the ones to the third generation fermions are currently accessible at the LHC and the Higgs decay into a pair of τ leptons is the only accessible leptonic decay mode. The $H \to \tau\tau$ is therefore considerably important since it allows to directly probe the Higgs Yukawa couplings.

This thesis work presents two measurements based on the data from proton-proton collisions collected by the ATLAS experiment during 2015 and 2016, and corresponing to an integrated luminosity of 36 fb^{-1}. Both the measurements exploit the $H \to \tau\tau$ decay channel. Firstly a measurement of the Higgs boson production cross section is presented. This analysis exploits all the available combination of leptonic ($\tau \to l\nu\bar{\nu}$ with $l = e\mu$) and hadronic ($\tau \to$ hadrons ν) τ decays. Two different kind of signal regions are defined by applying specific selection cuts: one region which targets events produced through vector boson fusion and one region which is enriched in gluon-gluon fusion events with an additional recoiling jet. A fit to the Higgs invariant mass distribution in all the signal

regions is used in order to extract the signal strength. In addition, measurements of the total Higgs production cross section as well as separate measurements of the vector boson fusion and the gluon-gluon fusion production cross section and a first attempt to cross section measurements, based on the simplified template cross section framework, have been carried out.
The second measurement described in this manuscript concerns a CP violation study in the same Higgs decay channel and therefore is closely related to the cross section measurement. Beside the coupling strength, it is crucial to measure the spin and CP quantum numbers of the Higgs boson in order to determine whether it is exactly the predicted particle by the Standard Model. In addition, it is interesting to look for CP violation since this is one of the three Sakharov conditions [7], required for explaining the baryon asymmetry in the universe.
This second analysis aims at studying the coupling structure of the vertex where the Higgs couples to two electroweak gauge bosons (HVV), through the vector boson fusion production mode. This is done in an Effective Field Theory framework that features only SM particles and symmetries at the electroweak scale. This is a minimal choice but it is well justified by the present data and in addition it reduces the possible interactions that Lorentz symmetry alone would allow and gives a powerful framework where constraints can be analyzed. The analysis strategy is in principle independent from the specific decay mode, however the $H \to \tau\tau$ is used due to its effectiveness in selecting VBF events. The variable used to probe CP violation is the so-called Optimal Observable [8, 9] which combines the informations from a multidimensional phase space into a single quantity. The Optimal Observable mean value is used for a direct test of CP invariance, while a fit to the full Optimal Observable distribution, allows to measure the $\tilde{d}$ parameter, which describes the strength of CP violation. I took part on the first publication using this technique on Run 1 data during my master thesis.
My contributions are focused especially on the fully hadronic τ decay channel ($\tau_{\text{had}}\tau_{\text{had}}$). Indeed, in this thesis the general analysis strategies will be presented but with particular emphasys on this subchannel.
Concerning the cross section measurement I performed the calculation of the theoretical uncertainties on the signal. These involved a first part of truth-level studies where I took care of the Monte Carlo samples production, of the evaluation strategy and of the software development for evaluating these uncertainties. A second step involved the evaluation of reconstruction-level uncertainties with new Monte Carlo techniques developed by a joint effort of many theoretical groups. This is explained in details in Chapter 5. In addition, I took care of the statistical analysis and the fitting strategy for the $\tau_{\text{had}}\tau_{\text{had}}$ channel.
Concerning the CP violation measurement, I was also in charge of the statistical analysis and the fitting strategy for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. In addition, I took care of signal regions optimization and systematic uncertainties, considering that the work done for the cross section analysis affects this analysis as well. Therefore, in this manuscript, more details will be given on the systematic uncertainties and the fitting strategy as well as on the discussion of the results.
I also co-supervised two bachelor thesis related to the $H \to \tau\tau$ studies: one thesis has been developed within the CP measurement, testing the sensitivity of the analysis using

different models with respect to the default one used in the analysis. The other thesis concerned the problem of the recostruction of the Higgs mass in the $H \to \tau\tau$ channel and explored alternative machine learning algorithms with respect to the actual likelihood-based algorithm used in ATLAS.
Moreover, I was involved in studies on the Pixel detector. I worked on studies on the errors flagged by the front-end chips in order to look at possible correlations with occupancy and beam conditions. This task also allowed me to become a qualified ATLAS author. In addition, I joined the Pixel Data Quality group which performs checks on the data, specifically looking at the Pixel detector status. In the meantime I took care of the maintanance and modifications of two packages of Pixel software, which are able to convert and monitor data and informations coming from the Pixel modules and which are part of the ATLAS official software. These studies are not part of the main body of this thesis and are described in Appendix A.
The thesis structure is organized as follows. Chapter 1 gives an introduction to the Standard Model theory and the Higgs mechanism, as well as the Higgs boson phenomenology at the LHC. Chapter 2 describes the Large Hadron Collider and the ATLAS experimental apparatus. The workflow of data and Monte Carlo samples is illustrated in Chapter 3. The algorithms for reconstructing the physics objects, which are of interest of the described analysis, are presented in Chapter 4. Finally, Chapter 5 and Chapter 6 describe the main topic of this work, illustrating the $H \to \tau\tau$ cross section measurement and the CP violation measurement, performed in the same channel, respectively.

Chapter 1

The Standard Model of Particle Physics

The Standard Model (SM) [10, 11, 12, 13] of particle physics is the theoretical model which describes the structure of matter and all fundamental interactions except gravity, through the framework of *quantum field theory*, which unifies relativistic dynamics and quantum mechanics.
The Standard Model is a non-abelian gauge theory invariant under the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$ where $SU(3)_C$ is the color symmetry which provides a complete description of *Quantum Chromodynamics* (QCD) and $SU(2)_L \times U(1)_Y$ is the symmetry of *Electroweak theory* which describes electromagnetic and weak interactions, unified at ~ 100 GeV scale. The gauge group of SM has then $8 + 3 + 1 = 12$ generators with non trivial commutation rules. These 12 generators are associated with 12 spin 1 particles, which are the mediators of the fundamental interactions, called *gauge bosons*: 8 gluons, $W^i (i = 1, 2, 3)$, B. The gluons are massless particles responsible for the strong interaction while $W^i (i = 1, 2, 3)$ and B are responsible for the electroweak interaction. The mass eigenstates for the electroweak gauge bosons correspond to the $W^{\pm}$, Z and γ particles. In the SM there are also *matter fields* in addition to the gauge fields, which describe spin $\frac{1}{2}$ particles called *fermions*, namely *quarks* and *leptons*, grouped in three *generations*.
Figure 1.1 summarizes the fundamental particles and their characteristics. In addition to the particles described above, the Higgs boson, which is the particle associated to the *Higgs field*, is added to the scheme and it is necessary for introducing the mass in the SM as will be described in Section 1.3.

The SM Lagrangian can then be expressed as the sum of an electroweak and a QCD Lagrangian, not taking into account for the moment the Higgs sector:

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} \tag{1.1}$$

1.1 Electroweak interaction

Electromagnetic and weak interactions, altough seem very different at low energy scale, are different aspect of the same force. The model that unifies these two interactions is the result of the work done in the 1960s by Glashow and later by Weinberg and Salam who laid the foundations of the SM [10, 11, 12].

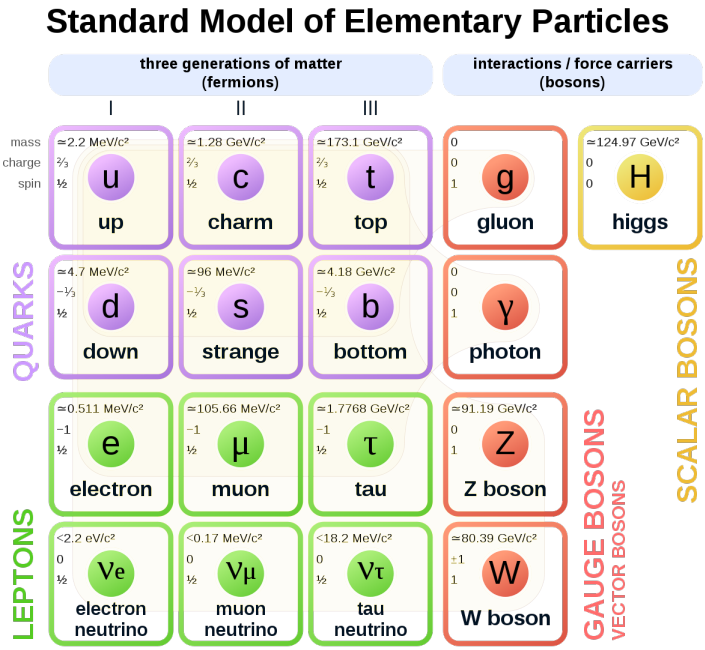


Figure 1.1: Illustration of all Standard Model particles with their features and quantum numbers [14].

The electroweak Lagrangian is invariant under the symmetry group $SU(2)_L \times U(1)_Y$ where the subscript L refers to $SU(2)$ *left-handed* fields and Y refers to the conserved quantity of*electroweak hypercharge* which is defined by the Gell-Mann–Nishijima formula:

$$Q = I_z + \frac{1}{2}Y \tag{1.2}$$

where Q is the electric charge and I_z is the third component of isospin.
The local gauge invariance is obtained by defining the covariant derivative D_μ:

$$\partial_\mu \rightarrow \mathcal{D}_\mu = \partial_\mu + igT_iW_\mu^i + ig'\frac{1}{2}YB_\mu \tag{1.3}$$

where the T^i are the three components of the weak isospin and Y is the weak hypercharge operator.
Thus the complete Lagrangian for the electroweak sector is:

$$\mathcal{L}_{EW} = i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi - \frac{1}{4}W^k_{\mu\nu}W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{1.4}$$

where $W^k_{\mu\nu}$ and $B^k_{\mu\nu}$ are the kinetic terms:

$$\begin{aligned} W^k_{\mu\nu} &= \partial_\mu W^k_\nu - \partial_\nu W^k_\mu - g\epsilon^{klm}W_{\mu,l}W_{\nu,m} \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{1.5}$$

The introduction of the covariant derivative leads to the following expression for the interaction part of the Lagrangian:

$$\mathcal{L}_{EW}^{int} = -g(\overline{\psi}_L \gamma^\mu T_i \psi_L) W_\mu^i - \frac{g'}{2} Y(\overline{\psi}\gamma^\mu \psi) B_\mu \tag{1.6}$$

The fields associated to the electroweak gauge group are therefore four: W_i^μ (i = 1,2,3) for $SU(2)$ and B^μ for $U(1)$. The three W^μ fields couple only to left-handed spinor fields while B^μ couples both to left-handed and right-handed spinors. These gauge fields however don't represent the known physical particles such as $W^\pm$, Z^0, and γ; the physical fields associated to vector bosons $W^\pm$ are obtained by a linear combination of W_1^μ e W_2^μ:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1^\mu \mp W_2^\mu) \tag{1.7}$$

The mixing between W_3^μ e B^μ provides the two neutral bosons, the photon and the Z:

$$\begin{pmatrix} Z^\mu \\ A^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_3^\mu \\ B^\mu \end{pmatrix} \tag{1.8}$$

where the mixing angle θ_W is the Weinberg angle, with $\sin^2\theta_W = 0.23122 \pm 0.000017$, according to the most recent measurement [14]. The electric charge is related to the g and g' coupling constants:

$$g \sin\theta_W = g' \sin\theta_W = e \tag{1.9}$$

and the coupling constant for the weak neutral current is:

$$g_Z = \frac{e}{\sin\theta_W \cos\theta_W}(T_3 - Q\sin^2\theta_W) \tag{1.10}$$

1.2 Quantum Chromodynamics

QCD Lagrangian is based on a $SU(3)$ symmetry group and has the following expression:

$$\mathcal{L}_{QCD} = -\frac{1}{4} G_\alpha^{\mu\nu} G_{\mu\nu}^\alpha + \sum_f \overline{q}_{f\alpha} i\gamma^\mu \mathcal{D}_{\mu\beta}^\alpha q_f^\beta \tag{1.11}$$

where the subscript f denotes the quark flavour, the indices α and β run over the color quantum numbers and $G_{\mu\nu}^i$ is the gluon kinetic term:

$$G_\alpha^{\mu\nu} = \partial^\mu G_\alpha^\nu - \partial^\nu G_\alpha^\mu - g_s f^{abc1} G_b^\mu G_c^\nu \tag{1.12}$$

In Equation 1.11 the covariant derivative is introduced, in order to retain the gauge invariance and it has the following expression:

$$\partial_\mu \to \mathcal{D}_{\mu\beta}^\alpha = \partial_\mu \delta_\beta^\alpha + i g_s \lambda_\beta^{i\alpha} G_\mu^i \tag{1.13}$$

[1] Structure constants of SU(3)

where λ^i are the $SU(3)$ Gell-Mann matrices and the index α runs over the color quantum number.
The QCD coupling constant α_s can be expressed as a function of the energy transferred in the interaction, as follows:

$$\alpha_s(Q^2) = \frac{g_s^2}{4\pi} = \frac{12\pi}{(11N_C - 2N_f)\ln(\frac{Q^2}{\Lambda^2_{QCD}})} \tag{1.14}$$

where Λ_{QCD} is an experimental parameter with value of ~ 200 MeV, N_C is the number of colours and N_f the number of flavours. The strength of the coupling is therefore a quantity that depends on the energy scale of the process and this is a general feature of the quantum field theories. It happens either in quantum electrodynamics, in the weak theory and in QCD: in the first case the coupling constant increases when the scale of the process gets larger. This behaviour can be physically explained as a screening effect of the electron charge by a cloud of e^+e^- generated from emission of virtual photons from the electron itself. This is the result of renormalization of the theory which treats the infinities appearing in the calculation of Feynman diagrams. In the weak theory and in QCD the situation is different because boson-boson self couplings are possible (this is a particularity of non-abelian gauge theories). For low energy scale (therefore for short distances) the coupling becomes larger and this is responsible for the phenomenon of *confinement* of quarks and gluons in the hadron. Only colorless particle such as hadrons are visible in nature, the quarks and gluons in the hadrons, which are called *partons*, are not visible. When the scale gets larger the coupling costant decreases and the partons are quasi free instead, this is what is called *asymptotic freedom*. In addition when attempting to extract a parton from a hadron, because of the increase of the potential energy, new quark-antiquark pairs and gluons are created and this process can proceed with the creation of more pairs (*fragmentation* and *hadronization*).
Infinities arising from loop Feynman diagrams (ultraviolet divergences) in QCD are reabsorbed by the renormalization procedure. This procedure introduces a energy scale from which the physical observables are independent and it is referred to as the renormalization scale, μ_R. In principle the value of μ_R chosen is arbitrary but usually the energy scale of the involved process is used and therefore we can replace Q^2 in Equation 1.14 with μ_R.
The perturbative QCD is fundamental for precision calculation in case of high energy hadronic collisions (as for the LHC). In this case the hadrons interacts revealing their inner structure. Therefore the center of mass energy depends on the fraction of the momentum carried by the interacting partons and thus it is lower than the proton-proton center of mass energy. The parton center of mass energy is:

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 s} \tag{1.15}$$

where $\sqrt{s}$ is the proton center of mass energy and x_1 and x_2 are the so-called Bjorken x, namely the fractions of proton momenta carried by the scattering partons. From the factorization theorem of the QCD [15] it is known that the total hadron-hadron cross section is given by the convolution of the partonic cross section and the Parton

Distribution Functions (PDFs):

$$\sigma_{pp\to X} = \sum_{a,b} \int dx_1 dx_2 f_a(x_1, \mu_F) f_b(x_2, \mu_F) \hat{\sigma}_{ab\to X}(\hat{s}, \mu_F, \mu_R) \tag{1.16}$$

where a and b run over all types of partons, $\hat{\sigma}_{ab\to X}$ is the partonic cross section which depends on $\sqrt{\hat{s}}$ and on μ_R and μ_F which are the renormalization and factorization scales respectively and, $f_a(x_1, \mu_F)$ and $f_b(x_2, \mu_F)$, are the PDFs, which represent the probability to find a parton with a momentum fraction x inside the hadron and depend only on the factorisation scale. The meaning of the renormalization scale has been previously mentioned while the factorization scale can be interpreted as the resolution with which the hadron is being probed. It is related to the collinear and soft infinities in QCD (infrared divergences) which are reabsorbed by redefining the dependence of PDFs from the scale of the process. The PDFs evolution is indeed determined by the DGLAP equations [16, 17, 18] while the perturbative QCD is not able to predict their initial values and they must be fixed from fits to experimental data. The expression for the DGLAP equations is the following:

$$\begin{aligned} \frac{df_{q_i}}{d\log(Q^2)} &= \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} \left[P_{qq}\left(\frac{x}{\xi}\right) f_{q_i}(\xi, Q^2) + P_{qg}\left(\frac{x}{\xi}\right) f_g(\xi, Q^2) \right] \\ \frac{df_g(x, Q^2)}{d\log(Q^2)} &= \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} \left[P_{gq}\left(\frac{x}{\xi}\right) f_{q_i}(\xi, Q^2) + P_{gg}\left(\frac{x}{\xi}\right) f_g(\xi, Q^2) \right] \end{aligned} \tag{1.17}$$

where the first equation is for quarks and the second for gluons. The P_{ab} functions are the splitting functions which accounts for the probability that a parton a with momentum fraction ξ emits a parton b with a fraction of momentum $z = \frac{x}{\xi}$. At Leading Order the splitting functions have the following expressions:

$$\begin{aligned} P_{qq}(z) &= \frac{4}{3}\left[\frac{1+z^2}{1-z}\right] \\ P_{gg}(z) &= 3\left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z)\right] \\ P_{gq}(z) &= \frac{4}{3}\left[\frac{1+(1-z)^2}{z}\right] \\ P_{qg}(z) &= \frac{1}{2}\left[z^2 + (1-z)^2\right] \end{aligned} \tag{1.18}$$

As for the renormalization scale we can replace Q^2 in Equations 1.17 with μ_F.
An example of PDF evolution can be seen in Figure 1.2.

1.3 Higgs mechanism

In a gauge theory the mass of all gauge bosons and fermions is zero, but it is experimentally known that particles in SM such as electroweak gauge bosons $W^{\pm}$ and Z^0 have mass. However an explicit mass term of the form $M^2 V^{\mu} V_{\mu}$ for bosons is not allowed in the SM

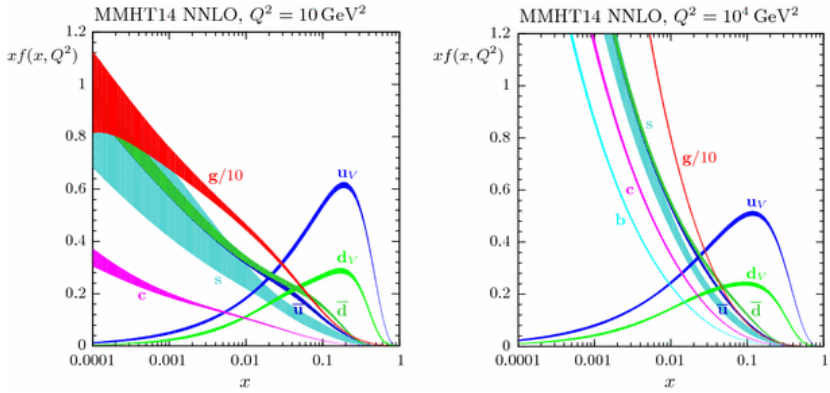


Figure 1.2: NNLO Parton Distribution Functions (PDFs) evaluated at $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 with associated 68% C.L. uncertainty bands for the MMHT2014 set [19].

lagrangian because it breaks local gauge invariance. In addition also the mass term for fermions introduced by hand in the lagrangian, which would be of the form $-m_f\bar{\psi}\psi$, is not gauge invariant and this is clear if we decompose the expression in helicity states:

$$-m_f\bar{\psi}\psi = -m_f(\bar{\psi}_R + \bar{\psi}_L)(\psi_L + \psi_R) = -m_f(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R) \tag{1.19}$$

Since ψ_L (left-handed, member of an isospin doublet, $I = \frac{1}{2}$) and ψ_R (right-handed, isospin singlet, $I = 0$) behave differently under rotations these terms are not gauge invariant:

$$\psi_L{}' \to \psi_L = e^{i\alpha(x)T + i\beta(x)Y}\psi_L \tag{1.20}$$

$$\psi_R{}' \to \psi_R = e^{i\beta(x)Y}\psi_R \tag{1.21}$$

In 1964 Higgs [3] and, Brout and Englert [4] in a indipendent way, proposed a mechanism which is able to introduce mass terms preserving at the same time invariance under $SU(2)_L \times U(1)_Y$.
The *Higgs Mechanism* of *spontaneous symmetry breaking* is introduced by considering an isospin doublet of complex scalar fields $\Phi(x)$ and a scalar potential $V(\Phi)$:

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} \tag{1.22}$$

$$V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 \tag{1.23}$$

where the complex scalar field has weak hypercharge $Y(\Phi) = 1$ and the coefficient λ is positive to ensure the potential V is limited from below.
The Lagrangian for the Higgs Sector can be written as:

$$\mathcal{L}_{Higgs} = (\mathcal{D}_\mu\Phi)^\dagger(\mathcal{D}^\mu\Phi) - V(\Phi) \tag{1.24}$$

where the covariant derivative is that of equation (1.3). The choice of $\mu^2 > 0$ gives a mass term to Φ and implies a trivial minimum of the potential at $\Phi = 0$. In this situation we have a vacuum at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ where the exact symmetry of the Lagrangian is preserved. In the situation of $\mu^2 < 0$ we have an infinite number of minima (vacuaa), as shown in Figure (1.3). From the infinite number we can choose the following minimum:

$$< \Phi >= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.25}$$

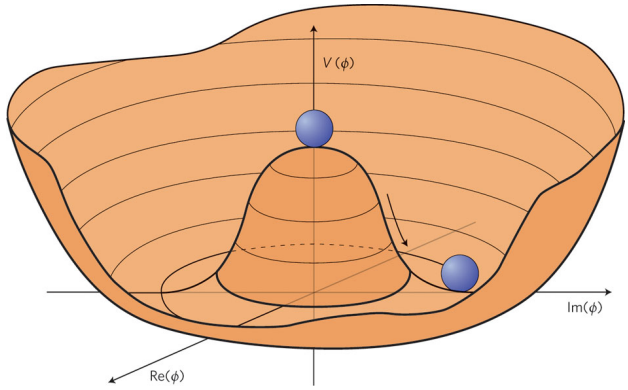


Figure 1.3: Higgs potential for a scalar field with $\mu^2 < 0$.

The choice of a particular value of the vacuum breaks the electroweak symmetry. The particle spectrum is obtained by studying the Lagrangian under small oscillations around the minimum and then the scalar doublet can be rewritten as:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + h + i\eta \end{pmatrix} \tag{1.26}$$

where v is the vacuum expectation value of Higgs potential with value $v = \sqrt{-\frac{\mu^2}{2\lambda}} > 0$, h is a scalar field, the Higgs field, and the other three fields are massless scalars. These three degrees of freedom corresponds to the Goldstone bosons due to the breaking of three out of four generators of the $SU(2)_L \times U(1)_Y$ symmetry group and they can always be brought to zero choosing a specific gauge called the *unitary gauge*. The resulting kinetic term for the scalar fields, after using hypercharge and isospin operators is (omitting the h-mixed terms):

$$\begin{aligned} \mathcal{L}_{kin} &= (\mathcal{D}^\mu \Phi)^\dagger (\mathcal{D}_\mu \Phi) \\ &= \frac{v^2}{8} \left| (g\tau^k W_\mu^k + g' B_\mu) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right|^2 \\ &= \frac{v^2}{8} \left[g^2 \left((W_\mu^1)^2 + (W_\mu^2)^2 \right) + (g W_\mu^3 - g' B_\mu)^2 \right] \end{aligned} \tag{1.27}$$

Using the relations 1.9 and 1.10 we can rewrite the Lagrangian in terms of physics fields as:

$$\frac{1}{8}v^2[g^2(W^+)^2 + g^2(W^-)^2 + (g^2 + g'^2)Z^2 + 0 \cdot A^2] \tag{1.28}$$

where we can identify the mass terms for the gauge bosons:

- $M_{W^+} = M_{W^-} = \frac{1}{2}vg$
- $M_Z = \frac{1}{2}v\sqrt{(g^2 + g'^2)}$
- $M_\gamma = 0$

We found that Higgs boson couples only to massive particles and the couplings to gauge bosons are proportional to M_V^2.
Fermions acquire mass through a Yukawa interaction between the scalar field Φ and the fermionic field:

$$\mathcal{L}_f = -\lambda_f[\bar{\psi}_L \Phi \psi_R + \bar{\psi}_R \bar{\Phi} \psi_L] \tag{1.29}$$

where λ_f are Yukawa couplings. These terms are invariant not only under $SU(2)_L$ but also under $U(1)_Y$ due to the Higgs weak hypercharge.
In the leptonic case, using the unitary gauge, we have:

$$\begin{aligned}\mathcal{L}_e &= -\lambda_e \frac{1}{\sqrt{2}}\Big[\begin{pmatrix}\bar{\nu}, & \bar{e}\end{pmatrix}_L \begin{pmatrix}0 \\ v+h\end{pmatrix} e_R + \bar{e}_R \begin{pmatrix}0, & v+h\end{pmatrix} \begin{pmatrix}\nu \\ e\end{pmatrix}_L \Big] \\ &= -\frac{\lambda_e v}{\sqrt{2}}\bar{e}e \; - \; \frac{\lambda_e}{\sqrt{2}}h\bar{e}e\end{aligned}$$

The Yukawa coupling costants are proportional to fermions mass, $\frac{\lambda_f}{\sqrt{2}} = \frac{m_f}{v}$ and the λ_f are free parameters, they are not predicted by the theory.
In the quark case, the Yukawa coupling mentioned before is not sufficient: it gives mass only to down components of the isospin doublet; for giving the mass also to the up component we need to introduce a doublet of complex fields and the Lagrangian with the Yukawa coupling for "up-like" particles is defined as follows:

$$\mathcal{L}_{up} = Y_{ij}^{up} \bar{\psi}_i^L \tilde{\Phi}^c u_j^R \tag{1.30}$$

where i and j are flavour indices, Y_{ij} is the Yukawa matrix which contains masses for the different flavours and Φ^c is the charge conjugate field of Φ.
The potential (1.23) can be expressed as:

$$V(\Phi) = const + 2\mu^2\Phi^2 + 2\sqrt{2}\frac{\mu}{\sqrt{\lambda}}\Phi^3 + \lambda\Phi^4 \tag{1.31}$$

where the cubic and the 4th order terms in Φ are the Higgs self-interaction terms. It is also possible to recognize a mass term for the Higgs:

$$m_H = \sqrt{2}\mu \tag{1.32}$$

The SM Lagrangian can then be completed by adding the Yukawa and Higgs terms:

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \tag{1.33}$$

1.4 Higgs boson phenomenology

The discovery of a new particle compatible with the SM Higgs boson was announced by the ATLAS and CMS experiments on July 4th 2012 [1, 2]. Since then many measurements have been performed at the LHC concerning the Higgs boson mass, cross section and couplings, and properties, confirming the SM nature of the Higgs. However before the Higgs discovery, theoretical limits on the Higgs boson mass were available, even though the theory does not predict the exact value of the mass. Moreover experimental limits came from searches at Tevatron and LEP. The knowledge about the Higgs boson before the LHC era will therefore be summarised in the following section, followed by the Higgs phenomenology description and measurements at the LHC.

1.4.1 Theoretical and experimental constraints on the Higgs boson mass

The theory, as seen in the previous section, is not able to predict Higgs mass as it depends on the λ parameter which is a free parameter of the spontaneous symmetry breaking. However there are some theoretical constraints which limit the Higgs mass range.
The first limit comes from the *unitarity* requirement. In particular the $W^+W^- \rightarrow W^+W^-$ scattering involves the quadratic gauge coupling with exchanges of Z, γ and possibly the Higgs Boson. If the boson is too heavy or does not exist, the unitarity is violated unless some new physics appears at the scale of 1 TeV or the Higgs is lighter than 800 GeV.
The *triviality bound* sets an upper limit to the Higgs mass. The SM remains valid as long as the Higgs self-coupling is finite (finite values of λ). As this coupling increases with energy, a cut-off Λ must be set, after which some new physics takes over. If the SM is required to remain valid up to the electroweak scale (10^2 GeV), then Higgs self-coupling is constrained and therefore the upper limit on the Higgs mass is 1 TeV. If its validity extends to the GUT (10^{16} GeV), then are allowed masses for the Higgs up to 160 GeV.
The *stability bound* sets a lower limit to the range of Higgs mass by the requirement that the minimum of the potential be lower than that of the unbroken theory and that the electroweak vacuum is stable. For small values of Higgs self-coupling and therefore for the Higgs mass, the dominant contribution comes from the top quark through the Yukawa coupling. Since this contribution is negative, the value of λ may become negative and at this point there is no minimum anymore.
Combining all these effects, the theoretical limits on the Higgs mass are represented in Figure 1.4.

Before the discovery of the Higgs at the LHC, the limits on its mass came from direct search at Tevatron and LEP. The first phase of LEP at the energy of the Z pole, excluded the mass range below 65.2 GeV. The upgraded LEP2 increased the center of mass energy up to 209 GeV, looking at the "Higgs-strahlung" production ($e^+e^- \rightarrow Z^* \rightarrow HZ$) and in

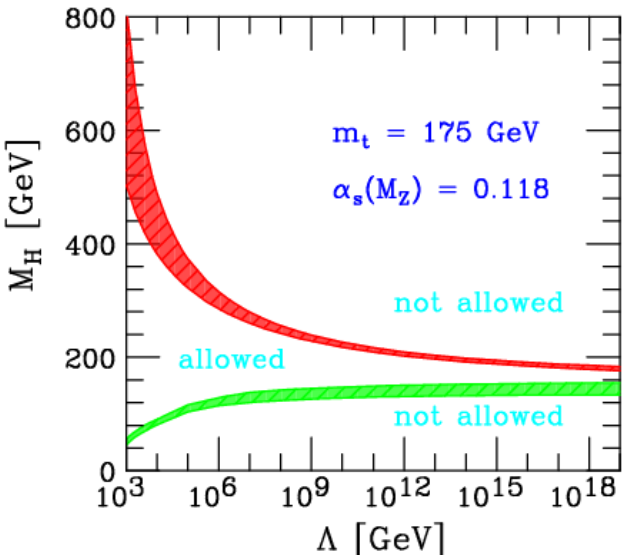


Figure 1.4: Theoretical limits on the Higgs boson mass from the triviality bound (upper limit) and the vacumm stability (lower limit) as a function of the cut-off Λ. The allowed regions lies between the two bands and the colored band represent the impact of various uncertainties [20].

this phase the range $m_H < 114.4$ GeV was excluded at the 95% confidence level(CL) [21]. Tevatron was a $p\bar{p}$ collider at Fermilab, which operated until 2011 reaching the center of mass energy of 1.96 TeV. Using the data from 10^{-1} fb^{-1} of integrated luminosity and combining the results from the two experiments, CDF and $D0$, the excluded regions at 95% CL are between 90 and 109 GeV and between 149 and 182 GeV [22]. The results from LEP and from the TEVATRON are summarized in Figure 1.5 and Figure 1.6.

The most stringent limits came from the electroweak precision data from the LEP Electroweak Working Group [23] and more recently by the GFitter collaboration [24] where constraints on Higgs mass are obtained by using a fit over all SM variables as shown in Figure 1.7.

1.4.2 Higgs boson production modes at the LHC

In a $p\bar{p}$ collider as the LHC, the Higgs can be produced through four dominant processes (leading-order Feynman diagrams are shown in Figure 1.8):

- **Gluon Fusion (ggF)**. The Higgs boson cannot couple to massless particles such as gluons, neverthless this coupling is allowed through a quark loop where the main contribution is from the top quark. It has the highest cross section due to the abundance of gluons in the proton and the large Yukawa coupling of the top quark. The theoretical calculation has reached the N^{3}LO accuracy and predicts

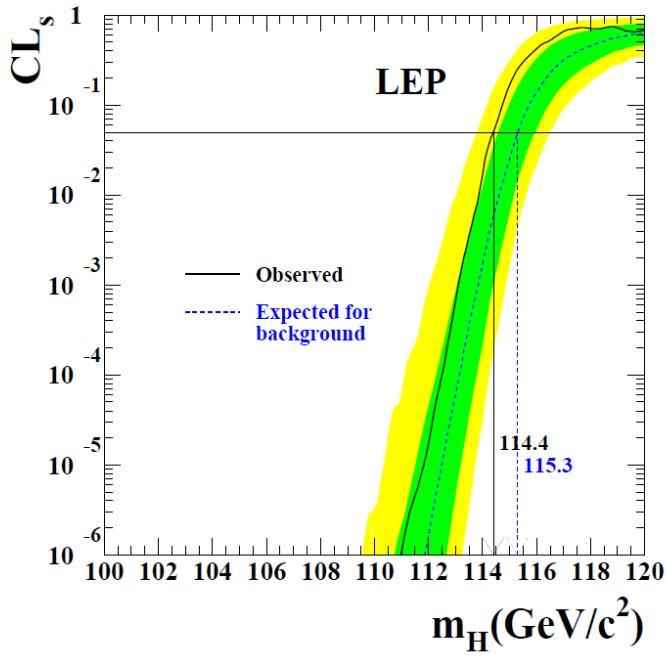


Figure 1.5: Confidence level of signal plus background hypothesis on the SM Higgs boson at LEP. Solid line: observation; dashed line: median background expectation. The dark and light shaded bands around the median expected line correspond to the 68% and 95% probability bands. The intersection of the horizontal line for $CL_s = 0.05$ with the observed curve is used to define the 95% confidence level lower bound on the mass of the Standard Model Higgs boson [21].

the following value for a ggF production cross section of a Higgs boson with $m_H = 125$ GeV at a center of mass energy of 13 TeV [25]:

$$\sigma_{ggF} = 48.58 \text{ pb } {}^{+4.6\%}_{-6.7\%} \text{ (theory)} \pm 3.2\% \text{ (PDF} + \alpha_s) \tag{1.34}$$

The systematic uncertainties referred to as "theory" are due to the missing higher order QCD terms, the EW corrections, the effects of missing b and c quark mass and the corrections due to the top quark mass effects whereas the "PDF + α_s" are due to the PDF set and the α_s value used.

- **Vector Boson Fusion (VBF)**. It is the second most abundant process at the LHC where the two quarks from the colliding protons irradiate a pair of electroweak gauge bosons that merge producing the Higgs. It is characterized by a clear experimental signature with two high-p_T jets located in the forward region with a large rapidity separation between them. The NNLO prediction for the VBF production cross

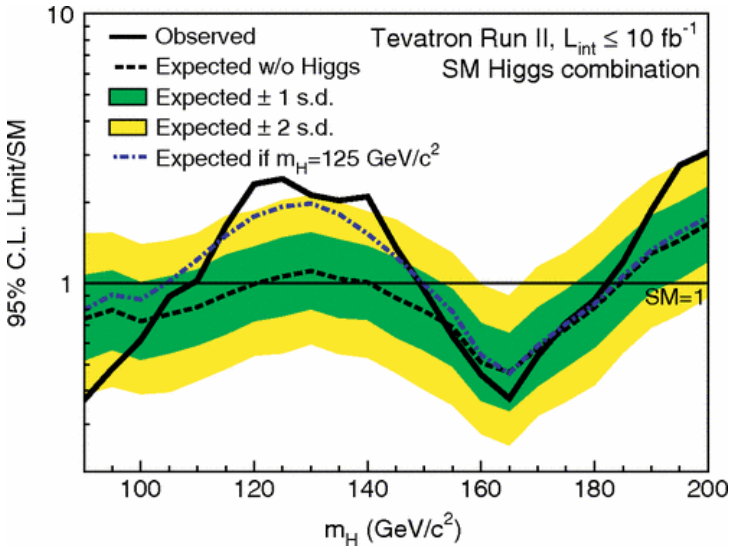


Figure 1.6: Expected (dashed line) and observed (solid line) 95% CL upper limits on SM Higgs boson at the Tevatron. Regions of massess between 90 and 109 GeV and between 149 and 182 GeV, for which the observed curve lies below 1, are excluded at 95% CL [22].

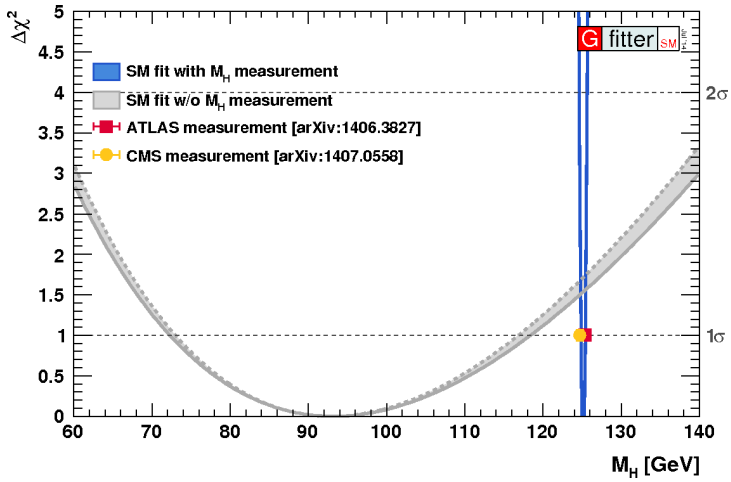


Figure 1.7: $\Delta\chi^2$ of the global electroweak fit without the M_H measurement shown as the grey band. Also shown is the result of the fit without the M_H measurements (grey band). The solid and dashed lines give the results when including and ignoring theoretical errors, respectively [24].

section of a Higgs boson with $m_H = 125$ GeV at a center of mass energy of 13 TeV is [25]:

$$\sigma_{VBF} = 3.782 \text{ pb } ^{+0.4\%}_{-0.3\%} \text{ (QCD scale)} \pm 2.1\% \text{ (PDF} + \alpha_s) \tag{1.35}$$

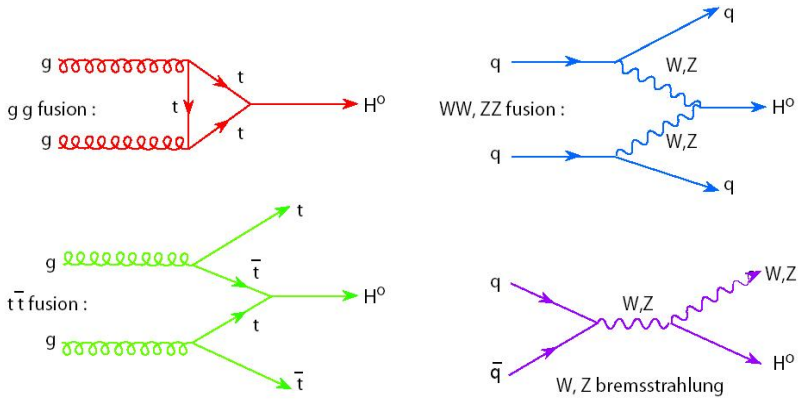


Figure 1.8: Feynman diagrams for the Higgs production mechanisms at the LHC.

The systematic uncertainties are here due to the missing higher order QCD terms (QCD scale) and to the PDF and α_s. The same uncertainties are also calculated for the following processes.

- **Associated production with Z and W (Higgs-strahlung)**. It is the third most abundant process where a produced high energy weak boson irradiates an Higgs. The final state is therefore composed by a Higgs boson and a W or Z, then the W or Z decays in jets or in leptons/neutrinos. The NNLO prediction for the ZH/WH production cross sections of a Higgs boson with $m_H = 125$ GeV at a center of mass energy of 13 TeV are [25]:

$$\begin{aligned} \sigma_{WH} &= 1.373 \text{ pb } {}^{+0.5\%}_{-0.7\%} \text{ (QCD scale)} \pm 1.9\% \text{ (PDF)} + \alpha_s) \\ \sigma_{ZH} &= 0.8839 \text{ pb } {}^{+3.8\%}_{-3.1\%} \text{ (QCD scale)} \pm 1.6\% \text{ (PDF)} + \alpha_s) \end{aligned} \tag{1.36}$$

- **Associated production with top-antitop quark pair (ttH)**. It is the last production process in terms of cross section among the ones measured at the LHC. The Higgs boson is produced in association wih a top-antitop quark pair. The final state is therefore composed by a Higgs boson and the $t\bar{t}$ pair that consequently decay in a b-quark plus a $W^{\pm}$. The NLO prediction for the ttH production cross section of a Higgs boson with $m_H = 125$ GeV at a center of mass energy of 13 TeV is [25]:

$$\sigma_{ttH} = 0.5071 \text{ pb } {}^{+5.8\%}_{-9.2\%} \text{ (QCD scale)} \pm 3.6\% \text{ (PDF)} + \alpha_s) \tag{1.37}$$

Similar to the ttH production mode, is also worth to mention the bbH process where a Higgs boson is produced in association with a bottom-antibottom quark pair. Its cross section, obtained by matching calculation at NNLO with the five flavour scheme and NLO with the four flavour scheme, is [25]:

$$\sigma_{bbH} = 0.488 \text{ pb } {}^{+20.2\%}_{-23.9\%} \text{ (QCD scale + PDF + } \alpha_s) \tag{1.38}$$

Decay	Branching Ratio	Error (%)
$H \to b\bar{b}$	5.77×10^{-1}	$^{+5.0}_{-4.9}$
$H \to W^+W^-$	2.15×10^{-1}	$^{+4.3}_{-4.2}$
$H \to gg$	8.6×10^{-2}	$^{+10.22}_{-9.98}$
$H \to \tau^+\tau^-$	6.32×10^{-2}	± 5.7
$H \to ZZ$	2.64×10^{-2}	$^{+4.3}_{-4.1}$
$H \to c\bar{c}$	2.68×10^{-2}	$^{+12.17}_{-12.21}$
$H \to \gamma\gamma$	2.28×10^{-3}	$^{+5.0}_{-4.9}$
$H \to Z\gamma$	1.54×10^{-3}	$^{+9.0}_{-8.9}$
$H \to \mu^+\mu^-$	2.19×10^{-4}	$^{+6.0}_{-5.9}$

Table 1.1: Higgs branching ratios for a Higgs of mass $m_H = 125$ GeV.

In Figure 1.9 the rates of the production mode, as explained above, are given as a function of the Higgs mass.

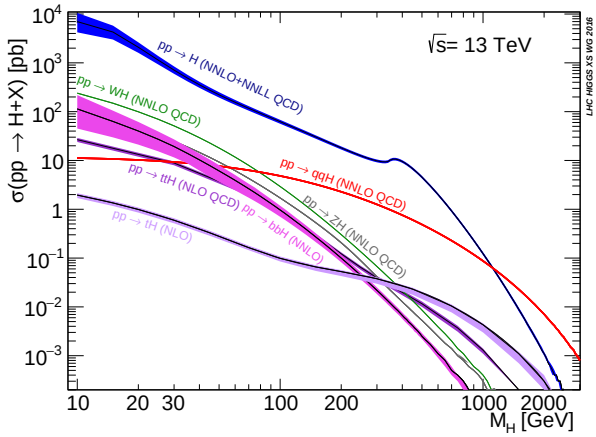


Figure 1.9: Cross sections for the Higgs production mechanism at a center of mass energy of 13 TeV as a function of the Higgs mass. Thickness of the curves indicates the level of uncertainty in the calculation [25].

1.4.3 Higgs boson decays

The Higgs boson couples to any massive particle with a strength of the coupling proportional to its mass. For this reason the Higgs decays mostly to the heaviest kinematically accessible particle pair. In table 1.1 are reported the branching ratios of a Higgs of mass $m_H = 125$ GeV and in Figure 1.10 the Higgs branching ratios as a function of the Higgs mass [25].

For a Higgs of mass about 125 GeV the decay with the highest probability is the decay in pair of b-quark but is dominated by QCD background and the problem of b-tagging.

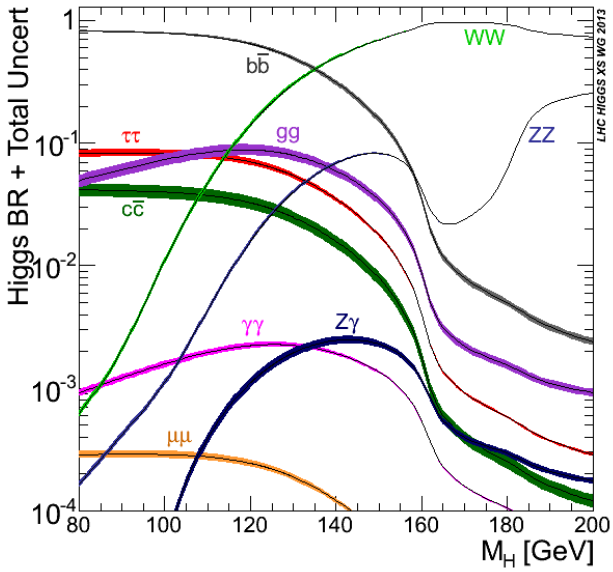


Figure 1.10: Higgs branching ratios as a function of the Higgs mass [25].

1.4.4 Higgs boson measurements

On July 4th 2012, the ATLAS and CMS collaborations announced the discovery of a new particle compatible with the predicted Standard Model Higgs boson [1, 2]. The ATLAS measurement used a dataset corresponding to the integrated luminosities of 4.8 fb^{-1} collected at $\sqrt{s} = 7$ TeV in 2011 and 5.8 fb^{-1} collected at $\sqrt{s} = 8$ TeV in 2012. Clear evidence for the production of a neutral scalar boson with a measured mass of $126.02 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ GeV was shown. This measurement combined the individual measurements in the $H \to ZZ^* \to 4l$, $H \to \gamma\gamma$ and $H \to WW^* \to e\nu\mu\nu$ channels for the 8 TeV data and $H \to ZZ^*$, $\gamma\gamma$, $WW*$, $b\bar{b}$ and $\tau^+\tau-$ for the 7 TeV data. The resonance was observed with a significance of 5.9 standard deviations (Figure 1.11).

Using the full Run 1 dataset, corresponding to an integrated luminosity of 25 fb^{-1}, ATLAS and CMS performed a combined mass measurement in the $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ channels [5]. They found a combined measured mass of the Higgs boson of $m_H = 125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ GeV, with a simultaneous fit to the reconstructed invariant mass peaks in the two channels and for the two experiments (Figure 1.12). ATLAS and CMS also performed a combined measurement of the Higgs couplings [6]. This measurement was performed using the reduced coupling strength scale factors of the Higgs to vector bosons and to fermions, which are defined respectively as $y_{V,i} = \sqrt{k_{V,i}} g_{V,i}/2v = \sqrt{k_{V,i} c m_{V,i}} v$ and $y_{F,i} = k_{F,i} g_{F,i}/\sqrt{2} = k_{F,i} c m_{F,i}/v$, following a leading order framework, the k-framework. The results are shown in Figure 1.13 and they are all compatible with the SM expectation within the uncertainties.

The Run 2 of the LHC started in 2015 and finished in 2018. During these years LHC operated at $\sqrt{s} = 13$ TeV and the ATLAS detector collected data corresponding to an

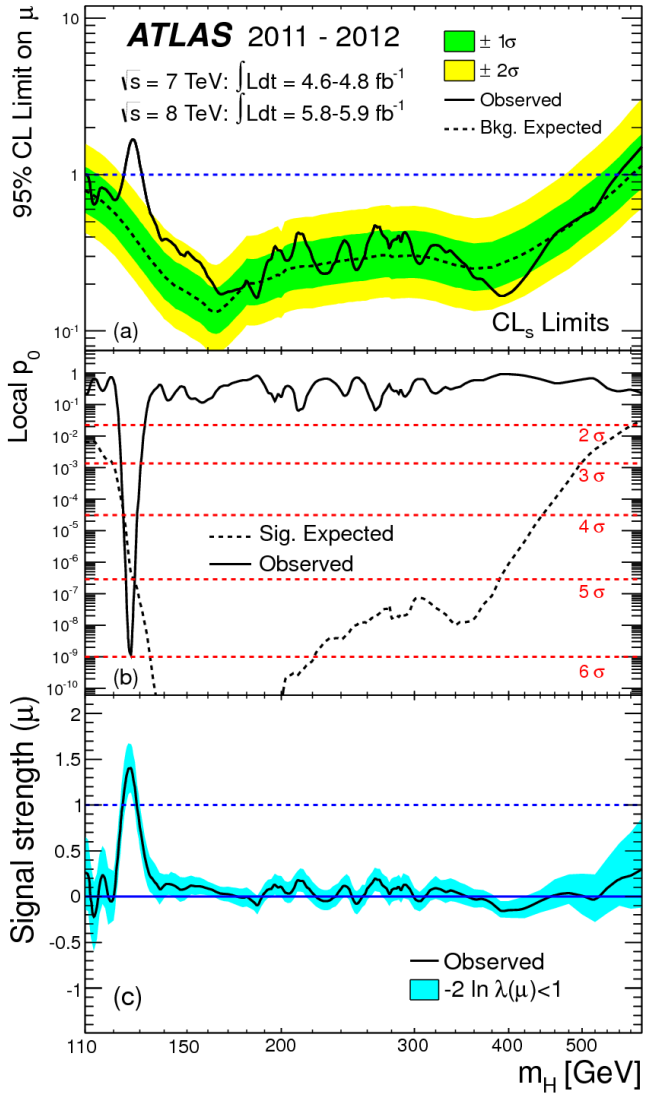


Figure 1.11: Combined search results by the ATLAS experiment during Run 1: (a) The observed (solid) 95% CL upper limit on the signal strength as a function of m_H and the expectation (dashed) under the background-only hypothesis. The dark and light shaded bands show the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the background-only expectation. (b) The observed (solid) local p_o as a function of m_H and the expectation (dashed) for a SM Higgs boson signal hypothesis ($\mu = 1$) at the given mass. (c) The best-fit signal strength μ as a function of m_H. The band indicates the approximate 68% CL interval around the fitted value [1].

integrated luminosity of 147 fb^{-1}. Several analyses have been performed with a new level of precision and there are still many of them ongoing. In the Higgs sector the studies have

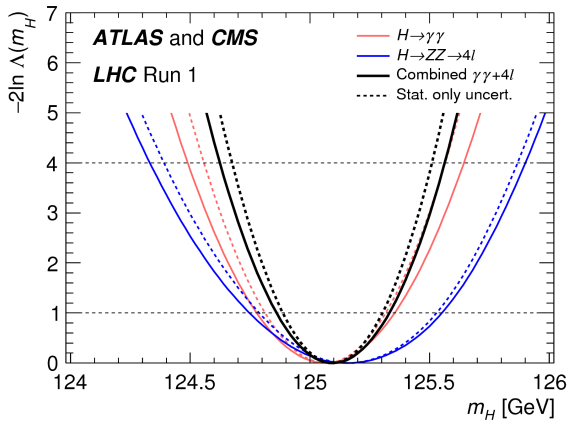


Figure 1.12: Scans of twice the negative log likelihood ratio $-2\ln\Lambda(m_H)$ as functions of the Higgs boson mass m_H for the ATLAS and CMS combination of the $H \to \gamma\gamma$ (red), $H \to ZZ \to 4l$ (blue), and combined (black) channels. The dashed curves show the results accounting for statistical uncertainties only, with all nuisance parameters associated with systematic uncertainties fixed to their best fit values. The 1 and 2 standard deviation limits are indicated by the intersections of the horizontal lines at 1 and 4, respectively, with the log likelihood scan curves [5].

been concentrated mostly on the measurements of Higgs boson couplings and properties as spin/CP quantum numbers looking for confirmation of the SM or any possible hints of Beyond the Standard Model (BSM) physics.

1.5 The limitations of the Standard Model

The SM is a successfull and well-tested theory and provides the description of many known phenomena. However it has some critical issues arising either from the theory or from some observed unexplained phenomena.

Hierarchy Problem

From the theoretical side this is the main problem [27]. In the SM there are two fundamental scales: the electroweak scale (~ 100 GeV) and the Planck mass scale($\sim 10^{18}$ GeV) where the gravity interaction becomes not negligible compared to the other gauge interactions. The huge difference between these two scales is somehow considered unnatural and the other related issue is that fine-tuning cancellation are required in the Higgs boson mass calculation. The Higgs mass has quadratically divergent radiative corrections coming from virtual effects of every SM particle. The physical mass of the Higgs is then:

$$m^2_{H,physical} = m^2_{H,bare} + \delta m^2_H \tag{1.39}$$

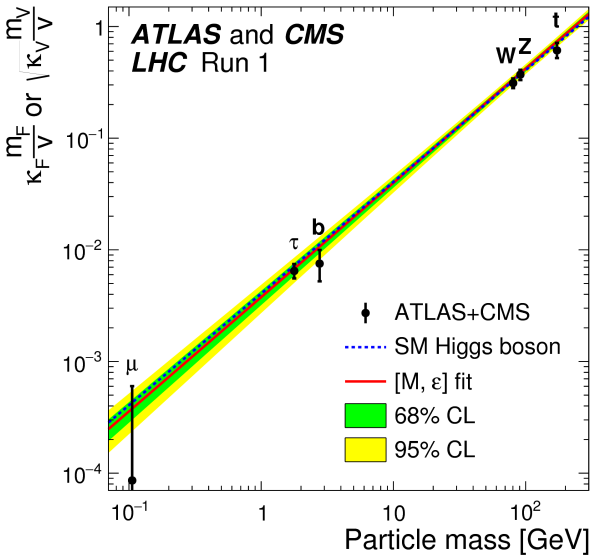


Figure 1.13: Best fit values as a function of particle mass for the combination of ATLAS and CMS data, with parameters defined as $k_F \frac{m_F}{v}$ for the fermions and as $\sqrt{k_V}\frac{m_V}{v}$ for the weak vector bosons where $v = 246$ GeV is the vacuum expectation value of the Higgs field. The dashed (blue) line indicates the predicted dependence on the particle mass in the case of the SM Higgs boson. The solid (red) line indicates the best fit result to the $[M, \epsilon]$ phenomenological model of Ref. [26] with the corresponding 68% and 95% CL band [6].

The correction, δm_H^2, due to a fermion loop is:

$$\delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda^2 + ... \tag{1.40}$$

where λ_f is the Yukawa coupling and Λ is the cut-off scale at which new physics should appear. The leading term is the one coming from the top quark loop and the omitted term is a logarithmic function of Λ. If we assume that the SM is valid until the Planck scale, Λ can be of the order of the Planck scale and this requires a precise cancellation in order to have a Higgs boson with mass ~ 125 GeV. SM theory requires indeed fine-tuning and this violates the *naturnalness* principle. This kind of theories are not incorrect in the sense of falsifying observations but they are considered problematic in the absence of a mechanism that would explain why the various parameters have exactly the needed values.

Gravity

The gravitational interaction is still not included in the SM as there is no quantum gravity theory. Since the gravity coupling constant is much smaller compared to the other interactions, the effect is negligible at the energies accessible today at accelerators. These effects become important only at the Planck scale.

Dark matter
Many experimental observations [28, 29] have found that only the $\sim 5\%$ of the Universe is composed by the known barionic matter while the remaining $\sim 27\%$ and $\sim 68\%$ is made up by *dark matter* and *dark energy* respectively. Recent WMAP [30] and Planck measurements [31] seems to suggest a cold dark matter hypothesis and in this sense there are many possible Dark matter candidates as Massive Compact Halo Objects (MACHOs), axions and Weakly Interacting Massive Particles (WIMP). MACHOs consist of large condensed objects that don't emit light such as black holes, neutron stars, white dwarfs. The search for these objects is performed through gravitational lensing and it seems to rule out this hypothesis. Axions are neutral pseudo-scalr particles with small masses and very low interaction cross sections for weak and strong interactions. Their existence would also solve the strong CP problem. Finally, the WIMP are neutral massive particles which interact only via gravity and with any other force but with the cross section of the order of the weak interaction. There are no candidates among the SM particles but many extensions of the SM are able to provide a suitable candidate for dark matter particle.

Neutrino
Neutrinos are predicted to be massless in the SM, however some recent measurements on neutrino oscillations suggest that they should have very small masses (< 1 eV). The neutrinos cannot constitute a valid dark matter candidate but many extensions of the SM predict the existence of heavy right-handed Majorana neutrinos that can have the requested features for being dark matter candidates.

Matter-Antimatter asymmetry
In the early universe matter and antimatter were produced in equal amounts and this is in contradiction with the asymmetry that we observed nowaday. In 1967 Sakharov [7] pointed out the three necessary conditions to explain the asymmetry:

- At least one B-number violating processes
- C and CP violation
- Interactions outside of thermal equilibrium

In particular the CP violation must be in a higher amount compared to the CP violation in the SM which is described by the CKM matrix [32].
For all these reasons many Beyond the Standard Model theories (BSM) have been developed since '70's like supersymmetry, technicolor, string theory, extra-dimensions. Many of them introduce additional scalar fields like additional Higgs doublets in the *Minimal Supersymmetric Standard Model* or provides new models for baryogenis.
Another way to study BSM physics effects in the Higgs sector is to use an *Effective Field Theory* (EFT) that features only SM particles and symmetries at the EW scale. This is a minimal choice but it is well justified by the present data and in addition it reduces the possible interactions that Lorentz symmetry alone would allow and gives a powerful framework where constraints can be analyzed. An EFT is a non renormalisable theory with dimensional coupling constants and it is an approximation at low energy of a theory

of higher energy where some new physics appears. An EFT requires an effective Lagrangian which is constructed with all the gauge-invariant local operators of the lowest dimension, for example:

$$\mathcal{L}_{eff} = \mathcal{L}_{SM} + \sum_i c_i O_i, \quad c_i \sim \frac{1}{\Lambda^{d_i-4}} \tag{1.41}$$

the $\mathcal{L}_{SM}$ has dimension-four operators, therefore the lowest dimension operators that extend SM are of dimension six with respect to the SM gauge symmetry. In this way the model and the Lagrangian are simple but general enough to describe any new physics effects coming from an higher scale in a model-independent way. It is known that the dimension-six basis contains 59 operators [33] and for this reason, searches for new physics always make a choice of an operator basis. This digression on EFT will be resume in Chapter 6 since one of the two analysis discussed in this work is developed within this framework.

1.6 CP violation

This Section will briefly describe the CP violation as this is relevant for the analysis that measures the CP quantum numbers of the Higgs, part of this thesis work that will be described in Chapter 6.
One of the three Sakharov hypothesis mentioned in Section 1.5 requires CP violation for explaining baryogenis. However the amount of CP violation requested is higher than what is predicted by the SM [32].

1.6.1 CP violation in the SM

The CP transformation combines the action of C, the *charge-conjugation* operator that acts directly on the state of the particle changing the sign of quantum charges and therefore particles and antiparticles are interchanged and P, the *parity* operator that inverts spatial cordinates in such a way that a left-handed system is transformed into a right-handed one and viceversa. C and P are conserved symmetries for electromagnetic and strong forces while they are separately violated by weak interaction. On the other hand the CP product is preserved in most of the weak processes and it is rarely violated.
The first evidence for CP violation was seen in 1966 by Val Fitch and Jim Cronin in neutral kaon decays [34]. The observed effect was small but was extremely important because it proved that matter and anti-matter are intrinsically different.
This phenomenon was explained in the context of the SM only later in 1973 by two Japanese physicists, Makoto Kobayashi and Toshihide Maskava [35], which extended the concept of Cabibbo matrix [36] and introduced the third family of quarks, which was unknown at that time.
In the quark sector, the eigenstates of the weak interaction differ from the eigenstates of mass. In order to conserve the universality of the interaction the CKM matrix was introduced. The CKM matrix is a unitary complex 3×3 matrix that connects a vector

of down-like quark mass eigenstates to a vector of the down-like interaction partners of the up-like quarks.

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.42}$$

A matrix with these characteristics can be parametrized in general by three angles and one complex phase:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{1.43}$$

It is convenient to use the Wolfenstein parametrization [37], where the CKM matrix is written through an expansion in powers of $\lambda = \sin\theta_c$ (θ_c is the Cabibbo angle with a value of $\sin\theta_c = 0.2253 \pm 0.0008$ [14]):

$$V_{CKM}^{Wolfenstein} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{1.44}$$

Here the three Euler angles and the phase are replaced by the four real quantities, λ, A, ρ and η. The phase δ is related to ρ and η through the relation $\tan\delta = \frac{\eta}{\rho}$, and it is responsible for all CP-violating phenomena in flavour-changing processes in the SM. It can be shown that CP invariance holds if the CKM matrix is real, $V_{ij} = V_{ij}^*$.
The elements of the CKM matrix can be measured through weak flavour changing decays and the most recent measurements show the following values [14]:

$$V_{CKM} = \begin{pmatrix} 0.97417 \pm 0.00021 & 0.2248 \pm 0.0006 & (4.09 \pm 0.39) \times 10^{-3} \\ 0.220 \pm 0.005 & 0.995 \pm 0.016 & (42.2 \pm 0.8) \times 10^{-3} \\ (8.2 \pm 0.6) \times 10^{-3} & (40.0 \pm 2.7) \times 10^{-3} & 1.009 \pm 0.031 \end{pmatrix} \tag{1.45}$$

The unitarity of CKM matrix imposes that $\sum_i V_{ij}V_{ik}^* = \delta_{jk}$ and $\sum_j V_{ij}V_{kj}^* = \delta_{ik}$. The six combinations can be represented as triangles in a complex plane, each with the same area. The most commonly used unitarity triangle arises from the following relation:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{1.46}$$

Equation 1.46 can be divided by the best-known product $V_{cd}V_{cb}^*$, thus obtaining a triangle where its vertices are exactly (0,0), (1,0) and $(\bar{\rho}, \bar{\eta})$ where $\bar{\rho} = \rho(1 - \frac{\lambda^2}{2})$ and $\bar{\eta} = \eta(1 - \frac{\lambda^2}{2})$. The representation of this triangle can be seen in Figure 1.14.

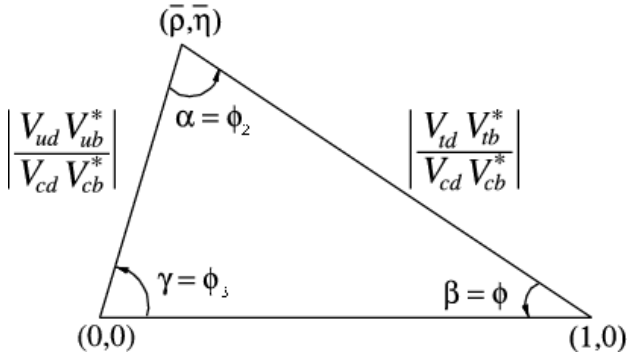


Figure 1.14: Unitarity triangle.

The angles of the unitarity triangle are then directly related to the complex phase:

$$\begin{aligned}
\beta = \phi_1 &= \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right) \\
\alpha = \phi_2 &= \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right) \\
\gamma = \phi_3 &= \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right)
\end{aligned} \tag{1.47}$$

Chapter 2

The LHC and the ATLAS detector

This Chapter is focused on the Large Hadron Collider [38] and the ATLAS experiment [39]. After a brief introduction to hadron colliders, a description of the LHC is given in Section 2.2 and the structure of the ATLAS detector is described in Section 2.3.

2.1 Hadron colliders

The Large Hadron Collider (LHC), described in Section 2.2, is a proton-proton (and heavy ions) circular accelerator and collider. Linear accelerators are easier to build since no beam steering system is needed but they have energy and luminosity limitations. In order to reach high energies a long linear accelerator is required and the beam can be used only once. Instead, in a circular accelerator only a small section is needed since the particles travels around the ring multiple times before reaching the designed energy and in addition the beam can be used several times. However the beams must be stable over a long period of time, therefore dipoles and quadrupoles must be used in order to bend and focus the beam.
The choice of particles to be accelerated is another crucial point. Elementary particles such as electrons and positrons could seem the better choice but they have some disadvantages, in particular a large amount of energy is loss due to synchrotron radiation according to the following relation:

$$\frac{dE}{dt} \propto \frac{E^4}{m^4 R} \tag{2.1}$$

where E and m are the particle energy and mass and R is the orbit radius. This implies that the energy loss is significantly huger for electrons: at fixed energy and collider orbit radius it is $(m_p/m_e)^4 \sim 10^{13}$ times higher than for the protons.
On the other side the usage of protons is not completely free from disadvantages. Protons are not elementary particles but they are made up by partons (quarks or gluons) which carry only a fraction of the total momentum of the proton. In this case there are two possible types of interactions: *soft* or *hard* collisions. The former are long distance collisions in which the protons behave as elementary particles and the particles in the final state are characterized by small transverse momentum. The hard collisions are instead short distance interactions in which the protons interact revealing their inner structure and these collisions are characterized by a high momentum exchange so that it is possible

to produce particles with high transverse momentum and mass.
Hard collisions have smaller cross sections compared to the soft ones but they are the ones of interest in most of measurements and in the search for new particles. As described in Section 1.2 the total hadron-hadron cross section is given by the factorization theorem of QCD and depends on the partonic cross section and on the PDFs.
Therefore the experiments at hadron colliders have to deal with some issues: first the background arising from soft collisions, which happens simultaneously with the hard ones, leading to the hadronization of many objects at low transverse momentum (*underlying event*) and second the fact that it is not possible to know which partons interacted and the center of mass energy at which the partonic collision happened and we have to rely on PDFs.

2.2 The Large Hadron Collider

The LHC, as mentioned in Section 2.1, is a proton-proton (and heavy ions) collider machine based at CERN, near Geneve. It uses the 27 km tunnel located underground between 50 m and 175 m depth that was built between 1984 and 1989 for the Large Electron-Positron (LEP) [40] collider machine. In 2000 LEP was dismissed to give way to the LHC. The idea of a hadron collider in the LEP tunnel was proposed in 1984 during a CERN and European Committee for Future Accelerators workshop. The LHC was then approved in December 1994 by the CERN Council and built between 2000 and 2008. The LHC is a two-ring synchrotron particle accelerator. It has 1232 dipole magnets which bend the beams into its circular orbit, whereas other 392 quadrupole magnets focus the beams. The dipoles magnetic field required to bend the proton path is 8.3 T. The resulting current and material requirements are beyond the capabilities of traditional electromagnets, thus, NbTi superconducting magnets that operate at 1.9 K are used instead. This is accomplished by using an elaborate cryogenics system, with liquid helium cooling the magnets to their operating temperature.
The first beam circulated in the LHC on September 10, 2008. Nine days later an accident during the commisioning phase caused a one year stop. Repairs and consolidation work allowed the accelerator to resume its program at the end of 2009. LHC succesfully operated between 2009 and 2012 at a center of mass energy of 7 and then 8 TeV (Run 1). After a 2 year of maintance shutdown the LHC started in 2015 a second period of data-taking operating at center of mass energy of 13 TeV and this second period ended in 2018 (Run 2).

2.2.1 The CERN accelerators and the experiments complex

The LHC is only the final step of the CERN accelerator complex as shown in Figure 2.1. A succession of four accelerators is used to accelerate the protons to the energy needed for injection into the LHC. Protons are first generated using hydrogen ionized by electric field. The first accelerating machine is the Linac 2 where the proton energy increases to 50 MeV. The next step is the Proton Synchrotron Booster (PSB) where the protons reach 1.4 GeV. The last two steps are the Proton Synchrotron (PS) and the Super Proton

Synchrotron (SPS). The injection of the particle beam in the LHC occurs at an energy of 450 GeV.

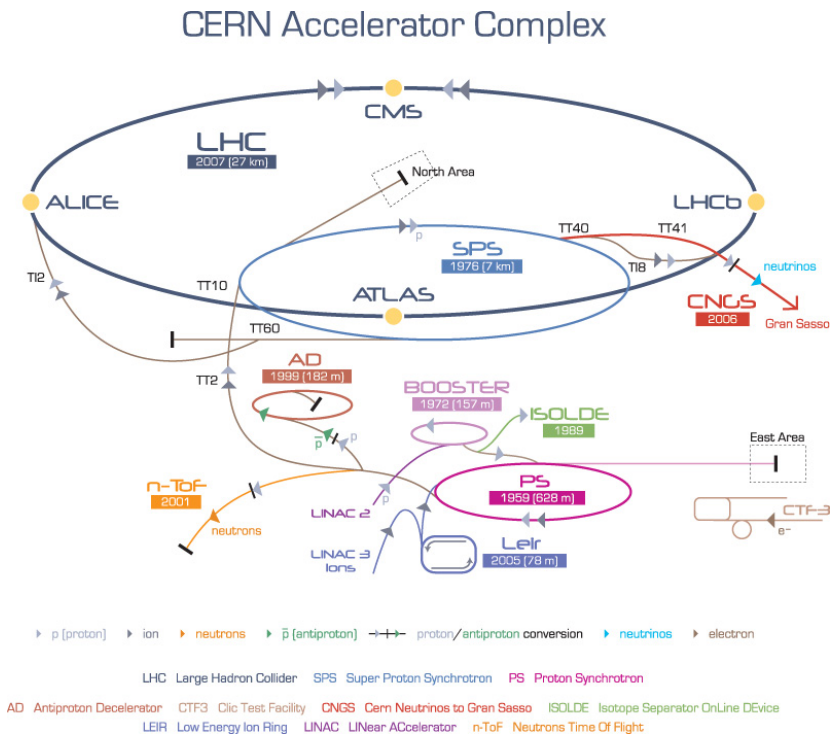


Figure 2.1: The CERN accelerator complex.

Along the LHC tunnel there are four detectors located in the collisions points, as also shown in Figure 2.1. These are ATLAS (A Toroidal LHC ApparatuS) [39], CMS (Compact Muon Spectometer) [41], ALICE (A Large Ion Collider Experiment) [42] and LHCb (Large Hadron Collider beauty experiment) [43]. ATLAS and CMS are general purpose detectors, built to be able to discover a broad spectrum of possible new physics and study the properties of the SM Higgs boson. LHCb is specialized for looking at hadrons made up of bottom quark to study b quark physics and to unveil the nature of CP violation. ALICE looks at heavy ion collisions to study the nature of the quark-gluon plasma, the state of matter believed to exist in the first fractions of time after the Big Bang.

2.2.2 Luminosity and pile-up

The total number of events at a collider is given by:

$$N = L\sigma, \quad L = \int_0^T \mathcal{L} dt \tag{2.2}$$

where σ is the inelastic $p-p$ cross section of the event and L is the integrated luminosity. $\mathcal{L}$ is the instantaneous luminosity that depends on some machine parameters:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F \tag{2.3}$$

N_b is the numbers of particles in each bunch, n_b is the number of bunches in each beam, f_r is the revolution frequency, γ is the relativistic Lorentz factor, ϵ_n is the transverse beam emittance, β^* is the focusing function at the collision point and F is the geometric luminosity reduction factor due to the crossing angle at the interaction point.
The beams are injected in the LHC in bunches with a length corresponding to 1.71 ns (reduced to 1.06 ns for collisions) and spaced by 25 ns. After the injection of all bunches, the beams are accelerated up to designed energy, 6.5 TeV during Run 2, and then focused. The LHC beam luminosity is not constant during the data taking period but decreseas with time due to the degradation of the intensity and emittance of the beams. This is mainly due to the collisions themselves in the LHC interaction points. The decay time of the beam luminosity is given by:

$$\tau_{nuclear} = \frac{N_{b,0}}{k \mathcal{L}_0 \sigma_{tot}} \tag{2.4}$$

where $N_{b,0}$ is the initial bunch intensity, $\mathcal{L}_0$ is the initial luminosity, k the number of interaction points and σ_{tot} is the proton-proton total cross section.
The integrated delivered luminosity by the LHC from 2011 is shown in Figure 2.2 per year. The integrated luminosity collected by ATLAS was 25 fb^{-1} during Run 1 (2011-2012) and 147 fb^{-1} during Run 2 (2015-2018).

Another important parameter is the mean number of inelastic interactions per bunch crossing. Increasing the instantaneous luminosity means increasing this parameter too. These events are called *pile-up events*, they mainly come from soft collisions and constitute a source of background. In particular the pile-up deteriorates the energy resolution for object measurements and makes the identification of vertices and the reconstruction of tracks more difficult. The pile-up event rate is therefore related to the instantaneous luminosity according to the relation:

$$\mu = \frac{\mathcal{L}\sigma}{n_b f_{rev}} \tag{2.5}$$

where μ is the pile-up event rate, namely the number of inelastic interactions per bunch crossing and n_b is the number of bunches. In Figure 2.3 are reported the mean number of interactions per bunch crossing which reached $< \mu >= 37.0$ in 2018.

2.3 The ATLAS experiment

The ATLAS experiment is one of the two general purpose experiments at the LHC, and is the biggest one in volume. The ATLAS detector has a cylindrical shape, providing almost full coverage of the solid angle. It has a diameter of 25 meters and it is 44 meters long with a weight of about 7000 tonnes. The ATLAS coordinates system, is a right-handed one,

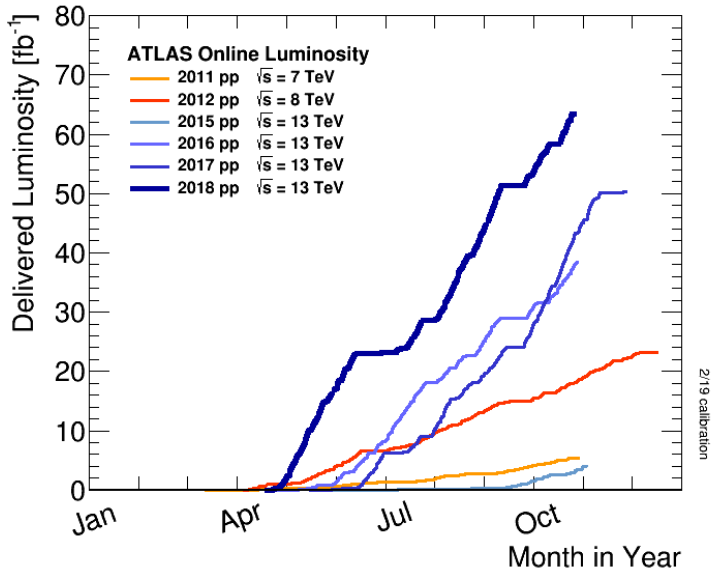


Figure 2.2: Cumulative luminosity versus day delivered to ATLAS during stable beams, for p-p collisions [44].

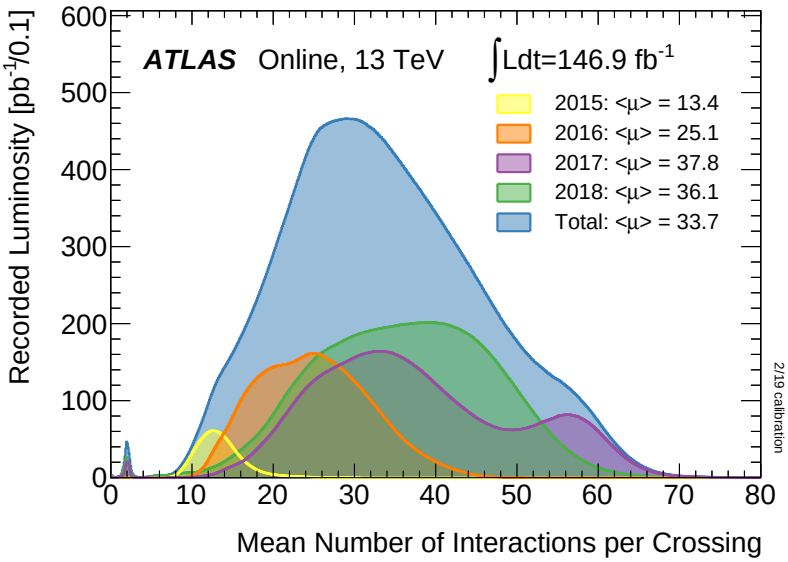


Figure 2.3: Luminosity-weighted distribution of the mean number of interaction per bunch crossing during 2015, 2016, 2017 and 2018 [44].

in which the z-axis is along the beam direction, the y-axis points upwards to the earth's surface and the x-axis points towards the center of the LHC ring. Due to the cylindrical

layout it is convenient to use a cylindrical coordinate system: the azimuthal angle ϕ runs in the transverse plane around the beam-line and the polar angle θ is measured from beam-axis (Figure 2.4).

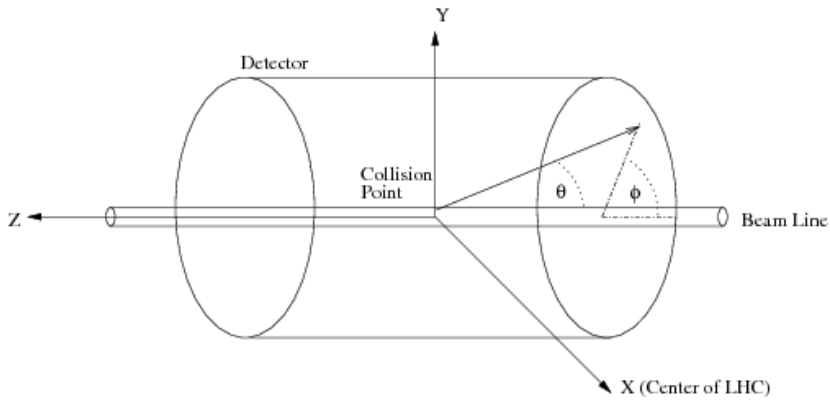


Figure 2.4: Illustration of the ATLAS coordinate system.

Most commonly the pseudorapidity η is used instead of the polar angle:

$$\eta = -\ln[\tan\frac{\theta}{2}] \tag{2.6}$$

and the distance in the $\eta - \phi$ space is defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{2.7}$$

The transverse quantities such as the transverse momentum, $\vec{p}_T$, the transverse energy E_T or the missing transverse energy E_T^{miss} are defined in the x-y plane.
The ATLAS detector is composed by layers of different subdetectors as shown in Figure 2.5:

- *Inner Detector*: the innermost part of the ATLAS detector, dedicated to tracking the charged particles and vertex reconstruction. It is immersed in a 2 T magnetic field produced by a magnetic solenoid.

- *Calorimeters*: there are three calorimeters in the ATLAS detector, namely electromagnetic, hadronic and forward calorimeters. They are placed just outside the solenoid and their purpose is to measure the energy of electrons, photons and hadrons.

- *Muon Spectrometer*: it is the outermost detector that surrounds the calorimeters. It is used for the identification and energy measurement of muons.

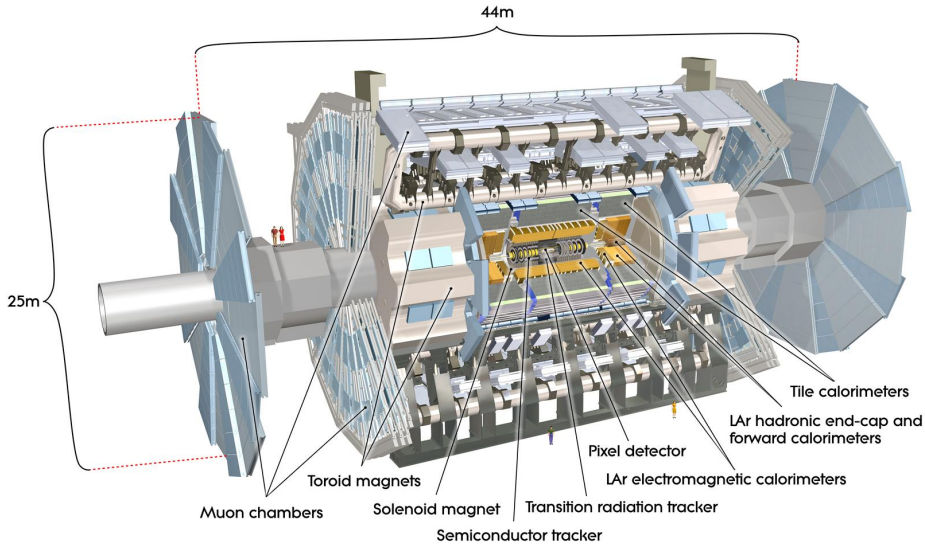


Figure 2.5: A view of the ATLAS detector with all its subdetectors.

2.3.1 Magnet system

A superconducting magnet system provides the magnetic field in which the inner detector is immersed. The magnet system [45] is composed by a central solenoid, a barrel toroid and two end-caps toroids Figure 2.6):

- **Solenoid**: the central solenoid magnet is located between the inner detector and the calorimeters and provides a 2 T axial magnetic field, generated by an 8 kA electric current, to the inner detector. It is a superconducting coil with 2.3 m diameter and 5.3 m length. The thickness is 5 cm corresponding to 0.66 radiation lenghts. In order to optimize the material budget the solenoid is placed inside the calorimeter cryostat and cooled down to a temperature of 4.5 K.

- **Toroids**: the magnetic field needed for the muon spectrometer is provided by a barrel toroid and two end-cap toroids, each of them consisting of 8 coils. The barrel has an inner diameter of 9.4 m, an outer diameter of 20.1 m and it is 25.3 m long while the end-cap have an axial length of 5 m and an outer diameter of 10.7 m. These magnets are located outside the calorimeters. The endcap toroids generate a magnetic field near the beam axis in order to deflect the particles with small polar angles and therefore together with the barrel they provide an almost full coverage of the solid angle.

2.3.2 The Inner Detector

The Inner Detector (ID) [46, 47] shown in Figure 2.7 is the innermost ATLAS subdetector. Its aim is to reconstruct tracks, vertices and measure the charged particles momentum. It

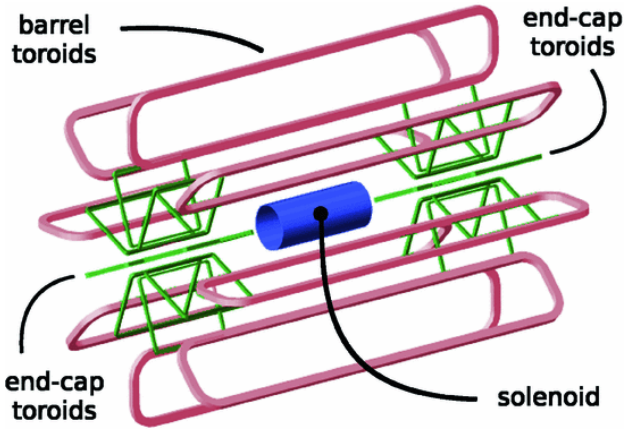


Figure 2.6: View of the ATLAS magnet system.

is designed to provide a robust pattern recognition, excellent momentum resolution and both primary and secondary vertex reconstruction capabilty. It has a cylindrical shape with a diameter of 2.1 m and a length of 6.2 m.
The ID is composed by a barrel and two end-caps with a coverage of $|\eta| \leq 2.5$. Due to its proximity to the interaction point, radiation hardness is a crucial requirement for all the components. In addition due to the high number of tracks per event, high granularity detectors are needed in order to reach high precision track parameter measurements and vertex reconstruction.

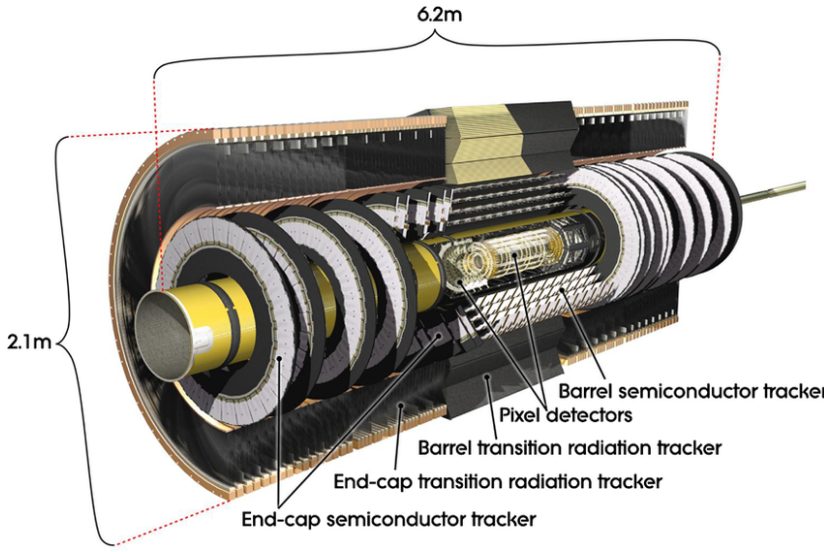
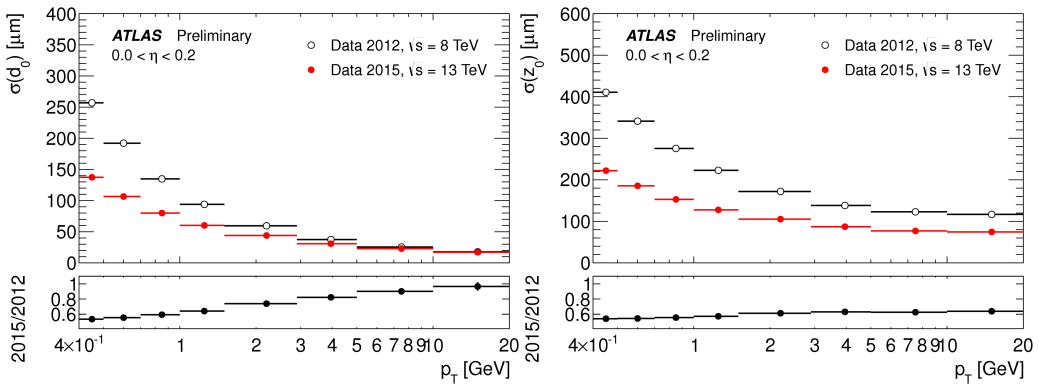


Figure 2.7: Cut-away view of the ATLAS Inner Detector.

The ID is made of three further subdetectors which exploit different technologies. The

structure is therefore the following (Figure 2.7):

- the **Pixel detector** [48] is the closest to the beam pipe. It consists of silicon sensors arranged in four cylindrical layers in the barrel region and three disks for each end-cap. The first cylindrical layer, placed at 33 mm from the interaction point, is called Insertable B-Layer (IBL) [49] and was installed during the first long shutdown after LHC Run 1, between a new smaller beam pipe and the original first layer of the Pixel detector. The IBL pixel dimensions are 50×250 μm^2. The other layers are located at radial distance of 50.5, 88.5, 122.5 mm respectively from the beam pipe, while the end-cap disks are placed at $\pm$49.5, $\pm$58 and $\pm$65 mm along the z-axis respectively. They have pixel dimensions of 50×400 μm^2.
  The Pixel detector is the one with the highest granularity and spatial resolution. These features allows an accurate vertex reconstruction and track impact parameter measurements, being very important in particular for identifying b-quark jets and tau leptons. The pixels have an intrinsic hit position resolution better than 12 μm in the transverse plane ϕ direction and 72 μm for IBL and 115 μm for the other layers along the η direction. The IBL was installed in order to have better track and vertex reconstruction performance at the luminosities expected during Run 2 and to mitigate the effect of radiation damage to the innermost layer of the Pixel detector. This can be seen in Figure 2.8 where the differences between Run 1 and Run 2 for the impact parameter resolution are shown.



(a) (b)

Figure 2.8: Transverse impact parameter (a) and longitudinal impact parameter (b) resolution as a function of the transverse momentum, p_{T}, with data taken in 2012 at $\sqrt{s} = 8$ TeV and in 2015 at $\sqrt{s} = 13$ TeV after the IBL insertion [50].

- the **Semi-conductor Tracker** (SCT) [51, 52] completes the high precision tracking of the Pixel detector. It is a silicon microstrip detector composed by four cylindrical layers outside the Pixel detector in the barrel region and nine disks for each end-cap. Each layer of the barrel is made up by two sensors mounted back to back with a 40 mrad stereo angle, in order to provide measurements of the two coordi-

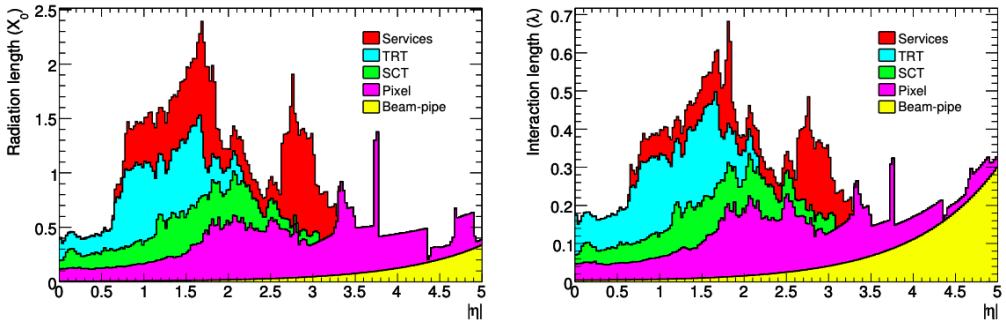


Figure 2.9: Material distribution (X_0, λ) at the exit of the ID envelope, including the services and thermal enclosures. The distribution is shown as a function of $|\eta|$ and averaged over ϕ.

nates. They are located at radial distances of 299, 371, 443 and 514 mm from the interaction point and provide 4 additional space points for each track. The intrinsic hit resolution is 17 μm in the transverse plane ($R-\phi$) and 580 μm along the z direction.

- the **Transition Radiation Tracker** (TRT) [53, 54] is the outermost part of the Inner Detector. It is composed of 4 mm straw-tubes filled with a 30 μm diameter gold-plated tungsten wire and filled with Xenon and Argon mixture. The space between the straws are filled with polymer fibres (barrel) and foils (endcaps) to create transition radiation. There are 50000 straws in the barrel and 250000 for each end-cap disposed radially in wheels. The TRT is located from 0.55 m to 1.08 m from the interaction point. It provides 36 additional points per track with an intrinsic resolution of 130 μm in the transverse plane($R-\phi$). The TRT provides then additional hits for track reconstruction and momentum measurement but it is also important for electron identification. Indeed the transition radiation is emitted when a highly relativistic particle passes across the boundaries of materials with different dielectric constants. Transition radiation photons are absorbed by the Xenon and Argon atoms depositing additional energy in the gas and leading to higher readout signals. This feature allows to strongly discriminate between electrons and pions in the range 1-200 GeV.

The overall ID momentum resolution is:

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% \cdot p_T(\text{GeV}) \oplus 1\% \tag{2.8}$$

In order to achieve high granularity and include readout and cooling system, the amount of ID material is quite sizeable and it can be seen in Figure 2.9 where it is shown in terms of radiation length X_0 and nuclear interaction lenght λ. Therefore the photons may convert into electron-positron pairs and electrons can loose part of their energy due to bremsstrahlung emissions before reaching the calorimeters.

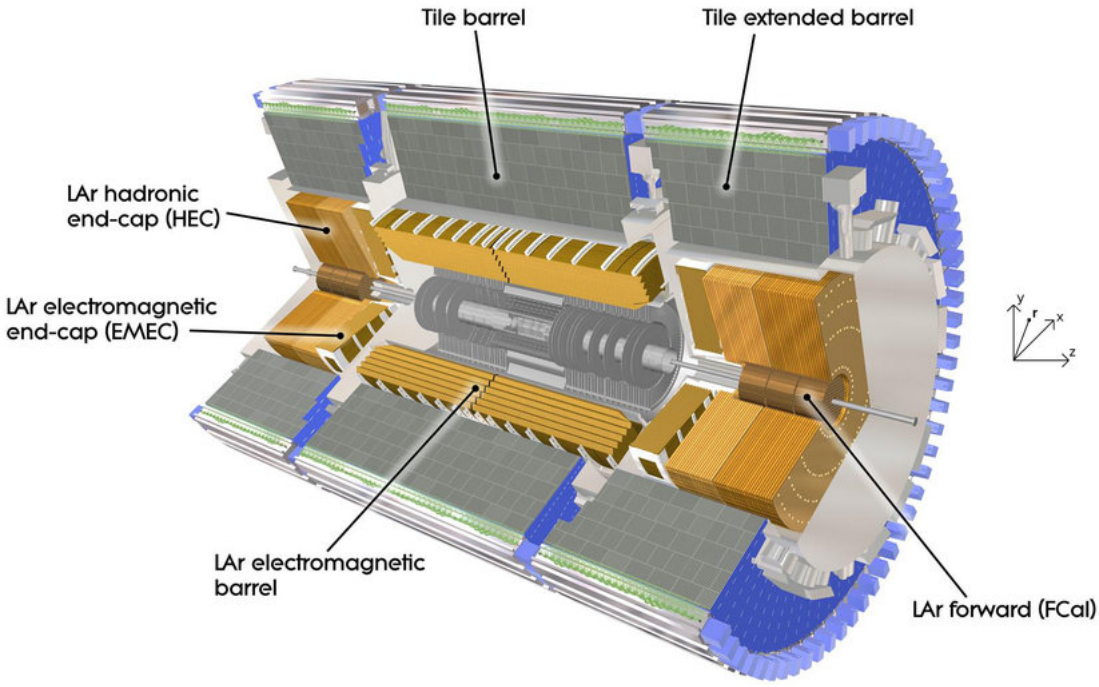


Figure 2.10: Structure of the ATLAS Calorimeters.

2.3.3 Calorimeters

The ATLAS calorimetric system, shown in Figure 2.10, is located after the ID and consists of three calorimeters: the Electromagnetic Calorimeter (EM) that covers the pseudorapidity region $|\eta| < 3.2$, the Hadronic Calorimeter (Had) that covers the region $|\eta| < 3.9$ and the Forward Calorimeter (FCal) that covers the region $3.1 < |\eta| < 4.9$. The main purpose of the calorimeters is to measure energy and direction of electrons, photons and hadrons and in addition to give a fundamental contribution ot the measurement of missing transverse energy.

A brief description of the three calorimeters is given in the following:

- The **Electromagnetic Calorimeter** (EM) [55] is 6.65 m long with an outer radius of 2.25 m. It is made up of a central barrel region which covers $|\eta| < 1.4$ and two end-caps which cover $1.4 < |\eta| < 3.2$. The EM is a sampling calorimeter which uses lead as aborber and Liquid Argon (LAr) as sensitive medium. The 2 mm layers of LAr are interleaved with copper electrodes, which collect the ionization charge generated by the electromagnetic shower, and with lead plates with a thickness of 1.5 mm (for $|\eta| < 0.8$ mm) or 1.13 mm (for $|\eta| > 0.8$ mm). Electrodes and ab-

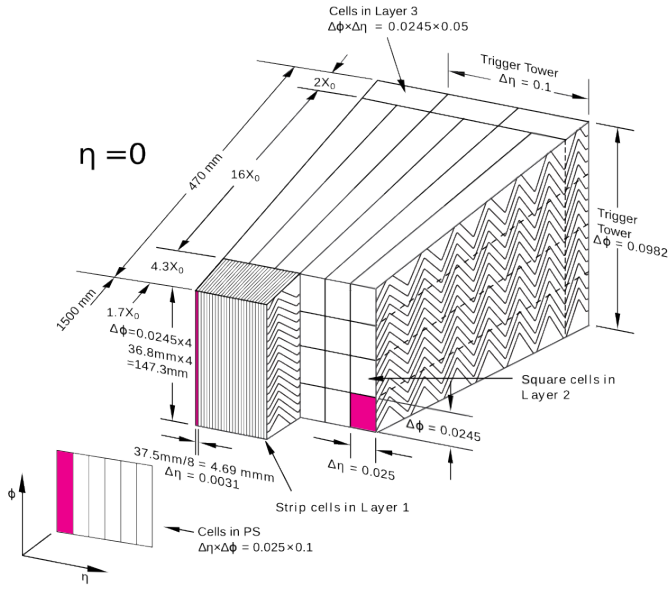


Figure 2.11: The ATLAS electromagnetic calorimeter accordion structure and segmentation.

sorbers are bent with an accordion geometry in order to provide a full coverage in ϕ without cracks due to the outgoing readout system. In order to contain completely the electromagnetic shower the EM calorimeter depth is about 25 radiation lenghts (X_0), approximately constant over η.
The EM calorimeter is composed of 3 longitudinal layers in the $0 < |\eta| < 2.5$ region and two in the higher η region, namely $2.5 < |\eta| < 3.2$. An additional layer (called presampler) is located in front of the EM calorimeter covering the region up to $|\eta| < 1.8$. It is useful for measuring the energy deposits of the electromagnetic showers which start before reaching the calorimeter. The segmentation of the EM calorimeter is shown if Figure 2.11.

The EM energy resolution is:

$$\frac{\sigma}{E} \approx \frac{10\% \div 17\%}{\sqrt{E}} \oplus 0.7\% \tag{2.9}$$

where E is expressed in GeV and the stochastic terms depens on η, the noise term is instead negligible.

- The **Hadronic Calorimeter** (Had) [56] is 6.10 meters long with an outer radius of 4.25 and covers the pseudorapidity region $|\eta| < 3.9$. In the barrel region ($|\eta| < 1.7$) a sampling calorimeter (called TileCal), with scintillating tiles and iron absorbers as active and passive medium respectively, is used. Figure 2.12 illustrates the components in a module of the tile calorimeter.

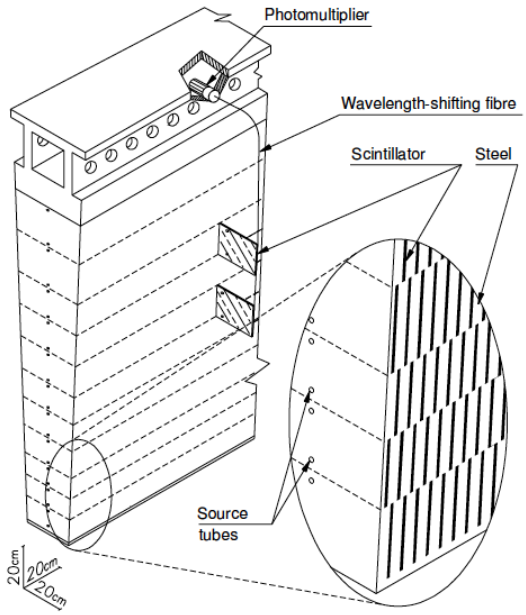


Figure 2.12: Illustration of the components in a module of the Tile Calorimeter.

In the end-cap region of the hadronic calorimeter (called HEC), Liquid Argon is used instead as active medium while copper and tungsten plates are used as absorber and it is placed in a separate cryostat together with the forward calorimeters.
The thickness of the hadronic calorimeter is 11 nuclear interaction length (λ) for $\eta = 0$. The energy resolution for hadronic jets, combined with electromagnetic calorimeter, is:

$$\frac{\sigma}{E} \approx \frac{50\%}{\sqrt{E}} \oplus 3\% \tag{2.10}$$

- The **Forward Calorimeter** (FCal) [57] is located at 4.7 m from the interaction point and covers the pseudorapidity region $3.1 < |\eta| < 4.9$. It provides measurements of both electromagnetic and hadronic showers in the forward region, therefore it is composed by three layers: an electromagnetic calorimeter and two hadronic calorimeters. LAr is used as active medium for all of them while copper is used as passive medium for the electromagnetic calorimeters and tungsten for the hadronic ones. The energy resolution is:

$$\frac{\sigma}{E} \approx \frac{100\%}{\sqrt{E}} \oplus 10\% \tag{2.11}$$

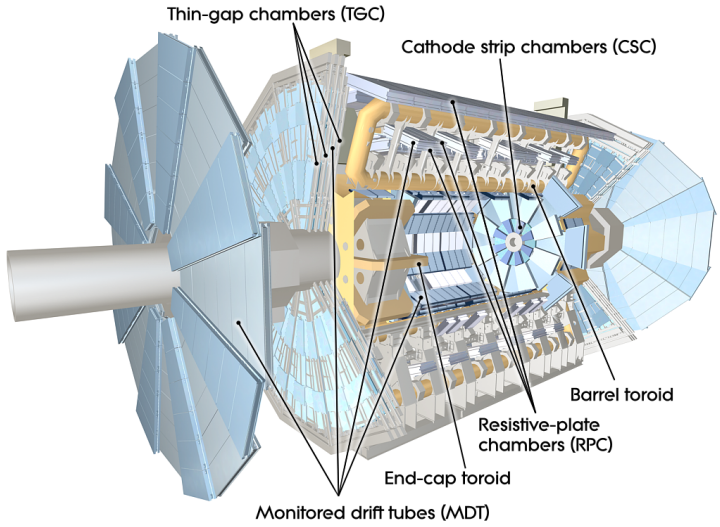


Figure 2.13: Layout of the ATLAS Muon Spectrometer.

2.3.4 The Muon Spectrometer

The Muon Spectrometer (MS) [58] is the outermost part of the ATLAS detector and surrounds the calorimeters. Generally all the particles produced in the collisions are absorbed by the calorimeters except muons (and neutrinos) which reach the MS. For this reason the purpose of the MS is to provide high precision measurement of the muon momentum, track reconstruction and trigger, using the magnetic field generated by toroid magnets. The MS is composed by a central barrel with three layers located at 5, 8 and 10 m respectively from the interaction point and four wheels for each end-cap at 7, 11, 15 and 22 m from the interaction point. It covers the pseudorapidity region $|\eta| < 2.7$. The structure of the MS is shown in Figure 2.13 and described in the following:

- Monitored Drift Tube (MDT) chambers are located in the barrel region and allow to precisely measure the muon momentum in the region $|\eta| < 2.7$. Each MDT chamber is made up by 3-8 layers of drift tubes with 30 mm radius and filled with a Ar/CO_2 mixture of gases. They have an average resolution of 35 μm per chamber.

- Cathode-strip chambers (CSC) are multi-wire proportional chambers, used for the muon momentum measurement in the forward region ($2 < |\eta| < 2.7$). CSCs are able to provide measurements of both coordinates and high timing resolution. The intrinsic resolution is of 60 μm in the bending plane and 4 mm in the transverse plane.

- Resistive Plate Chambers (RPC) in the barrel region ($|\eta| < 1.05$) and Thin Gap Chambers (TGC) in the end-caps ($1.05 < |\eta| < 2.4$) are used to provide fast muon trigger. The intrinsic time resolution is 1.5 ns for RPC and 4 ns for TGC.

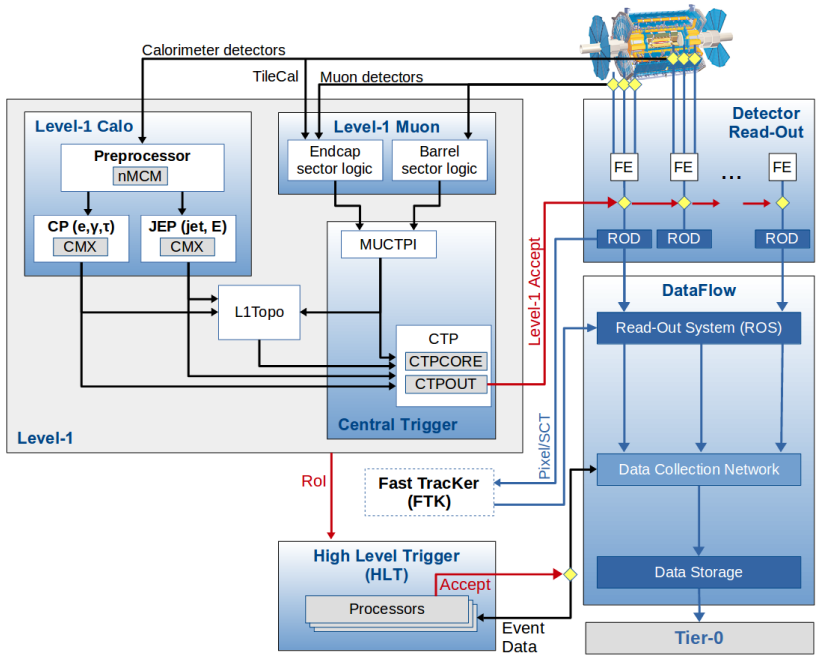


Figure 2.14: The ATLAS trigger and data acquisition system [61].

The overall transverse momentum resolution for the muon spectrometer is:

$$\frac{\sigma}{p_T} \approx \frac{0.29\text{GeV}}{p_T} \oplus 0.043 \oplus 4.1 \times 10^{-4}\text{GeV}^{-1} \times p_T \tag{2.12}$$

2.3.5 Trigger system

The LHC operates at very high rate: the proton bunches are spaced by 25 ns, which corresponds to 40 MHz frequency. The maximum rate that can be stored for offline data-analysis is about 1 kHz. For this reason ATLAS must have a very efficient trigger system [59, 60, 61] which should establish if the signal has the required characteristics in a very short time. In order to achieve such a reduction the trigger consists of two separate steps (Figure 2.14 shows a scheme of the trigger and data acquisition system):

- The **Level 1 Trigger** (L1) is an hardware-based trigger. The decision is formed by the Central Trigger Processor (CTP) using inputs coming from the calorimeters and the muon spectrometer. It searches for objects with high transverse momentum and also large missing energy and large total transverse energy. For each event, the L1 trigger selects some regions of physical interest called Regions Of Interest (ROI). The L1 decision is characterized by a latency of less than 2.5 μs and it reduces the event rate to about 100 KHz.

- The **High-Level Trigger** (HLT) is a software-based trigger. The events accepted by L1 are processed by the HLT, using all the information available within the ROIs,

exploiting the full detector granularity and trigger algorithms. The HLT reduces the output rate to about 1 KHz with a processing time of 4 s. The accepted events are then transferred as RAW data to local storage at the ATLAS site and then to the Tier-0 facility at CERN.

Chapter 3

Data and Monte Carlo samples

This Chapter will briefly describe how the data and Monte Carlo samples are produced in ATLAS. The specific samples used will be specified from time to time, this Chapter is just intended to give a general overview.

3.1 Data preparation

The data have to pass through some steps before being used by physics analyses. After the trigger selection mentioned in Chapter 2 the data are assigned to different streams that can be grouped in three main categories: *physics*, *express* and *calibration* [62]. Physics stream contains the dataset to be used for physics analyses, the express stream is a reduced set of the physics stream used mainly for data quality checks and the calibration stream contains data taken between colliding bunches used for calibration purposes. Data are then stored at the Tier-0 facility at CERN and pass through the Prompt Calibration Loop (PCL) [63], which is summarized in Figure 3.1.

The first step is the online monitoring from the ATLAS control room. A fraction of events from the express stream is reconstructed in order to produce monitoring histograms available within minutes and a dedicated DQ shifter checks regurarly these histograms, providing fast feedback and minimizing the data losses. The second step is done using the express stream and in particular the CosmicCalo stream (which contains events triggered by a cosmic ray signature in the calorimeter in empty bunches). These streams are sufficient to provide a broad coverage for potential problems. If some issues are found the data preparation team can be early warned and perform further investigations. The last step concerns the data calibration. Many calibrations are performed in order to maximise the detector and physics performance.
Data quality assesment is performed at several levels during these steps and relies on a dedicated data quality monitoring infrastucture. Shifters and experts look at the histograms and summarise the results in a dedicated database. Data passing all the data quality assessments enter in the Good Run List (GRL) that indicates the subset of data suitable for physics analyses and it is directly used by analysers to exclude data that do not pass the data quality criteria. Only the events that enter in the GRL are considered in the computation of integrated luminosity. Figure 3.2 reports the cumulative luminosity for the Run 2 delivered by the LHC, recorded by ATLAS and passing the quality requirements in order to be used by physics analyses.

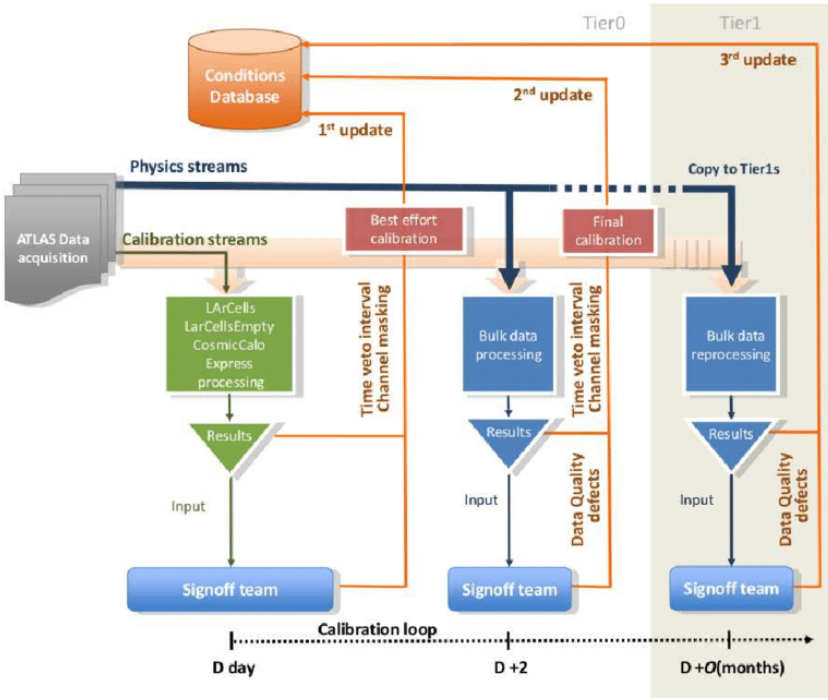


Figure 3.1: The workfklows of the ATLAS data preparation prompt calibration loop [63].

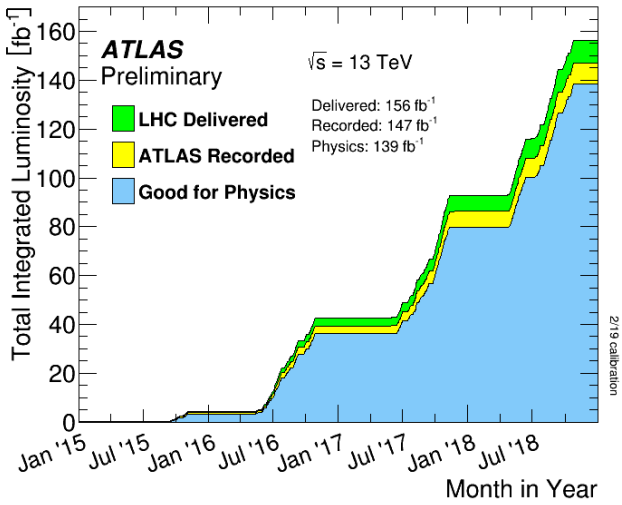


Figure 3.2: Cumulative luminosity as a function of time delivered by the LHC (green), recorded by ATLAS (yellow) and certified to be good quality data (blue) for the Run 2 data-taking (2015-2018) [44].

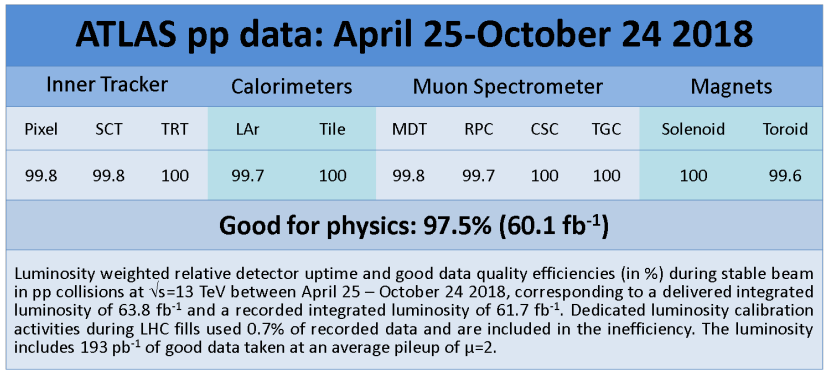


Figure 3.3: Fraction of integrated luminosity, passing all the data quality requirements, for each subystem for the period April 25 - October 24 2018. It corresponds to a integrated recorded luminosity of 61.7 fb^{-1} [64].

As an example Figure 3.3 also shows the data quality efficiencies in the various subsystems during 2018.

3.2 Data format and software framework

The whole process of processing and reconstruction of data mentioned in Section 3.1 is performed using a common and robust analysis framework called ATHENA [65]. This is used by all the ATLAS groups also for simulations and physics analysis. It consists of many packages written in C++ programming language and can be executed using Python scripts.
Different data formats are handled by this framework and used in ATLAS:

- **Raw data**: contains directly the output of the ATLAS detector after the HLT trigger. It comes in the bytestream format and the event size is about 1.6 MB.
- **Event summary data (ESD)**: contains the output of the reconstruction process. Both the detector information and reconstructed objects like electrons, muons and jets are stored in these files. The approximate size is some MB per event.
- **xAOD**: it is a subset of ESD with the physical objects of analysis interest and reduced detector information. The event size is around 300-500 KB.
- **DxAOD**: it contains a reduced set of informations derived from the xAOD. The information is restricted to only objects and relative informations useful for a specific analysis or performance group. The derivations are also run centrally using ATHENA.
- **Ntuples**: each analysis group produced then their own ntuples, starting from the derivation.

The huge amount of data recorded by ATLAS is used by data processing and analysis activities in all around the world. In order to deal with this the Worldwide LHC Computer Grid (WLCG) was developed. It is a global network of computer centers that provide the resources to store, distribute and process the data generated by the LHC. This is realized with 170 centers across 41 countries. These computers are arranged in *Tiers*: Tier-0 is the CERN computing centre where the RAW data are stored while the Tier-1 and the Tier-2 are located around the world, in particular the Tier-1 facilities are able to store copy of the RAW data and to provide the processing capability and Tier-2 are mainly used for performing calibration, simulation and analysis.

3.3 Monte Carlo generators and simulation

The proton-proton collisions happening at the LHC are very complex processes since the proton are not elementary particles. Comparing the theory predictions with the experimental data is essential and not straightforward. The link between theory and experiments are the Monte Carlo generators which simulate the particle physics processes. Having an accurate modelling of both expected signals and backgrounds is of paramount importance for the LHC physics.
The Monte Carlo (MC) simulation goes through different steps, that are shown in Figure 3.4 (from the hard scattering to the hadronization) and can be summarized as follows:

- **Hard scattering**: the first step is the calculation of the hard scattering matrix element between two interacting partons. As discussed in Chapter 2 the main feature of QCD is that the total cross section is given by the convolution of partonic cross section and the parton distribution functions (PDFs). Thus these two ingredients are needed for the matrix element calculation: the partonic cross section which is calculated analitically and the PDFs. The partonic cross section is calculated at fixed order prediction and thus the precision depends on the truncation of the perturbative series expansion. The current precisions available are: Leading Order (LO), Next-To-Leading order (NLO) and Next-Next-To-Leading Order (NNLO). The PDFs are instead not determined by perturbative QCD but rather by fits to experimental data. The theory is only able to predict their evolution. There are currently different PDF sets available obtain by using different approaches, namely different datasets and algorithms: NNPDF3.0 [67], MMHT2014 [19], CT14 [68].

- **Parton Shower**: partons are coloured objects, hence they can radiate gluons both before (*initial state radiation*) or after (*final state radiation*) the hard interaction. These gluons can, in turn, split further into quark-antiquark pairs and gluons and so on and this radiation might result in additional jets in the final state. This additional radiation is simulated by a parton shower algorithm. The parton shower treats emissions starting from the hard interaction down to the hadronization scale where the perturbative QCD breaks down ($\sim$ 1 GeV). It is implemented as a markovian process which step by step generates the particles with their kinematic variables. The effect of this extra radiation should be unitary, thus should not change the total cross section but it affects jet properties and multiplicities. Because of the

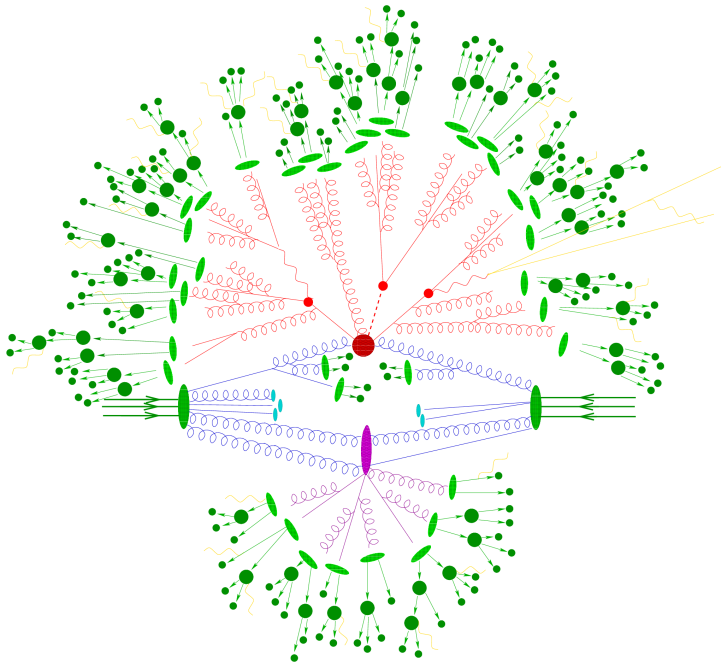


Figure 3.4: Sketch of a hadron-hadron collision as simulated by a Monte Carlo event generator. The red blob in the center represents the hard collision, surrounded by a tree-like structure representing Bremsstrahlung as simulated by parton showers. The purple blob indicates a secondary hard scattering event. Parton to hadron transitions are represented by light green blobs, dark green blobs indicate hadron decays, while yellow lines signal soft photon radiation [66].

latter effect the parton shower must be carefully matched with the matrix element calculation in order to do not double count the extra radiation.

- **Underlying event**: the partons of the interacting protons that do not participate in the hard scattering can interact as well, producing secondary interactions with low transfer momentum which define the *underlying event*. There are dedicated algorithms for estimating underlying events, usually embedded in the same software as the parton shower.

- **Pile-up**: the events coming from other interacting protons from the same bunch crossing have to be taken into account as well.

- **Hadronization**: the partons produced in the final state are then transformed into colorless hadrons that can also further decay. We need a non-perturbative algorithm in order to perform this step, called *hadronization*. The algorithms available are also embedded in the parton shower program.

- **Detector simulation**: after the complete event is generated, we need to simulate the interaction with the detector in order to compare the MC simulation to data. The behaviour of the detector is simulated with the Geant4 software [69].

- **Reconstruction**: the final step is the application of the physics objects reconstruction algorithms as it is done for data, in order for the MC simulations to be totally comparable with data.

All the steps described previously can be performed using different standalone MC generators. However the whole workflow (for some generators) is also embedded in the ATLAS ATHENA framework and the user can produce MC simulations using scripts called *JobOptions* that set the specific required features.

Chapter 4

Physics objects reconstruction in ATLAS

The particles produced in the p-p collisions interact with the material of the ATLAS detector, producing a signal that is acquired and stored. The information coming from the various subdetectors is then combined in order to build physics objects like electrons, photons, muons, taus, jets and E_T^{miss}. This process, which allows to identify the particles and measure their physics properties, is called *object reconstruction*. How the reconstruction of all the physics objects in ATLAS is performed, is described in this Chapter.

4.1 Tracks reconstruction in the Inner Detector

In this section the track reconstruction in the ID will be illustrated as this is a common procedure to all the physics objects reconstruction.
Tracks are reconstructed using various algorithms [70]. The main one is the inside-out algorithm which searches for 3-hits seeds in the silicon detectors. The seeds are the starting point for a combinatorial Kalman filter algorithm [71] which build tracks by adding hits to the seed and moving away from the interaction point. Overlapping between track candidates is resolved and the tracks are extended to the TRT.
Primary vertices are reconstructed using an iterative vertex finding algorithm. An iterative χ^2 fit is performed using vertex seeds and tracks closer to the seed. Tracks displaced more than 7σ from the vertex seed are not considered as belonging to this vertex but used to seed a new vertex. This procedure is repeated until no new vertex is found.

4.2 Electrons and photons

Electrons and photons interact with the EM calorimeter giving rise to energy deposits. In addition, electrons are characterized by tracks in the inner detector while photons could have an associated track (more specifically speaking, a secondary vertex with one or two tracks) if they are converted photons, namely photons originated by an electron.
The reconstruction procedure is followed by an identification process where the candidates are selected against background using additional information from tracks, track-to-cluster matching and shape of the electromagnetic shower. Furthermore, electrons and photons can be required to be isolated from other activity in the calorimeter.
The reconstruction, isolation and identification [72, 73, 74] will be described in the following.

4.2.1 Electron reconstruction

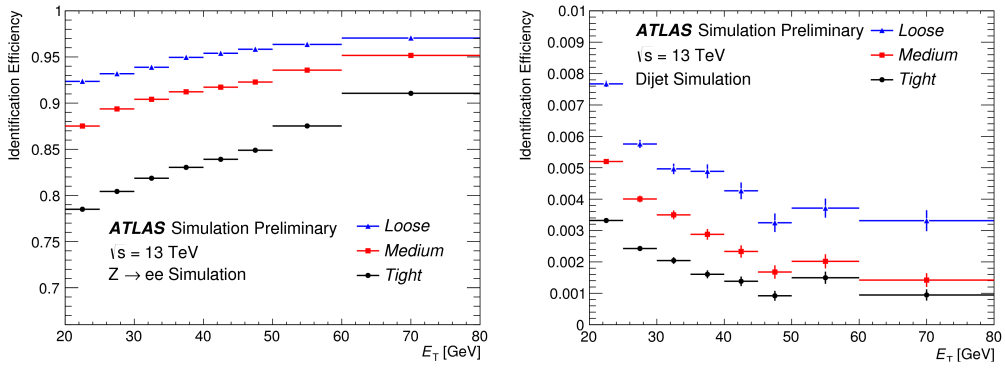
The electron reconstruction consists of several steps. First a scan is performed using a sliding window with a size corresponding to the granularity of the EM calorimeter middle layer. *Seed clusters* with energy above 2.5 GeV are identified and then the clusters are formed around the seeds using a cluster algorithm. The efficiency of the cluster reconstruction depends on the energy. It is 95% at $E_T = 7$ GeV and more than 99% above $E_T = 15$ GeV. The track reconstruction in the silicon detectors is then performed and they are matched to the EM clusters using the distance in η and ϕ between the extrapolation of the track in the calorimeter and the cluster barycenter. Tracks associated to electron clusters are refitted using the Gaussian Sum Filter algorithm [74]. The matching procedure is then repeated for the refitted tracks and electron clusters are re-formed. The energy of electron clusters is corrected to the original electron energy using multivariate techniques based on Monte Carlo samples.
The four momentum of the electrons is at the end computed using informations from both the energy clusters and the tracks where the energy is given by the deposit in the calorimeter and the η and ϕ directions are taken from the track. In addition the electron associated tracks are also required to be compatible with the primary vertex of the hard collision in order to reduce the background from conversions or secondary particles.

4.2.2 Electron identification

In order to reject background objects like hadronic jets or converted photons, algorithms for electron identification are applied. The identification algorithms use different quantities in order to discriminate between a signal-like object or a background-like object. These variables are related to the electron cluster and track measurements including shower shapes information, information from the transition radiation tracker, track-cluster matching related quantities and variables measuring bremsstrahlung effects. The identification algorithm used during Run 2 is based on a likelihood method. It is a multivariate technique where several properties of the electron candidates are evaluated at the same time. The probability for an object to be signal-like or background-like is then calculated. The optimization of the algorithm is performed using events from Monte Carlo simulations of $Z \to ee$, dijet and $J/\psi \to ee$ and three identification operating points can be determined: *Loose*, *Medium* and *Tight*. These operating points are defined in order of increasing background rejection and such as that the samples selected by them are subsets of one another. The efficiency of the different operating points as a function of E_T is shown in Figure 4.1.

4.2.3 Electron isolation

Prompt electrons from electroweak processes are usually not accompanied by nearby particles. Therefore, in addition to the identification, an electron could be required to fulfill isolation requirements, in order to further reject the background like electrons originating from converted photons produced in hadron decays, electrons from heavy flavor hadron decays and light hadrons misidentified as electrons. The isolation variables quantify the



(a) (b)

Figure 4.1: Efficency to identify electrons from $Z \to ee$ events (a) and the efficiency to identify hadrons as electrons, namely background rejection, using dijet events (b). The identification efficiencies are shown for the three operating points: loose, medium and tight [74].

energy of the particles around the electron candidate, in particular two variables are defined:

- Calorimetric isolation energy $E_T^{cone0.2}$: defined as the sum of transverse energies of topological clusters within a cone of $\Delta R = 0.2$ around the candidate electron cluster.
- Track isolation $p_T^{varcone0.2}$: defined as the sum of transverse momenta of all tracks within a cone of $\Delta R = \min(0.2, 10 \text{ GeV}/E_T)$ around the candidate electron track and originating from the primary vertex of the hard collision.

Different selection requirements based on the quantities $E_T^{cone0.2}/E_T$ and $p_T^{varcone0.2}/E_T$ are defined in order to select isolated candidates. The resulting operating points are divided in two classes: efficiency targeted operating points where varying requirements are used in order to have a given isolation efficiency and fixed requirement operating points where the upper tresholds on the isolation variables are constant. Fixed requirement operating points are tipically used in physics analysis with low E_T electrons and requiring high background rejection whereas efficiency targeted operating points are prefered in analysis with large E_T electrons in order to maintain high signal efficiency.

4.2.4 Photon reconstruction

The reconstruction of photons follows several steps and proceed in parallel with the reconstruction of electrons, since they have very similar signature in the EM calorimeter. First seed clusters are identified using a sliding window algorithm. The procedure is the same as that of the electrons and the efficiency is greater than 99% for photons with $E_T > 20$ GeV. Once the clusters are reconstructed, inner detector tracks loosely matched

with the clusters are identified, using the distance in η and ϕ between the extrapolated track in the calorimeter and the cluster barycenter. They are then refitted using the Gaussian Sum Filter algorithm and retained for the reconstruction of electrons and converted photons.
Double track conversion vertex candidates are reconstructed from pairs of oppositely charged tracks in the inner detector. For each track a likelihood-based algorithm determine whether the candidate is likely to be an electron. A fit is then performed to find the conversion vertex. Conversion vertices are matched to the clusters as in the loose track matching, therefore the coordinates of the extrapolated track is compared to the coordinates of the cluster barycenter.
Finally the candidates are classified as electron, unconverted photon or converted photon in the following way:

- Clusters with neither an associated track nor a converted vertex are considered unconverted photons.
- EM clusters matched to a conversion vertex are classified as converted photon candidates. If this candidate has also been reconstructed as an electron, the electron track is evaluated against the track originating from conversion vertex and in case additional information like hits in the pixel detector is used to determine the nature of the candidate.
- Converted photon candidates are recovered from reconstructed electron candidates if the electron candidates has no hits in the silicon detectors and low transverse momentum, $p_T < 2$ GeV.

4.2.5 Photon identification

In order to discriminate between prompt photons and photons from hadron decays, photon identification algorithms are applied. These algorithms are based on a set of discriminating quantities, related in particular to the shower shape in the EM calorimeter and the lower leakage fraction in the hadronic calorimeter. Photons generally produce narrower energy deposits in the EM calorimeter than background photons from jets and have smaller leakage fraction in the hadronic calorimeter because of the absence of additional hadrons.
A multivariate technique is employed in order to define two operating points: *Loose* and *Tight*. The loose selection is based only on the shower shapes in the second layer of the EM calorimeter and on the energy deposit in the hadronic calorimeter while the tight working point adds information from the strip layer of the calorimeter. The efficiency ranges from 97% at $E_T = 20$ GeV to above 99% at $E_T > 40$ GeV for the loose selection for both converted and unconverted photons while it is 85% at $E_T = 40$ GeV and above 90% at $E_T \sim 1$ TeV for the tight selection for both converted and unconverted photons.

4.2.6 Photon isolation

Photons are tipically required to pass isolation criteria. The requirement is to have a low E_T^{iso} ($E_T^{iso} < 4$ GeV) where this quantity is calculated considering three dimensional

topological clusters of calorimeter cells reconstructed in a cone of size $\Delta R = 0.4$ around the photon candidate. The photon transverse energy and the energy coming from pile-up and underlying event are subtracted from the calculation of E_T^{iso} and a correction factor is applied in order to take into account photon energy outside the cluster.

4.2.7 Electron and photon energy calibration

It is fundamental for many physics measurements to precisely know the electron and photon energy. The energy calibration is based on both data and Monte Carlo simulations and can be summarised as follows:

- *Non-uniformity detector response*: data are corrected for time-dependent defects in the detector. These corrections concern high-voltage regions of the EM calorimeter and different gains of the calorimeter cells readout.
- *Monte Carlo based calibration*: the optimization of the energy calibration is performed using multivariate techniques on single particle Monte Carlo simulations. The technique is based on a Boosted Decision Tree (BDT) with gradient boosting. The EM clusters properties are calibrated to the original electron and photon energy in the Monte Carlo samples. The calibration coefficients are calculated separately for electrons, converted and unconverted photons.
- *In-situ corrections*: after the two corrections described above a disagreement between data and Monte Carlo can still be observed. Simulations of $Z \to ee$ events are used for comparison between data and simulations. Energy scale correction factors are then applied to data and correction factors of the resolution are applied to simulations.

4.3 Muons

Muons are able to reach the Muon Spectrometer, which is the outermost part of the ATLAS detector, therefore they are characterized by tracks both in the MS and in the ID, and their reconstruction proceed separately in these two subdetectors. The information from individual subdetectors is then combined to form the muon tracks. The reconstruction in the ID is performed as described in Section 4.1 while in this Section the focus will be on the reconstruction in the MS and the combined muon reconstruction [75].

4.3.1 Reconstruction

Muons reconstruction in the MS proceed with the searching for hit patterns inside each muon chamber in order to form segments. In the MDT chambers a search for hits aligned on a trajectory in the bending plane of the detector is performed and the segments are reconstructed by using a fit to the hits found in each layer. The RPC or TGC hits measure the coordinate orthogonal to the bending plane and segments in the CSC are reconstructed using a separate search in the η and ϕ coordinate. Muon track are then built by fitting together hits from segments in different layers.

The muon tracks built in the MS are then combined with the ID tracks and this combined reconstruction is performed using various algorithms according to the information provided by the ID, MS and also the calorimeters. Four muon types can be then defined as follows:

- *Combined (CB) muons*: tracks are built independently in the ID and in the MS and then a combined track is constructed performing a fit that uses hits from both the ID and the MS. Most muons are first reconstructed in the MS and then extrapolated inward and matched to an ID track.
- *Segment-tagged (ST) muons*: a reconstructed track in the ID is extrapolated to the MS and then associated with at least one local track segment in the MDT or CSC chambers. This kind of muons are only defined when they cross only one layer of the MS chambers.
- *Calorimeter-tagged (CT) muons*: a track in the ID can be identified as a muon if it is matched to an energy deposit in the calorimeter compatible with a minimum ionizing particle. This type has the lowest purity but it helps to recover acceptance in the region where the MS is only partially instrumented to allow for cabling and services to the calorimeters and the ID.
- *Extrapolated (ME) muons*: the muon trajectory is reconstruced based only on the MS track and a loose requirement on compatibility with originating from the interaction point. This allows to reconstruct muons in an acceptance region not covered by the ID ($2.5 < |\eta| < 2.7$).

The overlap between different muon types is resolved before producing the muon collection used by physics analysis. If two muons have the same ID track, preference is given to CB muons, then to ST and finally to CT muons. The overlap with ME muons is instead resolved selecting the track with better fit quality and larger number of hits.

4.3.2 Identification

In order to select muons against background, which is mainly made up by muons from pion and kaon decays, further quality requirements are applied. Background muons originating from in-flight decays of charghed hadrons are usually characterzed by a "kink" topology in their reconstructed track, therefore the fit quality of the resulted track is poor and the momentum measurements in the ID and in the MS may not be compatible. In order to discriminate between prompt muons and background muons, the following variables are used:

- q/p significance: it is the absolute value of the difference between the ratio of the charge and momentum of the muons measured in the ID and in the MS divided by the sum in quadrature of the corresponding uncertainties.
- ρ': it is the absolute value of the difference between momentum measurements in the ID and in the MS divided by the p_{T} of the combined track.

	$4 < p_T < 20$ GeV		$20 < p_T < 100$ GeV	
Selection	ϵ (prompt-muons) [%]	ϵ (misidentification) [%]	ϵ (prompt-muons) [%]	ϵ (misidentification) [%]
Loose	96.7	0.53	98.1	0.76
Medium	95.5	0.38	96.1	0.17
Tight	89.9	0.19	91.8	0.11
High-p_{T}	78.1	0.26	80.4	0.13

Table 4.1: Efficiency for prompt muons and misidentification rate for hadrons for the four different identification working points in the low p_{T} range ($4 < p_T < 20$ GeV) and high p_{T} range ($20 < p_T < 100$ GeV). The results are obtained using $t\bar{t}$ Monte Carlo samples [75].

- χ^2 of the combined track fit.

Furthermore specific requirements on the number of hits in the ID and in the MS are used.
Four identification selections are then provided as follows (where the *Loose*, *Medium* and *Tight* working points are defined in an inclusive way):

- *Loose*: this selection maximises the reconstruction efficiency and it is optimized for reconstructing Higgs boson candidates in the four lepton final state. All the muon types are used in this case.
- *Medium*: this is the default selection for muons in ATLAS and minimizes the systematic uncertainties. Only CB and ME muons tracks, which satisfies certain hits requirements and have the $q/p\ significance$ less than seven, are used.
- *Tight*: this selection maximises the purity to the detriment of efficiency. Only CB muons satisfying the Medium selection criteria are used. Additional requirements on hits, χ^2 of the combined track fit, on ρ' and $q/p\ significance$ are applied.
- *High-p_T muons*: this selection maximises the momentum resolution for tracks with transverse momentum above 100 GeV. This is optimized for searches for high-mass Z' and W' resonances.

The reconstruction efficiencies for the selections mentioned above are computed using $t\bar{t}$ simulation and reported in Table 4.1.

4.3.3 Isolation

In addition to the identification criteria a muon can be required to be isolated. This is mostly the case for muons produced from the decay of heavy particles such as W, Z and Higgs boson whereas muons from semileptonic decays are usually embedded in jets.
Two variables are used in order to quantify the detector activity around the muon candidate and asses muon isolation:

- *Track isolation*, $p_T^{varcone30}$: defined as the scalar sum of the transverse momenta of the tracks with $p_T > 1$ GeV in a cone of size $\Delta R = \min(10\ \mathrm{GeV}/p_T^{\mu}, 0.3)$ around the muon candidate, excluding the muon track itself.

- *Calorimeter isolation*, $E_T^{topocone20}$: defined as the sum of the transverse energy of topological clusters in a cone of size $\Delta R = 0.2$ around the muon candidate, excluding the muon itself.

Isolation criteria are then defined using the quantities $p_T^{varcone30}/p_T$ and $E_T^{topocone20}/E_T$ and different operating points are provided.

4.3.4 Calibration

As for electrons and photons, it is important to have a precise knowledge of muon momentum, therefore the muon momentum scale and resolution are studied using $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ events. Scale factors are computed in order to correct the muon transverse momenta reconstructed both in the ID and in the MS. They are derived with a method which compares Monte Carlo simulations of $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ events and data in order to reconstuct J/ψ and Z mass peaks.

4.4 Jets

Due to QCD properties, partons coming from p-p collisions cannot be observed in nature but only colorless particles such as hadrons can. As described in Chapter 1, partons undergo through the processes of fragmentation and hadronization, hence what is possible to observe in the final state are bunches of collimated hadrons, called jets.
Jets reconstruction is crucial for many physics analysis since jets are a distinctive signature of hard interactions between partons, occuring in many processes. Jets reconstruction and calibration [76, 77] will be described in the following sections.

4.4.1 Reconstruction

Jets are reconstructed using the anti-k_T algorithm [78], which runs over topological energy clusters found inside the calorimeters. For each pair of clusters i and j, the following quantities are calculated:

$$\begin{aligned} d_{i,j} &= \min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2}\right)\frac{\Delta R_{i,j}^2}{R^2} \\ d_{i,B} &= \frac{1}{p_T^2 i} \end{aligned} \tag{4.1}$$

The $d_{i,j}$ variable represents the distance between object i and object j, while $d_{i,B}$ is the distance between the object and the beam. The variable R is the radius parameter, which in case of the ATLAS reconstruction, is usually sets to 0.4. The clustering proceeds identifying the smallest distance between $d_{i,j}$ and $d_{i,B}$ and if it turns out to be $d_{i,j}$, the object i and j are merged into the same cluster, while if it is $d_{i,B}$ the cluster is declared to be a jet and removed from the objects over which the algorithm is running. The procedure is then iterated until there is no object left.

4.4.2 Calibration

After the reconstruction, a calibration procedure is performed in order to restore the jet energy scale to that of *truth jets* reconstructed at the particle-level energy scale. These jets, used as reference, are built in simulation using anti-k_T algorithm with $R = 0.4$ using stable final state particles. In addition they are required to have $p_T > 7$ GeV and $|\eta| < 4.5$.
The calibration proceeds through different steps and corrects the four-momentum by scaling the jet p_{T}, energy and mass. The various stages are summarised as follows:

- **Origin corection**: the momentum is recalculated in order to match with the position of the primary vertex rather than the center of the detector, while keeping the jet energy constant.
- **Pile-up correction**: the excess energy due to pile-up is removed. This is achieved by using an area-based p_{T} density subtraction algorithm and then a residual correction is derived using Monte Carlo simulations.
- **Absolute Monte Carlo based calibration**: the reconstructed jet four-momentum is corrected using the absolute jet energy scale and η calibration, derived from dijets Monte Carlo simulations using reconstructed truth jets after the application of the origin and pile-up corrections.
- **Global sequential calibration**: the calorimeter response and the jet reconstruction are sensitive to fluctuations in the jet particle composition and the distribution of energy within the jet. For instance, the average particle composition and the shower shape of a jet varies between a quark or a gluon initiated jet. Residual corrections are derived using information from calorimeters and also from ID and MS.
- ***Insitu* calibration**: data and Monte Carlo simulations can still present residual differences due to inaccuracy in the different steps of Monte Carlo simulations, namely simulation of the hard scatter, parton shower and underlying event, pile-up, interaction with the detector as well as detector response and detector material. Differences are quantified using as reference other well-measured object and balancing the jet p_{T} consequently. The corrections derived are then applied to data.

4.4.3 Jet vertex fraction and jet vertex tagging

In order to further reduce backgrounds jets arising from pile-up events, techniques based on the association of a vertex to a jet, are employed [79]. The jet vertex fraction (JVT) is the variable used to identify the primary vertex (PV) from which the jet originated. JVF is defined as the ratio of the scalar sum of p_{T} of tracks that originated from a given primary vertex to the scalar sum of p_{T} of all tracks in the jet, independently of their origin:

$$\mathrm{JVF}(\mathrm{jet}_i, \mathrm{PV}_j) = \frac{\sum_m p_T(\mathrm{track}_m^{\mathrm{jet}_i}, \mathrm{PV}_j)}{\sum_n \sum_l p_T(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_n)} \tag{4.2}$$

where m runs over all tracks originating from PV_j matched to jet_i, n over all primary vertices in the event and l over all tracks originating from PV_n matched to jet_j. Applying a lower limit on JVF helps to remove jets which are not associated to the hard-scatter primary vertex, therefore to suppress backgrounds from pile-up jets. However this selection depends strongly on the number of primary vertices in the event. To avoid this issue a multivariate combination of the JVF variable and the number of primary vertices (N_{PV}) is used. The resulting discriminating variable, called Jet Vertex Tagger (JVT), is employed instead of the JVF and a lower limit on that is imposed for pile-up jets suppression.

4.4.4 B-tagging

Jets originated by a b quark can be identified due to the relatively long lifetime of hadrons containing this particle, which is $\sim 1.3 \times 10^{-12}$ s ($c\tau \sim 500$ μm). This leads to a significant mean flight path length with at least one displaced secondary vertex. B-tagging algorithms exploit such features in order to discriminate b-jets from c-jets and light flavor jets.
The algorithm used in ATLAS is called MV2c10 [80] and it is based on a BDT which uses the following input variables: informations on the transverse and longitudinal impact parameter, the presence of a secondary vertex and its properties, jets momentum and pseudorapidity and informations on possible heavy hadrons decay chains within the jet. The BDT is performed using simulated $t\bar{t}$ events and different working points are then defined.

4.5 Tau

Tau lepton has a mass of 1.777 GeV and has a very short mean lifetime, 2.9×10^{-13} s, corresponding to a decay length of 87 μm, thus it is not visible in the detector but accessible only through the measurement of its decay products. The τ can either decay leptonically ($\tau \to l\nu_l\nu_\tau, l = e, \mu$) with a branching ratio of $\sim 35\%$ or hadronically ($\tau \to$ hadrons ν_τ, labelled as τ_{had}) with a branching ratio of $\sim 65\%$. Hadronic tau decay products can contain one charged hadron or three charged hadron, therefore they can be further split into two categories called 1-prong and 3-prong. The hadrons constitute the only visible part of tau decay products, for this reason they will be referred to as $\tau_{had-vis}$.
Leptonic tau decays will not be discussed in this section since they are reconstructed as electrons and muons, already described in Section 4.2 and Section 4.3 respectively. The reconstruction and identification procedures of hadronically decaying τ [81, 82] will instead be described in the following as this is also relevant for this thesis work, that will be illustrated in details in Chapter 5 and Chapter 6.

4.5.1 Reconstruction

Tau candidates are reconstructed starting from jets built using the procedure described in Section 4.4, namely using an anti-k_t algorithm with radius parameter $R = 0.4$. In addition they are required to have $p_T > 10$ GeV and be within the ID acceptance, i.e.

$|\eta| < 2.5$. Tau candidates in the transition region between the barrel and the end-cap are vetoed.
A primary vertex is then associated to the tau candidate through the tau vertex (TV) algorithm. The TV algorithm uses as input all tau candidates tracks in the region $\Delta R < 0.2$ around the jet seed. The p_{T} of all the tracks is then summed and the primary vertex with the largest fraction of p_{T} sum is chosen as TV. Tracks are associated to two different region according to their distance from the tau candidate: a *core* region ($0 < \Delta R < 0.2$) and an *isolation* region ($0.2 < \Delta R < 0.4$). The direction of the tau candidate is calculated using the vectorial sum of the clusters within $\Delta R < 0.2$ of the jet barycenter and using the TV as origin.
The geometrical overlap with other objects within $\Delta R < 0.2$ is also resolved by selecting only one object with the following priorities: muons, electrons, $\tau_{had-vis}$ and jets.

4.5.2 Identification

The main background to hadronic tau lepton decays arises from jets quark or gluon initiated. In order to reject this background and correctly identifying hadronic taus, further requirements are applied. In particular the identification is based on multivariate techniques which exploit BDT algorithms. Two separate BDTs are trained for tau candidates associated with one track or with three tracks using simulated $Z/\gamma^* \rightarrow \tau\tau$ events for signal and dijets events for background. Three working points, called *Loose*, *Medium* and *Tight*, are identified, targeting efficiencies of 0.6, 0.55 and 0.45 respectively for the 1-track case and 0.5, 0.4 and 0.3 respectively for the 3-track case. The efficiencies are designed to be independent of p_{T}. Discriminating variables, based on the narrow shower in the calorimeter, the number of tracks and the displaced tau vertex, are used to distinguish the $\tau_{had-vis}$ candidates from jets. The complete list of BDT input variables together with their meaning is the following:

- Central energy fraction, f_{cent}: fraction of the calorimeter transverse energy deposited in the region $\Delta R < 0.1$ with respect to all energy deposited in the region, $\Delta R < 02$ around the $\tau_{had-vis}$ candidate.
- Leading track momentum fraction $f^{-1}_{leadtrack}$: sum of transverse energy of all clusters in the core region of the $\tau_{had-vis}$ candidate divided by the transverse momentum of the highest-p_{T} charged partice in the core region.
- Track radius, $R^{0.2}_{track}$: ΔR distance of the associated tracks to the $\tau_{had-vis}$ candidate direction, using only tracks in the core region.
- Leading track IP significance, $|S_{leadtrack}|$: absolute value of transverse impact parameter of the highest-p_{T} track in the core region divided by its uncertainty.
- Fraction of tracks p_{T} in the isolation region: scalar sum of the p_{T} of tracks associated with the $\tau_{had-vis}$ candidate in the isolation region divided by the sum of all tracks associated with the $\tau_{had-vis}$ candidate.
- ΔR_{max}: maximum ΔR between a track associated with the $\tau_{had-vis}$ candidate and the $\tau_{had-vis}$ candidate itself.

- Transverse flight path significance, S_T^{flight}: decay length of the secondary vertex in the transverse plane divided by its uncertainty.

- m_{track}: invariant mass calculated from the four-momentum of all tracks in the core and isolation regions, assuming a pion mass for each track.

- Fraction of EM energy from charged pions, $f_{EM}^{track-HAD}$: fraction of the electromagnetic energy of tracks associated with the $\tau_{had-vis}$ candidate in the core region.

Figure 4.2 shows the performance of the tau identification algorithm for both 1-prong and 3-prong decays. The efficiency for misidentifying QCD jets is displayed versus the identification efficiency for $\tau_{had-vis}$.

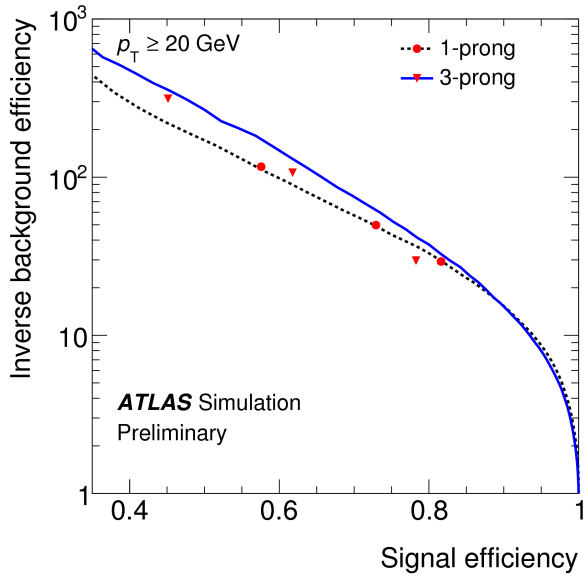
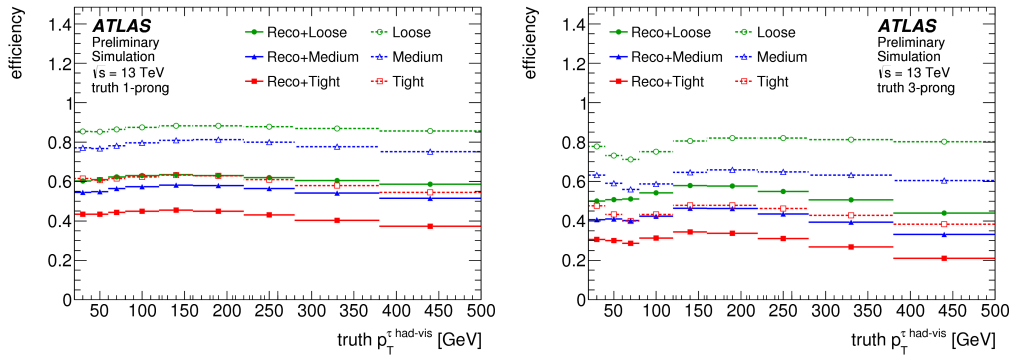


Figure 4.2: Inverse background efficiency is shown as a function of signal efficiency. The two lines refer to 1-prong and 3-prong decays and for each line the loose, medium and tight working points are shown. The working points do not lie exactly on the line because they are optimized to have a reduced p_T dependency on the efficiency [81].

Furthermore the working points are optimized to be p_T independent. Figure 4.3 shows the residual p_T dependecies for the signal efficiency both for the 1-prong and 3-prong case.

4.5.3 Energy calibration

In order to account for differences between the tau energy measured in the detector and the average value of the energy carried by the tau decay products obtained with simulations, a calibration of the tau energy is applied. In Run 2 two calibrations were available: the baseline calibration and a calibration based on a Boosted Regression Tree (BRT) method.



(a) (b)

Figure 4.3: Efficiency for $\tau_{had-vis}$ identification (open symbols) and combined reconstruction and identification efficiency (full symbols) as a function of the p_{T} for the 1-prong (a) and 3-prong (b) $\tau_{had-vis}$ candidates [81].

The baseline correction is calculated as:

$$E_{\mathrm{calib}} = \frac{E_{LC} - E_{\mathrm{pileup}}}{\mathcal{R}(E_{LC} - E_{\mathrm{pileup}}, |\eta|, n_p)} \tag{4.3}$$

The E_{LC} is the first calibration applied, where the tau energy is calibrated to the local hadronic calorimeter (LC) scale which mostly corrects for the calorimeter non-compensation and for the energy deposited in dead material. E_{pileup} refers to the pile-up offset correction, which is subtracted. Finally at the denominator there is the detector response calibration $\mathcal{R}$, that is computed separately for 1-prong and 3-prong decays (n_p) and for various $|\eta|$ bins.
The resulted energy resolution is excellent at high p_{T} but it degrades at low p_{T}, therefore a new method, called Tau Particle Flow (TPF), has been developed by ATLAS. This new technique is able to improve the tau energy resolution at low p_{T} thanks to the superior measurement of the charged pion momentum from the tracking system. Furthermore this method is able to reconstruct the individual charged and neutral hadrons in the tau decay. A new energy calibration has been therefore introduced which combines the information from the baseline and the TPF calibrations together with additional calorimeter and tracking informations, using a multivariate analysis based on a BRT method. The BRT is trained using simulated $Z/\gamma^* \to \tau\tau$ events with the following list of input variables:

- n_{PV}: number of primary vertices in the event.
- μ: average interaction per bunch crossing.
- Cluster shower depth, λ_{centre}: distance of the cluster shower centre from the calorimeter front face measured along the shower axis.
- Cluster second moment in λ, $\langle\lambda^2\rangle$: second moment of the distance of a cell from the shower centre along the shower axis.

- Cluster first moment in energy density, $\langle\rho\rangle$: cluster first moment in energy density $\rho = E/V$ where E and V are the energy and the volume of the cluster respectively.
- Cluster presampler fraction, $f_{presampler}$: fraction of cluster energy deposited in the barrel and endcap presamplers.
- Cluster EM-like probability, P_{EM}: probability of the cluster to be electromagnetic-like.
- n_{track}: number of tracks associated with the $\tau_{had-vis}$ candidate.
- n_{π^0}: number of reconstructed π^0 associated with the $\tau_{had-vis}$ candidate.
- Relative difference of pion energies, γ_π: relative difference of the total charged pion energy and the total neutral pion energy.
- η_{calo}: calorimeter-based pseudorapidity.
- p_T^{interp}: transverse momentum interpolated from calorimetric corrections to energy measurement and TPF reconstruction.
- p_T^{LC}/p_T^{interp}: ratio of the local hadron calibration transverse momentum to p_T^{interp}.
- p_T^{TPF}/p_T^{interp}: ratio of the TPF reconstruction transverse momentum to p_T^{interp}.

The use of the BRT improves significantly the $\tau_{had-vis}$ energy resolution, in particular for low p_{T}, as shown in Figure 4.4.

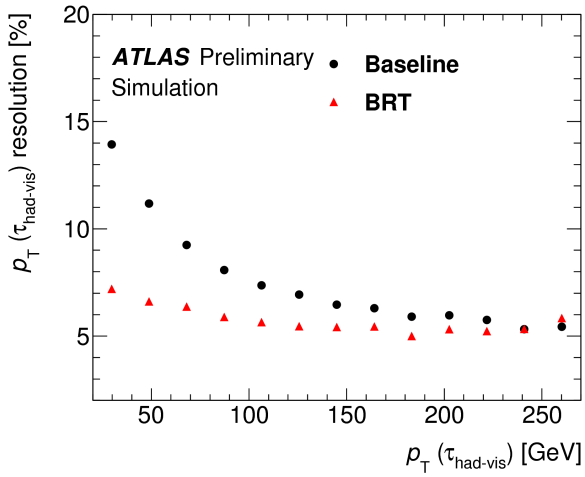


Figure 4.4: The resolution of the baseline and the BRT $\tau_{had-vis}$ energy calibration as a function of p_{T}. The resolution is defined as half of the 68% central intervals of the ratio of the calibrated p_{T} to the true p_{T} [82].

4.6 Missing transverse momentum

In a hadron collider the conservation of momentum in the plane transverse to the beam axis implies that the sum of transverse momenta of the final state particles is zero. An imbalance may indicate the presence of neutral, weakly interacting and stable particles, which are not detectable. The only particle with those features in the SM is the neutrino. Indeed missing transverse momentum is crucial for analysis involving neutrinos in the final state but also for many searches of Beyond the Standard Model physics, where exotic particles with the same behaviour are predicted. In addition, fake E_T^{miss} can arise from interacting SM particles that escape the acceptance of the detector or fail to be reconstructed. Reconstruction of missing transverse energy [83] is described in the following.

4.6.1 Reconstruction

The E_T^{miss} vector is calculated using the components along the x and y axes:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,\mu} + E_{x(y)}^{miss,soft} \tag{4.4}$$

where each term is the negative vectorial sum of energy deposits and/or tracks. The last term in Eq. 4.4 is the *soft term* where all the signals from the hard interaction not associated to a physics object are taken into account, whereas all the other terms are associated to fully reconstructed and calibrated physics objects and are collectively referred to as *hard term*.
The various terms are obtained as follows:

- $E_{x(y)}^{miss,\gamma}$: photons fulfilling the tight requirement are selected. Furthermore they must have $p_T > 25$ GeV and $|\eta| < 1.37$ or $1.52 < |\eta| < 2.37$.

- $E_{x(y)}^{miss,e}$: electrons are selected using the medium working point. In addition they must have $p_T > 10$ GeV and $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$.

- $E_{x(y)}^{miss,\tau}$: it is given by the contribution of hadronically decaying τ. They are required to pass the medium selection and must have $p_T > 20$ GeV and $|\eta| < 1.37$ or $1.52 < |\eta| < 2.37$.

- $E_{x(y)}^{miss,jets}$: it is obtained using calorimetric jets reconstructed with the anti-k_t algorithm with $R = 0.4$. They must have $p_T > 20$ GeV and an additional requirement on the jet vertex tagger variable, $JVT > 0.64$, is applied to jets with $|\eta| < 2.47$ and $p_T < 50$ GeV in order to remove pile-up jets.

- $E_{x(y)}^{miss,soft}$: as already mentioned, this term takes into account all the momentum of all particles not associated with any physical object. It is estimated from tracks with $p_T > 0.4$ GeV and $|\eta| < 2.5$. The tracks must be matched with the primary vertex in order to reduce contribution from pile-up jets.

From the x and y components the vector $\vec{E}_T^{miss}$, the magnitude E_T^{miss} and the azimuthal angle ϕ^{miss} are calculated as follows:

$$\begin{aligned} \vec{E}_T^{miss} &= (E_x^{miss}, E_y^{miss}) \\ E_T^{miss} &= \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \\ \phi^{miss} &= \arctan(E_y^{miss}/E_x^{miss}) \end{aligned} \tag{4.5}$$

Chapter 5

Cross section measurement of the Higgs decaying into a pair of tau leptons

This thesis work has been performed within the $H \to \tau\tau$ cross section and CP violation analysis where the former aims at measuring the cross section of the Higgs decaying in two tau leptons and the latter measures the Higgs boson CP properties in the same decay channel. The two analysis are closely related and they will be illustrated in this Chapter and in Chapter 6 respectively.

5.1 Introduction

After the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments, the studies have been focused on measuring its properties such as its coupling strengths and spin and CP quantum numbers. As already described in Section 1.4.4, the measurements performed during Run 1 of the LHC at a center of mass energy of $\sqrt{s} = 7$ and 8 TeV have not shown any deviation from the SM predictions.
Among fermion couplings, only the ones to the third generation fermions are currently accessible at the LHC. In particular both the coupling to bottom quarks [84] [85], through the $H \to b\bar{b}$ decay channel, and the coupling to top quarks [86] [87], through the ttH production mode, have been observed. They are both challenging since the former requires high b-tagging efficiency and the latter has a small cross section. In this context, the $H \to \tau\tau$ decay channel is considerably important since it allows to direcly measure the Higgs Yukawa coupling in the leptonic decay mode (limits have been set to the $H \to \mu\mu$ but there is not an evidence yet). A first measurement of the Higgs coupling to τ leptons have been performed by ATLAS during Run 1 [88], using ~ 25 fb^{-1} of data collected at centre of mass energies of $\sqrt{s} = 7$ and 8 TeV. The measured signal strength, defined as the measured cross section value over the SM expectation, was found to be $\mu = 1.43^{+0.43}_{-0.37}$, therefore consistent with the predicted Yukawa coupling in the SM. The signal excess over the expected background was found with an observed (expected) significance of 4.5 (3.4) standard deviations. In addition, the observation of $H \to \tau\tau$ decay has already been established by the ATLAS and CMS combined measurement using Run 1 data with a significance of 5.5 standard deviations [6]. Nevertheless the $H \to \tau\tau$ will benefit from the higher Run 2 statistics in order to reach higher precision results and it has also good potentiality for measuring the Higgs properties as it will be illustrated in Chapter 6.

The analysis, presented in this Chapter [89], has been performed using the 2015 and 2016 data recorded by the ATLAS detector at the LHC, corresponding to an integrated luminosity of 36 fb^{-1}. The tau lepton, with his brief lifetime of $\tau = 2.9 \times 10^{-13}$, is not visible in the detector, and it is accessible only through its decay products. All combinations of possible decay modes have been exploited in the analysis: both leptonic ($\tau \to l\nu\bar{\nu}$ with $l = e\mu$) and hadronic ($\tau \to$ hadrons ν). The three decay channels will henceforth be referred to as $\tau_{\text{lep}}\tau_{\text{lep}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$. The overall analysis strategy is the same across the various decay channels but the background composition is different, therefore different selections and background estimations are performed according to the specific decay channel.

5.2 Data and Monte Carlo samples

The data used in this analysis are taken from proton-proton collisions at $\sqrt{s} = 13$ TeV during 2015 and 2016, corresponding to an integrated luminosity of 3.2 fb^{-1} in 2015 and 32.9 fb^{-1}, for a total of 36.1 fb^{-1} integrated luminosity. Several triggers have been used for recording data for the analyisis: single light lepton, two light leptons and two hadronically decaying taus (more details in Section 5.5.1). Only the data which passed the data quality requirements are used.
The signal and background processes have been simulated using different Monte Carlo generators. For the signal, the main production processes of the Higgs at the LHC are considered: ggF, VBF, VH associated production and $t\bar{t}H$ associated production. The $b\bar{b}H$ and single top productions are found to be negligible. The Higgs decay to a pair of W bosons is also considered as a background to this analysis and thus it is simulated in the same way for these production processes. The Higgs mass is always assumed to be 125 GeV.
The ggF Higgs production is simulated at next-to-next-to-leading order (NNLO) accuracy in QCD using Powheg-Box v2 NNLOPS [90, 91, 92, 93, 94] generator with the MiNLO [95] approach. The VBF and VH associated production are simulated at next-to-leading-order (NLO) using Powheg-Box v2 as well, with the MiNLO approach. The $t\bar{t}H$ production is simulated using Madgraph5_aMC@NLO v2.2.2 [96] at NLO accuracy in QCD. All the matrix element generators for these signal samples are interfaced with the Pythia 8.212 [97] program used for simulation of parton shower, hadronization and underlying event (UEPS). In order to estimate parton shower and underlying event signal uncertainties, alternative samples interfaced with the Herwig 7.0.3 [98] program have been produced as will be described with more details in Section 5.7.1. The PDF4LHC15 [99] parametrization for the pdf set is used for these production processes. In addition, the AZNLO [100] tune, developed by the ATLAS collaboration, is used with the CTEQ6LI [101] pdf set for the simulation of non-perturbative effects. Only for the $t\bar{t}H$ production process the NNPDF30LO [102] pdf parametrization is used in the matrix element and the NNPDF23LO [103] pdf parametrization for the UEPS model. Photos ++ 3.52 [104] is used for QED emssions from electroweak vertices and charged leptons.
The ggF sample is normalized to the cross section prediction calculated at next-to-next-to-next-to-leading-order (N^{3}LO) QCD accuracy with NLO EW corrections applied. The

Process	Monte Carlo generator	UEPS	PDF	Cross section order
ggF	Powheg-Box v2	Pythia 8.212	PDF4LHC15 NNLO	N^{3}LO QCD + NLO EW
VBF	Powheg-Box v2	Pythia 8.212	PDF4LHC15 NLO	∼NNLO QCD + NLO EW
VH	Powheg-Box v2	Pythia 8.212	PDF4LHC15 NLO	NNLO QCD + NLO EW
$t\bar{t}H$	MG5_aMC@NLO v2.2.2	Pythia 8.212	NNPDF30LO	NLO QCD + NLO EW
W/Z + jets	Sherpa 2.2.1	Sherpa 2.2.1	NNPDF30NNLO	NNLO
VV/Vγ^*	Sherpa 2.2.1	Sherpa 2.2.1	NNPDF30NNLO	NLO
$t\bar{t}$	Powheg-Box v2	Pythia 6.428	CT10	NNLO+NNLL
Wt	Powheg-Box v1	Pythia 6.428	CT10F4	NLO

Table 5.1: Summary of Monte Carlo generators used for the production of signal and background samples. Matrix element generators are listed together with parton shower, underlying event and hadronization (UEPS) generators and pdf set. In addition, the perturbative order of the total cross section is given.

VBF production is normalized to the cross section calculated with NNLO QCD accuracy plus NLO EW corrections applied. The VH samples are normalized to cross sections calculated at NNLO in QCD with NLO EW corrections applied.
Backgrounds are also mainly estimated from Monte Carlo simulations. Samples of EW production of W/Z bosons, W/Z boson production with associated jets and diboson production were simulated with the Sherpa 2.2.1 [105] generator, where the matrix elements are calculated with the Comix [106] and OpenLoops [107] generators and then matched with the Sherpa UEPS model using the ME+PS@NLO prescription [108]. For W/Z production with associated jets the matrix elements is calculated for up to two partons at NLO in the final state and four partons at LO. The overall normalization is taken from a NNLO calculation from Fewz [109, 110]. For all these background samples the NNPDF30NNLO PDF set is used. Theory uncertainties related to the $Z \to \tau\tau$ background are estimated using an alternative sample generated with Madgraph5_aMC@NLO at LO interfaced with Pythia 8.186 as will be described in Section 5.7.4.
For the $t\bar{t}$ process the Powheg-Box v2 generator is used with the CT10 pdf set. The cross section is taken from a NNLO calculation including soft gluon resummation at next-to-next-to-leading-log (NNLL) from Top++ 2.0 [111] program. For single top quark production, Wt, the simulation is performed with Powheg-Box v1 generator. The latter uses the four flavor scheme for the NLO matrix element calculation with the fixed four flavor pdf set CT10F4. The parton shower program used for these processes involving top quark is Pythia 6.428 [112] with the CTEQ6L1 pdf set. The top mass is assumed to be 172.5 GeV.
The ATLAS detector response is also simulated for all the samples, using the Geant4 [69] program. The effect of pile-up interactions is included by overlaying minimum bias events simulated with Pythia 8.186 on each generated signal and background event. All the simulated samples are summarised in Table 5.1.

5.3 Object Reconstruction

The final state of $H \to \tau\tau$ involves different physics objects, therefore it is necessary to reconstruct τ leptons, electrons, muons, missing transverse momentum, $E_{\mathrm{T}}^{\mathrm{miss}}$, and jets. The object reconstruction is based on what is already described in Chapter 4 and will be illustrated in this Section.
Electron candidates are reconstructed with energy deposits clusters in the EM calorimeter and matched with a track in the inner detector. Electrons passing the loose requirement and with $p_T > 15$ GeV are considered. In addition, they must be in the acceptance region of the inner detector, $|\eta| < 2.47$, where the transition region between the barrel and the end-cap is excluded ($1.37 < |\eta| < 1.52$). The trigger efficiency for single electrons in the analysis is between 90% and 95%. Overlap between different objects is always resolved. In the case of electrons they are removed if they share their track with a muon candidate or if their distance to a jet is within $0.2 < \Delta R < 0.4$.
Muon candidates are reconstructed by matching a track in the muon spectrometer with a track in the inner detector. They have to fullfill the loose identification requirement and must have $p_T > 10$ GeV and $|\eta| < 2.5$. The trigger efficiency for single muons in the analysis is 70%-80% in the barrel and 90% in the end-caps. Muon overlapping with jets is always resolved and they are discarded if their distance to a jet is $\Delta R < 0.4$ with the following exceptions: if $\Delta R < 0.2$ or the muon track is associated to the jet and if the jet has either less than three tracks or less than twice the transverse momentum of the muon, it is removed instead.
In the case of $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels selections, electrons and muons are required to pass the medium identification criteria and to fullfill the isolation requirement, therefore no additional high-p_{T} track must be in a cone around the track and no significant energy deposit in a cone around the calorimeter cluster.
Jets are reconstructed from topological clusters in the hadronic calorimeter using the anti-k_T algorithm with radius parameter $R = 0.4$. In order to reject pile-up jets the JVT algorithm is used for jets with $p_T < 50$ GeV and $|\eta| < 2.4$. A similar algorithm called "forward JVT" is applied to all jets with $p_T < 50$ GeV in the forward region, $|\eta| > 2.5$. b-jets are also selected using a multivariate algorithm applied to all jets in the $|\eta| < 2.5$ region. A working point corresponding to an efficiency of 85% for b-jets and rejection factors of 2.8 and 28 for c-jets and light-flavor jets respectively, is used. A jet is removed if it is within $\Delta R < 0.2$ with respect to an electron or an hadronically decaying τ candidate.
The missing transverse momentum vector is calculated as the negative vectorial sum of the p_{T} of all the fully reconstructed and calibrated physics objects, including also a soft term which takes into account all the tracks associated with the hard scattering but not associated to any reconstructed objects. The missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$ is the modulus of this vector).
Finally, τ leptons must be reconstructed from their decay products. Leptonic tau decays are reconstructed as electrons and muons as already described. The visible part of hadronic tau decays ($\tau_{\mathrm{had\text{-}vis}}$) is instead reconstructed from jets with $p_T > 10$ GeV and $|\eta| < 2.5$, excluding the transition region between the barrel and the end-cap calorimeters. The background from quark or gluon-initiated jets is rejected using a multivariate

algorithm. $\tau_{\text{had-vis}}$ are required to pass the loose identification criteria. In addition, they must have $p_T > 20$ GeV and to have one or three associated tracks and an absolute electric charge of one. The trigger efficiency of single $\tau_{\text{had-vis}}$ selected in the analysis is 95% for 1-prong and 85% for 3-prong taus. If the $\tau_{\text{had-vis}}$ candidate is within a cone of $\Delta R < 0.4$ with an electron candidate, it must have a low electron identification likelihood score, otherwise it is discarded. In this analysis further tight requirements are applied. In the $\tau_{\text{lep}}\tau_{\text{had}}$ selection, the $\tau_{\text{had-vis}}$ candidates must pass the medium identification criteria leading to an efficiency of 55% and 40% for real 1-prong and 3-prong $\tau_{\text{had-vis}}$ respectively. In addition, a multivariate algorithm is applied if a 1-prong $\tau_{\text{had-vis}}$ and an electron are selected, in order to reject $Z \to ee$ background. In the $\tau_{\text{had}}\tau_{\text{had}}$ selection the candidates are required to pass the tight identification criteria instead, corresponding to a selection efficiency of 45% and 30% for real 1-prong and 3-prong $\tau_{\text{had-vis}}$ respectively.

5.4 Missing Mass Calculator

Reconstructing the Higgs invariant mass in the $H \to \tau\tau$ decay channel is challenging because each τ decay, both leptonic and hadronic, involves neutrinos in the final state. The reconstruction of the ditau invariant mass ($m_{\tau\tau}$) is therefore performed in this analysis using the Missing Mass Calculator (MMC) [113] algorithm.
The MMC method allows to perform a complete reconstruction of the event kinematics and to have improved invariant mass and neutrino momentum resolution, as opposed to other methods like the collinear approximation [114], which is limited by the assumption of having the visible and invisible τ decay products collinear and consequently it is not able to reconstruct $m_{\tau\tau}$ for each event topology.
The starting assumption of the MMC is that the transverse missing energy is only due to neutrinos from the τ decays. Under this assumption, the full reconstruction of the event requires to solve the following system:

$$\begin{aligned}
E_{T_x}^{miss} &= p_1^{miss}\sin\theta_1^{miss}\cos\phi_1^{miss} + p_2^{miss}\sin\theta_2^{miss}\cos\phi_2^{miss} \\
E_{T_y}^{miss} &= p_1^{miss}\sin\theta_1^{miss}\sin\phi_1^{miss} + p_2^{miss}\sin\theta_2^{miss}\sin\phi_2^{miss} \\
m_{\tau_1}^2 &= (m_1^{miss})^2 + (m_1^{vis})^2 + 2\sqrt{(p_1^{vis})^2 + (m_1^{vis})^2}\sqrt{(p_1^{miss})^2 + (m_1^{miss})^2} \\
&\quad - 2p_1^{vis}p_1^{miss}\cos\Delta\theta_1^{vm} \\
m_{\tau_2}^2 &= (m_2^{miss})^2 + (m_2^{vis})^2 + 2\sqrt{(p_2^{vis})^2 + (m_2^{vis})^2}\sqrt{(p_2^{miss})^2 + (m_2^{miss})^2} \\
&\quad - 2p_2^{vis}p_2^{miss}\cos\Delta\theta_2^{vm}
\end{aligned} \tag{5.1}$$

where $E_{T_x}^{miss}$ and $E_{T_y}^{miss}$ are the x and y components of the $\vec{E}_T^{miss}$ vector, $p_{1,2}^{vis}$, $m_{1,2}^{vis}$, $\theta_{1,2}^{vis}$, $\phi_{1,2}^{vis}$ are the momenta, invariant masses, polar and azimuthal angle respectively of the τ visible decay products and m_τ is the τ invariant mass.The quantities $p_{1,2}^{miss}$, $m_{1,2}^{miss}$, $\theta_{1,2}^{vis}$, $\phi_{1,2}^{vis}$ are instead the momenta, invariant masses, polar and azimuthal angle respectively of the invisible τ decay products. Finally, $\Delta\theta_{1,2}^{vm}$ is the angle between $\vec{p}^{miss}$ and $\vec{p}^{vis}$ for each of the two τ leptons. The variables related to the invisible decay products are the unkowns of the system. In the case of the hadronic τ decays there

is only one neutrino involved and the m^{miss} is set to zero, thus reducing the number of the unknowns. Therefore, there are 6 to 8 unknowns depending on the number of τ leptonic decays but only four equations relating them and the system is then under-constrainted. Even though it is not possible to calculate the exact solution, not all the solutions are equally likely. The idea of the MMC is to use additional kinematic information in order to distinguish between less likely and more likely solutions and compute the $m_{\tau\tau}$ invariant mass with these additional constraints. A possible variable which provides this additional knowledge is the distance between the direction of the visible and invisible τ decay products, $\Delta R = \sqrt{(\eta^{vis} - \eta^{miss}) + (\phi^{vis} - \phi^{miss})^2}$. This quantity depends on the τ decay type (leptonic, 1-prong hadronic or 3-prong hadronic) and on the τ momentum. The ΔR distance is used as probability density function in order to build a global event probability.
The method starts from considering the (ϕ_1^{miss}, ϕ_2^{miss}, m_1^{miss}, m_2^{miss}) parameter space which is reduced to (ϕ_1^{miss}, ϕ_2^{miss}, m_1^{miss}) if only one τ decays leptonically and to (ϕ_1^{miss}, ϕ_2^{miss}) if both τ decay hadronically. For each point in this grid, it is possible to define the vectors $\vec{p}_{1,2}^{miss}$ and calculate the ΔR distance between the $\vec{p}_{1,2}^{vis}$ and the calculated $\vec{p}_{1,2}^{miss}$. In order to evaluate the probability of each decay topology the ΔR distribution is used and since this quantity depends only on the τ decay type and on the τ momentum and not on the source of τ (this is true if we neglect τ polarization), the shape of the distribution is derived from $Z/\gamma^* \to \tau\tau$ simulations. Fits to the distributions are performed in order to fully parametrize ΔR and this yields to the probability density function $\mathcal{P}(\Delta R, p_\tau)$, which can be used to evaluate the probability of a particular τ decay topology. The global event probability is then defined as follows:

$$\mathcal{P}_{event} = \mathcal{P}(\Delta R_1, p_{\tau 1}) \times \mathcal{P}(\Delta R_2, p_{\tau 2}) \tag{5.2}$$

A scan is performed over each point of the grid, calculating $m_{\tau\tau}$ and weighting it by the corresponding probability, $\mathcal{P}(\Delta R_1, p_{\tau 1}) \times \mathcal{P}(\Delta R_2, p_{\tau 2})$. The position of the maximum of the $m_{\tau\tau}$ distribution is then used as the final estimator for a given event.
The MMC is currently the best technique available in terms of resolution for reconstructing the Higgs invariant mass in the $H \to \tau\tau$ decay channel. The MMC performances are shown in Figure 5.1 and they are evaluated using $Z \to \tau\tau$ and $gg\mathrm{F} \to H \to \tau\tau$ simulated events. The $H \to \tau\tau$ sample is simulated assuming $m_H = 115$ GeV. In all cases the peak of the distribution is within $\sim 2\%$ of the true mass. The resolution, in the range of $(1.0 \pm 0.4) \times m_{\tau\tau}^{true}$, is from $\sim 8\%$ to $\sim 13\%$ according to the τ decay type (the leptonic decay mode is worse due to the weaker constraintts on the system). The fraction of events that cannot be reconstructed because the Equations 5.1 are not solved for each grid point, ranges from $\sim 1\%$ to $\sim 3\%$ and thus demonstrating that the MMC has also high reconstruction efficiency.

An example of mass reconstructed with the MMC technique in the case of the specific analyisis described in this Chapter, is shown in Figure 5.2. Distributions of $m_{\tau\tau}$ for $Z \to \tau\tau$ and $H \to \tau\tau$ processes after preselection (see Section 5.5) are shown for each decay channel. The standard deviation of the reconstructed ditau mass is 17.0 GeV, 15.3 GeV and 14.7 GeV for $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ and $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ respectively.

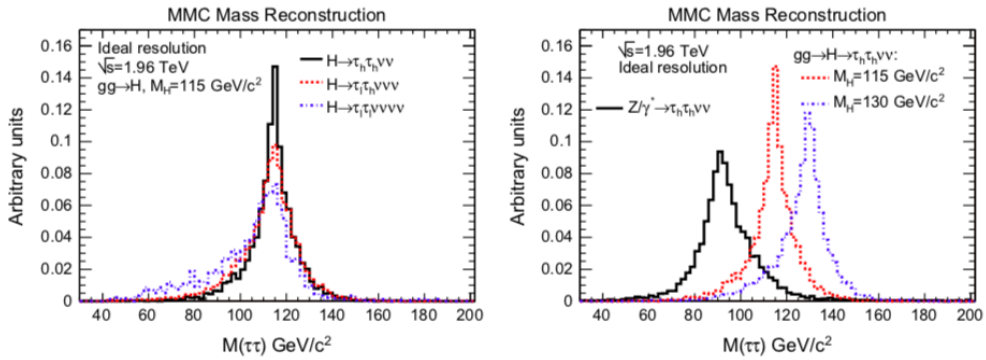


Figure 5.1: MMC performances are shown. The left plot shows the reconstructed $m_{\tau\tau}$ assuming $m_H = 115$ GeV for each of the three decay channels: $\tau_{\text{lep}}\tau_{\text{lep}}$ is the blue dashed-dotted line, $\tau_{\text{lep}}\tau_{\text{had}}$ is the red dashed line and finally $\tau_{\text{had}}\tau_{\text{had}}$ is the black solid line. The right plot shows the reconstructed mass $m_{\tau\tau}$ in the $\tau_{\text{had}}\tau_{\text{had}}$ decay channel both for $Z/\gamma^* \to \tau\tau$ and for $gg\text{F} \to H \to \tau\tau$ events assuming $m_H = 115$ GeV and $m_H = 130$ GeV [113].

5.5 Event selection

After the object selection, additional requirements are applied in the analysis. The triggers and the event preselection will be described in Section 5.5.1. Events passing the preselection cuts are further categorized into signal regions (SRs) which have enhanced signal to background ratio. In addition, control regions (CRs) are defined with considerable contribution of only a specific background in order to constraint its normalization. Finally, validation regions (VRs) are used for validate the modelling of the $Z \to \tau\tau$ irreducible background. Description of SRs, CRs and VRs is given in Section 5.5.2.

5.5.1 Preselection

As already mentioned, specific triggers are applied for recording data for this analysis. Depending on the trigger, different transverse momentum cuts are employed for selecting electron, muon and τ_{had} candidates. The p_{T} tresholds are different for 2015 and 2016 data, in particular they increase due to the high luminosity and pile-up during 2016. Summary of the triggers and p_{T} requirements can be found in Table 5.2. In the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel, triggers for two light leptons are applied only if the highest p_{T} lepton does not pass the single light lepton trigger p_{T} requirement. In addition, since 2016, the Di-$\tau_{\text{had-vis}}$ trigger requires a jet with $p_T > 25$ GeV calibrated at trigger level with $|\eta| < 3.2$.

Preselection cuts are then applied with some differences across the decay channels. All channels are required to have the exact number of loose leptons or $\tau_{\text{had-vis}}$, otherwise the event is rejected. In addition, an electron, muon or $\tau_{\text{had-vis}}$ must fulfill tighter identification and/or isolation criteria according to what is described in Section 5.3 and they must be of opposite charge.

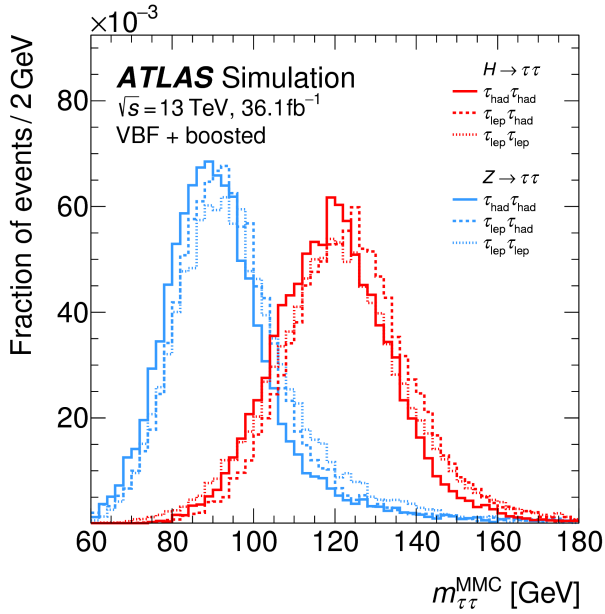


Figure 5.2: Invariant mass of the ditau system reconstructed with the MMC technique after event preselection of the $H \to \tau\tau$ cross section analysis. The blue curves are $Z \to \tau\tau$ events while the red ones are $H \to \tau\tau$ events (split by τ decay channel) [113].

Channel	Trigger	p_{T} treshold [GeV] 2015	p_{T} treshold [GeV] 2016
$\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ & $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	Single electron	25	27
	Single muon	21	27
$\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$	Di-electron	15/15	18/18
	Di-muon	19/10	24/10
	Electron+muon	18/15	18/15
$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$	Di-$\tau_{\mathrm{had\text{-}vis}}$	40/30	40/30

Table 5.2: Summary of the triggers used in the three decay channels together with the p_{T} requirement both for 2015 and 2016 data-taking.

$\tau_{\text{lep}}\tau_{\text{lep}}$		$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
$ee/\mu\mu$	e/μ		
$N_{e/\mu}^{\text{loose}} = 2$, $N_{\tau_{\text{had-vis}}}^{\text{loose}} = 0$		$N_{e/\mu}^{\text{loose}} = 1, N_{\tau_{\text{had-vis}}}^{\text{loose}} = 1$	$N_{e/\mu}^{\text{loose}} = 0, N_{\tau_{\text{had-vis}}}^{\text{loose}} = 2$
e/μ: Medium, gradient iso.		e/μ: Medium, gradient iso.	
		$\tau_{\text{had-vis}}$: Medium	$\tau_{\text{had-vis}}$: Tight
Opposite charge		Opposite charge	Opposite charge
$m_{\tau\tau}^{\text{coll}} > m_Z - 25$ GeV		$m_T < 70$ GeV	
$30 < m_{ll} < 75$ GeV	$30 < m_{ll} < 100$ GeV		
$E_T^{miss} > 55$ GeV	$E_T^{miss} > 20$ GeV	$E_T^{miss} > 20$ GeV	$E_T^{miss} > 20$ GeV
$E_T^{miss,hard} > 55$ GeV			
$\Delta R_{\tau\tau} < 2.0$		$\Delta R_{\tau\tau} < 2.5$	$0.8 < \Delta R_{\tau\tau} < 2.5$
$\lvert\Delta\eta_{\tau\tau} < 1.5\rvert$		$\lvert\Delta\eta_{\tau\tau} < 1.5\rvert$	$\lvert\Delta\eta_{\tau\tau} < 1.5\rvert$
$0.1 < x_1 < 1.0$		$0.1 < x_1 < 1.4$	$0.1 < x_1 < 1.4$
$0.1 < x_2 < 1.0$		$0.1 < x_2 < 1.2$	$0.1 < x_2 < 1.4$
$p_T^{j_1} > 40$ GeV		$p_T^{j_1} > 40$ GeV	$p_T^{j_1} > 70$ GeV, $\lvert\eta_{j_1}\rvert < 3.2$
$N_{\text{b-jets}} = 0$		$N_{\text{b-jets}} = 0$	

Table 5.3: Summary of the event preselection requirements for the the three decay channels.

In order to select events which contain neutrinos, the $E_{\text{T}}^{\text{miss}}$ is required to be $E_{\text{T}}^{\text{miss}} > 20$ GeV. In the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel with two same flavour leptons in the final state, this requirement is tighten in order to reject $Z \to ll$ background. For the same reason, further cuts on the invariant mass of the two leptons and on the missing transverse energy calculated only with the hard terms ($E_{\text{T}}^{\text{miss, hard}}$) are applied. Additional requirements on the angular distance between the tau candidates and their difference in pseudorapidity are applied in order to reject non-resonant background. Requirements are also applied on the fraction of the τ candidate momentum carried by the visible decay product, $x_i = p_i^{\text{vis}}/(p_i^{\text{vis}} + p_i^{\text{miss}})$. which is calculated in the collinear approximation, thus the visible and invisible decay products are assumed to be collinear. The reason for this requirement is to suppress events with $E_{\text{T}}^{\text{miss}}$ not compatible with that of a ditau decay. The transverse mass calculated from $E_{\text{T}}^{\text{miss}}$, light lepton and $\tau_{\text{had-vis}}$ is required to be lower than 70 GeV in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel to reject W+jets events with leptonic decays. In the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel the mass calculated with the collinear approximation is required to be $m_{\tau\tau}^{coll} > m_Z - 25$ GeV to further reduce $Z \to ll$ events and ensure orthogonality with the $H \to WW$ decay channel which can have the same final state as $\tau_{\text{lep}}\tau_{\text{lep}}$.
Further requirements are applied on the jets. Since this analysis targets in the signal region VBF events or ggF events produced in a boosted regime, at least one jet is required with minimal p_{T} requirement. In the $\tau_{\text{had}}\tau_{\text{had}}$ channel this requirement is raised to $p_T > 70$ GeV in order to achieve uniform trigger efficiency (as already mentioned in this section, the 2016 trigger for $\tau_{\text{had-vis}}$ requires at least one jet). Finally in the $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels no jets tagged as b with $p_T > 25$ GeV are required in order to suppress the top quark background. All these preselection cuts are listed in details in Table 5.3.

Signal region		Inclusive	$\tau_{\text{lep}}\tau_{\text{lep}}$	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
VBF	High-$p_T^{\tau\tau}$	$p_T^{j_2} > 30$ GeV $\lvert\Delta\eta_{jj}\rvert > 3$	–		$p_T^H > 140$ GeV $\Delta R_{\tau\tau} < 1.5$
	Tight	$m_{jj} > 400$ GeV $\eta_{j_1}\dot{\eta}_{j_2} < 0$ central leptons	$m_{jj} > 800$ GeV	$m_{jj} > 800$ GeV $p_T^{\tau\tau} > 100$ GeV	Not-VBF high-$p_T^{\tau\tau}$ $m_{jj} > (1550 - 250\dot{\Delta}\eta_{jj})$ GeV
	Loose		Otherwise		
Boosted	High-$p_T^{\tau\tau}$	Fail VBF $p_T^{\tau\tau} > 100$ GeV	$p_T^{\tau\tau} > 140$ GeV $\Delta R_{\tau\tau} < 1.5$		
	Low $p_T^{\tau\tau}$		Otherwise		

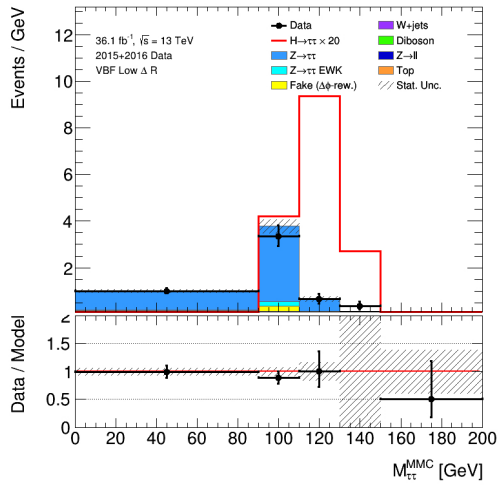
Table 5.4: Summary of the signal regions definitions for each decay channel. Two VBF and two boosted signal regions are exploited for each decay channel, apart from $\tau_{\text{had}}\tau_{\text{had}}$ channel which uses an additional VBF region (VBF high-p_T^H).

5.5.2 Signal, control and validation regions

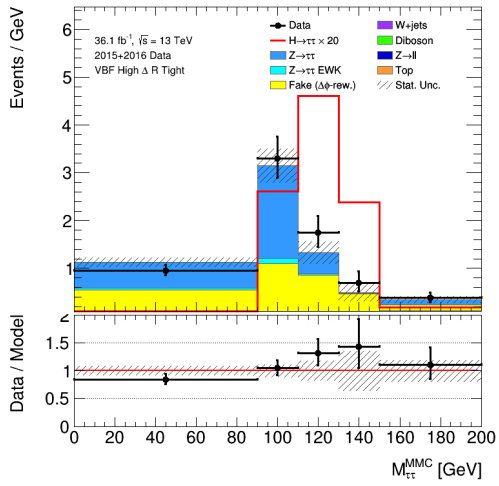
After applying the preselection requirements, further cuts are employed in order to select regions with enhanced signal to background ratio. Two kinds of signal regions are defined which exploit signal sensitive event topologies: a region called "VBF" and a region called "boosted". The VBF category targets events produced through VBF Higgs production mode and it is characterized by two high-p_{T} jets which have a large pseudorapidity gap ($|\Delta\eta_{jj}| > 3$) and have large invariant mass ($m_{jj} > 400$ GeV). In addition, the selected τ candidates must have η values that lie between those of the two jets. This region has a sizeable contamination of ggF events which amounts up to $\sim 30\%$ of the total expected Higgs signal.
The boosted category targets events which are produced through ggF with an additional recoiling jet. These events must fail the VBF selection and have a boosted Higgs boson ($p_{\text{T}}^{\tau\tau} > 100$ GeV). This boosted regime, which allows to select ggF plus one additional jet events, is motivated by a reduced $Z \to \tau\tau$ background contamination (on average the boson has an higher p_{T} for $H \to \tau\tau$.) The boosted category contains also contributions from other Higgs production mode, in particular sizeable contribution from VBF and VH production up to $\sim 20\%$ of the expected Higgs signal. The two signal regions are then defined in a mutual exclusive way and events that do not lie neither in the VBF nor in the boosted category are not used in the analysis. The VBF and boosted inclusive signal regions are further splitted into subregions in order to improve the sensitivity: $p_T^{\tau\tau}$, $\Delta R_{\tau\tau}$ and m_{jj} are used to define further selections, ending up with a total of 13 signal regions as it is shown in Table 5.4.

m_{MMC} distributions in the various signal regions for the $\tau_{\text{had}}\tau_{\text{had}}$ channel are shown as an example in Figure 5.3 and Figure 5.4 for VBF and boosted regions respectively.

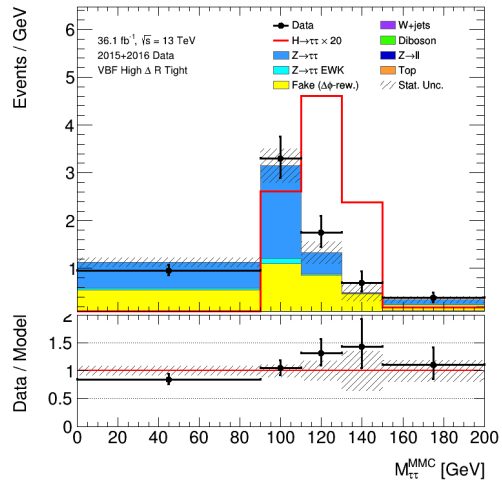
Six control regions are defined in a way that the contribution of a particular background is dominant and they are used to constraint background normalization. Regions of this kind are defined for the $Z \to ll$ and top quark backgrounds. In particular two $Z \to ll$ regions are determined for $\tau_{\text{lep}}\tau_{\text{lep}}$ by applying the VBF inclusive and the boosted inclusive selection respectively plus requiring the m_{ll} to be $80 < m_{ll} < 100$ GeV. These



(a) VBF high-p_{T}



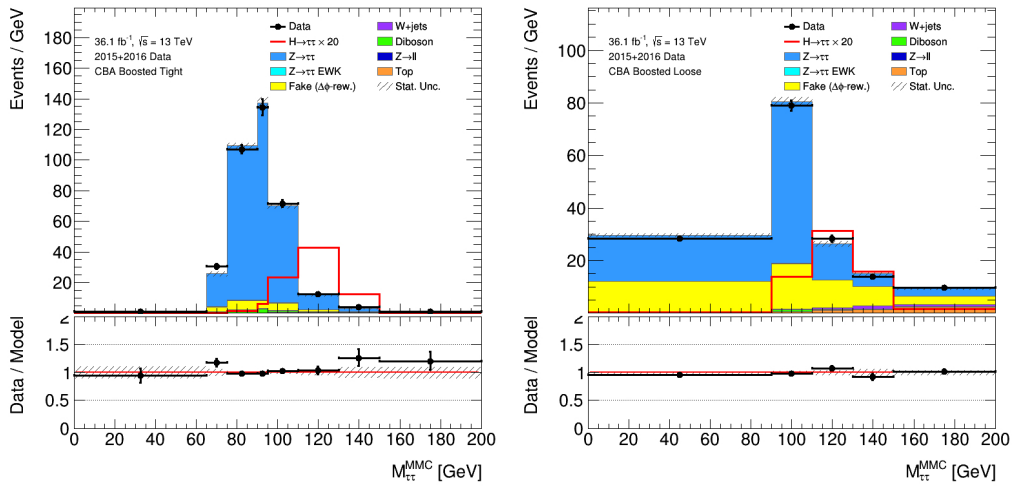
(b) VBF tight



(c) VBF loose

Figure 5.3: m_{MMC} distribution in the various VBF $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ signal regions: VBF high-p_{T} (a), VBF tight (b) and VBF loose (c). The red line correspond to the Higgs expected signal multiplied by 20 and the filled colored histrograms correspond to the various background predictions which will be discussed in details in Section 5.6. The uncertainties are represented by the hashed band which include only statistical uncertainties.

regions consist of more than 90% of $Z \to ll$ events. The top quark background is characterized by the presence of b-jets. Therefore four CRs are defined by inverting the b-jet veto requirement respectively in the VBF and boosted region for $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$, and in the VBF and boosted region for $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$. In addition, in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ regions, the requirement on the transverse mass, $m_T < 70$ GeV, is replaced by $m_T > 40$ GeV, to further enhance the



(a) Boosted tight (b) Boosted loose

Figure 5.4: m_{MMC} distribution in the two boosted $\tau_{\text{had}}\tau_{\text{had}}$ signal regions: boosted tight (a) and boosted loose (b). The red line correspond to the Higgs expected signal multiplied by 20 and the filled colored histrograms correspond to the various background predictions which will be discussed in details in Section 5.6.The uncertainties are represented by the hashed band which include only statistical uncertainties.

contribution of top quark background. The contribution of top quark events in these CRs is $\sim 80\%$ in $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\sim 60-70\%$ in $\tau_{\text{lep}}\tau_{\text{had}}$. Top quark and $Z \to ll$ backgrounds are instead negligible in $\tau_{\text{had}}\tau_{\text{had}}$. Summary of these CRs definitions can be found in Table 5.5.

The normalization of the major irreducible background, which is $Z \to \tau\tau$, is instead directly determined from the fit to data as will be discussed later in Section 5.8.2. In order to validate the modelling of this background, a $Z \to \tau\tau$ validation region is defined (one for each signal region). Since it is difficult to select a pure $Z \to \tau\tau$ sample, this VR is composed of $Z \to ll$ events with similar kinematics to $Z \to \tau\tau$ events. This is achieved by considering the $\tau_{\text{lep}}\tau_{\text{lep}}$ same flavor preselection region but with the following differences: the requirements on $m_{\tau\tau}^{\text{coll}}$, $E_{\text{T}}^{\text{miss}}$ and $E_{\text{T}}^{\text{miss,hard}}$ are dropped and the m_{ll} cut is inverted to $m_{ll} > 80$ GeV. The other requirements on the τ lepton decays are replaced by requirements on the light leptons. Finally, the requirements on the $p_{\text{T}}^{\tau\tau}$ are also replaced by requirements on the p_{T} of the Z boson which is obtained from the p_{T} of light leptons. Figure 5.5 summarises the composition of all the signal and control regions where the meaning of the backgrounds and their estimation will be described in Section 5.6.

5.6 Background estimation

The different final states of the three decay channels imply different background compositions and also need different strategies for background estimation. Both simulation and

Channel	Region	Selection
$\tau_{\text{lep}}\tau_{\text{lep}}$	VBF $Z \to ll$ CR	$\tau_{\text{lep}}\tau_{\text{lep}}$ VBF incl. selection, $80 < m_{ll} < 100$ GeV (SF)
	boosted $Z \to ll$ CR	$\tau_{\text{lep}}\tau_{\text{lep}}$ boosted incl. selection, $80 < m_{ll} < 100$ GeV (SF)
	VBF top CR	$\tau_{\text{lep}}\tau_{\text{lep}}$ VBF incl. selection, inverted b-jet veto
	boosted top CR	$\tau_{\text{lep}}\tau_{\text{lep}}$ boosted incl. selection, inverted b-jet veto
$\tau_{\text{lep}}\tau_{\text{had}}$	VBF top CR	$\tau_{\text{lep}}\tau_{\text{had}}$ VBF incl. selection, inverted b-jet veto, $m_T > 40$ GeV
	boosted top CR	$\tau_{\text{lep}}\tau_{\text{had}}$ boosted incl. selection, inverted b-jet veto, $m_T > 40$ GeV

Table 5.5: Definitions of the CRs in the various channels, used to constraint $Z \to ll$ and Top backgrounds.

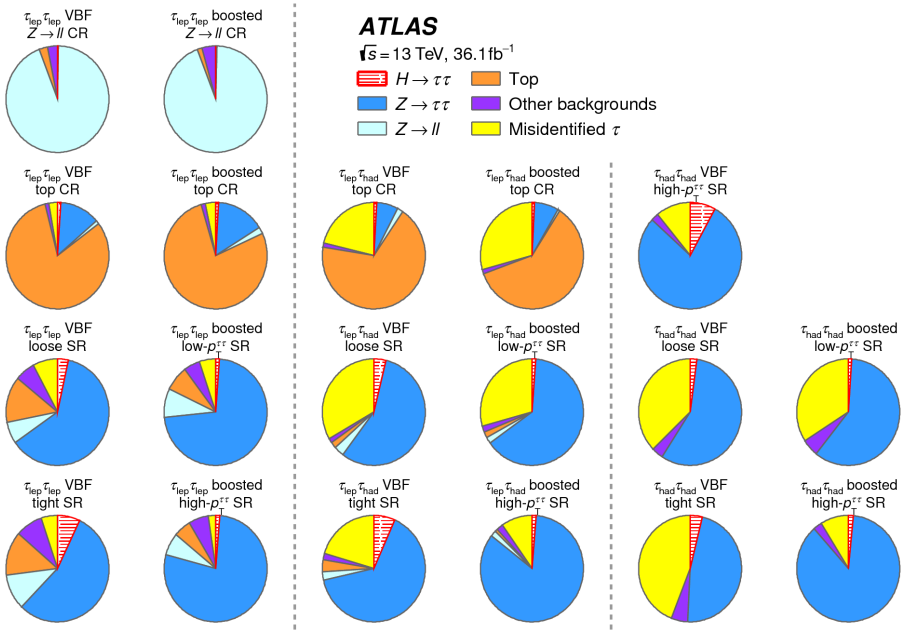


Figure 5.5: Expected signal and background composition of the 13 signal regions and the 6 control regions used in the analysis [89].

data-driven techniques are used to model the backgrounds in the three decay channels and a description of all of these methods can be found in this Section.

5.6.1 $Z \to \tau\tau$ background

The Drell-Yan process $Z/\gamma^* \to \tau\tau$ constitutes the main irreducible background and its contribution is in the range of 50%-90% of the total background yield depending on the signal region. The separation between $Z \to \tau\tau$ and $H \to \tau\tau$ is limited by the $m_{\tau\tau}^{\text{MMC}}$ resolution, indeed the $Z \to \tau\tau$ invariant mass distribution tail lies on the Higgs signal region (considering $m_H = 125$ GeV).
The $Z \to \tau\tau$ background is estimated using the same strategy across the different channels, namely using simulations from SHERPA at NLO and alternative MADGRAPH5_aMC@NLO

samples are used for estimating UEPS theoretical uncertainties. The normalization of this major background is directly retrieved from the fit to data and no specific control region is defined. Therefore, in order to verify the correctness of the $Z \to \tau\tau$ background modelling, a validation region is used instead. The purpose of this study is to validate the Z boson and the jet kinematics where the latter is particularly challenging to simulate for the MC generators.
The validation is performed using $Z \to ll$ events since it is difficult to select a pure $Z \to \tau\tau$ sample. The VR is then built using a selection as much as possible closer to the signal regions where the variables referred to τ are replaced with variables referred to muons or electrons, as already described in Section 5.5.2. Going more into detail, this is obtained by requiring:

- Two leptons with same flavor and opposite charge (e^+e^- or $\mu^+\mu^-$)
- $m_{ll} > 80$ GeV
- $|\Delta\eta_{ll}| < 1.5$, $\Delta R_{ll} < 2.5$
- At least one jet with $p_T > 40$ GeV for $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ and 70 GeV for $\tau_{\text{had}}\tau_{\text{had}}$
- Transverse momentum of the leading $p_T^{l_1} > 40$ GeV for $\tau_{\text{had}}\tau_{\text{had}}$

The events are then further categorised into VBF-like and boosted-like regions based on the selection for the signal regions:

- VBF-like: a second jet with $p_T^{j_2} > 30$ GeV, $m_{jj} > 400$ GeV, $|\Delta\eta_{jj}| > 3$, $\eta_{j_1} \times \eta_{j_2} < 0$
- Boosted-like: fail VBF selection, $p_T^{ll} > 100$ GeV

Figure 5.6 shows some examples of variables related to jet kinematics or used in the analyisis for categorization. The agreement between the data and the SHERPA simulation is generally good within the statistical and systematic uncertainties. The only variable which shows some trend is the m_{jj} distribution and therefore a reweighting based on the behaviour of this variable was investigated. The reweighting was intended to act only on the respective normalizations of the VBF sub-categories in each channel. As expected from the m_{jj} behaviour, it tends to reduce the size of $Z \to \tau\tau$ in the categories with high m_{jj} and to increase the amount of $Z \to \tau\tau$ events in the low m_{jj} categories. However, as Table 5.6 shows, it was found to have a negligible impact on the result, therefore no specific treatment is applied.

5.6.2 $Z \to ll$ background

Decays of Z boson into electrons or muons are significant particularly for the $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channel whereas they are basically negligible for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Mismeasurements of missing transverse energy can also bias the reconstructed $m_{\tau\tau}^{\text{MMC}}$ of these light lepton pairs, leading to values similar to those expected for the Higgs signal.

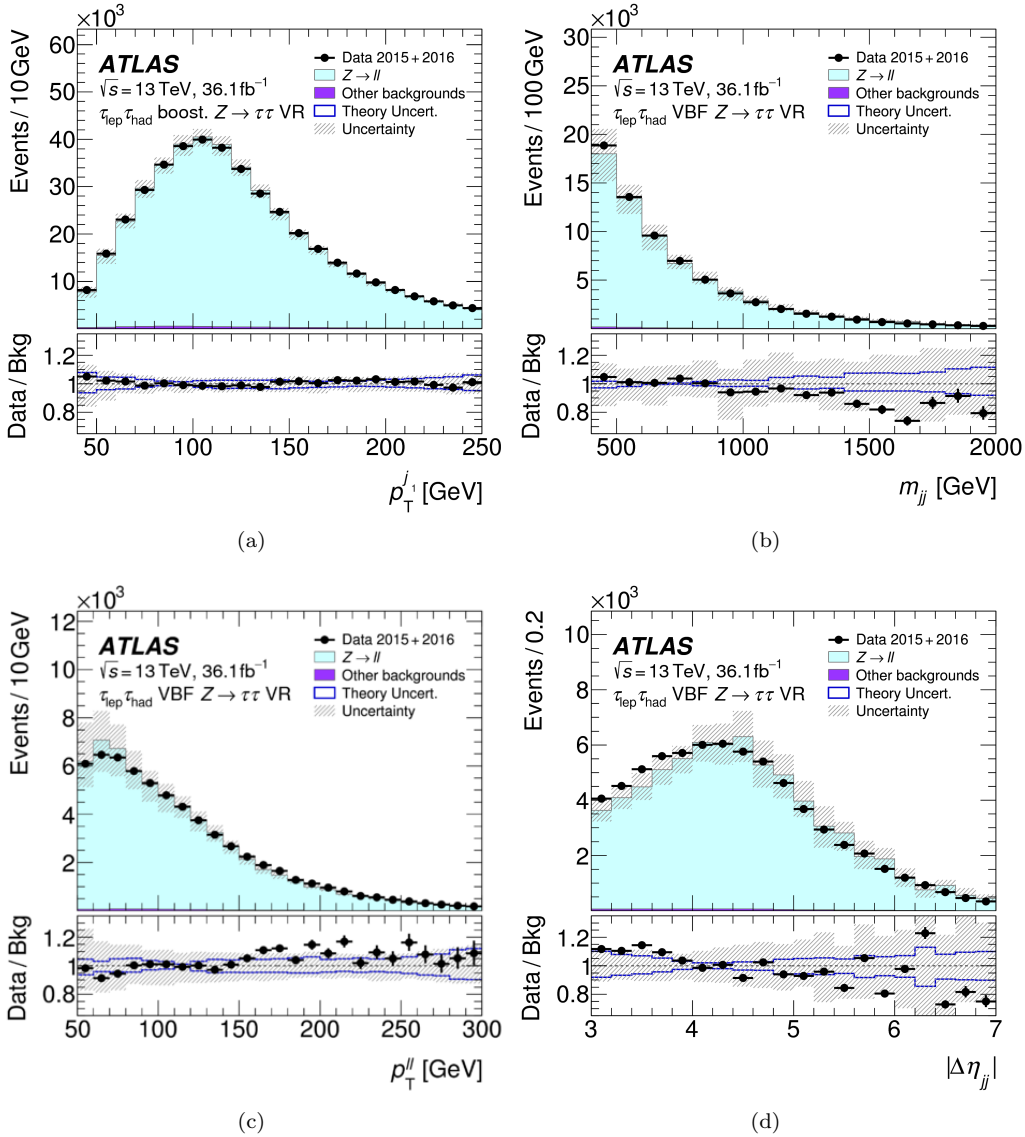


Figure 5.6: Observed and expected distributions in the $Z \to \tau\tau$ validation region, after performing the fit, for $\tau_{\text{lep}}\tau_{\text{had}}$ boosted inclusive category (a) and $\tau_{\text{lep}}\tau_{\text{had}}$ VBF inclusive category. The variables are examples of jet-related quantities or they are related to the event categorization: transverse momentum distribution of the highest p_{T} jet (a), invariant mass distribution of the system composed by the two highest p_{T} jets (b), transverse momentum distribution of the system composed by the two highest p_{T} leptons (c) and pseudorapidity separation betweeen the two highest p_{T} hets (d). All of these distributions shows good agreement between data and MC simulations, only m_{jj} shows a slightly trend in the tail but performing a reweighting based on the m_{jj} variable has a negligible impact on the result[89].

Channel	m_{jj} reweighting applied	nominal
$\tau_{\text{lep}}\tau_{\text{lep}}$	1.230	1.183
$\tau_{\text{lep}}\tau_{\text{had}}$	2.8126	2.8129
$\tau_{\text{had}}\tau_{\text{had}}$	2.844	2.838

Table 5.6: Impact of the m_{jj} reweighting on the expected significance in each channel.

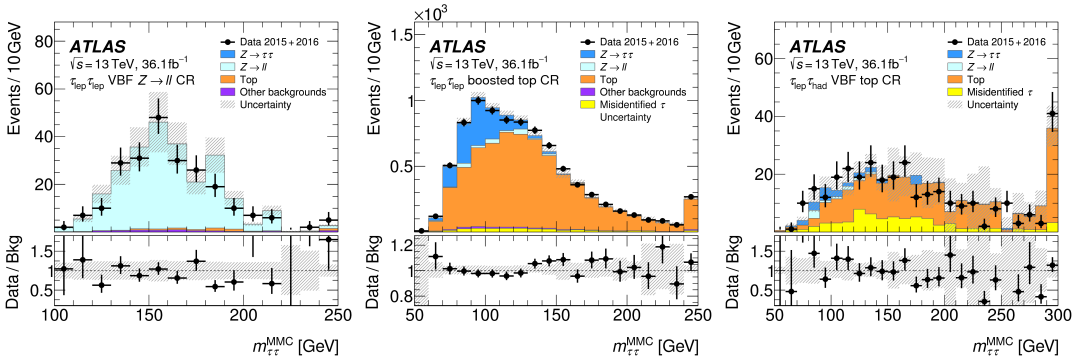
This background is estimated using simulations and, in order to constraint its normalization, a dedicated control region is used in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel. Figure 5.7 (a) shows a good modelling of the $m_{\tau\tau}^{\text{MMC}}$ in the $Z \to ll$ CR. In the other two channels the $Z \to ll$ yield is normalized to its theoretical prediction. In particular for the $\tau_{\text{lep}}\tau_{\text{had}}$ channel the main contribution comes from $Z \to ee$ events where an electron is misidentified as a τ_{had} but this background is significantly reduced by the veto algorithm applied to selected 1-prong τ_{had} candidates. The usage of a $Z \to ll$ CR in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel is therefore not necessary.

5.6.3 Top background

The production of $t\bar{t}$ pairs or of single top quarks, usually referred to as "top quark background" is a significant background for the $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels. This is due to the top quark decay chain, $t \to Wb$, $W \to l\nu, \tau\nu$. Events where a τ lepton decay product is misidentified are estimated using data-driven techniques as described in Section 5.6.4. The remaining top background is estimated using simulations and the normalization is constrainted using CRs both in the $\tau_{\text{lep}}\tau_{\text{lep}}$ and in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. Figure 5.7 (b-c) shows good modelling of the $m_{\tau\tau}^{\text{MMC}}$ distribution in the $\tau_{\text{lep}}\tau_{\text{lep}}$ boosted Top CR and in in the $\tau_{\text{lep}}\tau_{\text{had}}$ VBF top CR respectively.

5.6.4 Background from misidentified τ

A relevant background contribution comes from hadronic jets which can be misidentified as τ_{had} or even as electrons or muons. This background represents a significant amount since it contributes to up 40%-50% of the total background depending on the signal region, particularly for $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels. All of these backgrounds are estimated using data-driven techniques and different methods are applied according to the specific decay channel.
In the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel the main sources of this background arise from QCD jets (multi-jets), W bosons produced in association with jets and semileptonically decaying $t\bar{t}$ events. Dedicated regions are defined to estimate these backgrounds from data. They are obtained by "inverting" the isolation criteria, namely requiring that the light lepton with the second-highest p_{T} does not satisfy the gradient isolation criterion. In addition, if the light lepton is an electron, the identification criteria is relaxed to loose. Templates are then built in these data regions and the $Z \to ll$ and top quark contributions, that produce real light leptons, are estimated using MC and substracted from these samples. Scale factors are then derived for each template and applied, correcting for the inverted



(a) (b) (c)

Figure 5.7: Example of $m_{\tau\tau}^{\text{MMC}}$ distributions, after performing the fit, in the CRs: $\tau_{\text{lep}}\tau_{\text{lep}}$ VBF $Z \to ll$ CR (a), $\tau_{\text{lep}}\tau_{\text{lep}}$ boosted top CR (b) and $\tau_{\text{lep}}\tau_{\text{had}}$ VBF top CR (c). Data and predictions are in agreement within the uncertainties which include statistical and systematics uncertainties [89].

isolation requirement. This background from fake leptons is estimated in the same way in the top CR. Different systematic uncertainties both in the shape and normalization are taken into account due to the usage of this method. A non-closure systematic is estimated using events where the leptons are required to have the same charge and this leads to an uncertainty between 20% and 65% depending on the region. An uncertainty on the heavy flavour content is estimated using isolation correction factors that are computed with samples selected inverting the b-jet requirements. The resulted uncertainty is up to 50%. Other minor systematics refer to the fractional composition of the fake background in top quark decays, W+jets production and QCD jets.

In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, this background is mainly composed by QCD jets and W bosons produced in association with jets. The method used in this channel is a "fake-factor" method. Fake-factors are defined as the ratio of the number of events where the highest-p_{T} jet is identified as a medium τ_{had} candidate to the number of events where the highest-p_{T} jet passes a loose $\tau_{\tau_{\text{had}}}$ identification but fails the medium one. These factors depend on the p_{T} and track multiplicity of the τ_{had} candidate and on wether the jet is quark or gluon-initiated. Therefore two regions are defined: one called "W-enhanced" which is dominated by W+jets events and hence by quark-initiated jets, and defined by inverting the $m_T < 70$ GeV requirement and another region called "multijet-enhanced" which is dominated by QCD jets events and hence by gluon-initiated jets and defined by inverting the light lepton isolation. Other background sources like Z bosons produced in association with jets and semileptonic decays of $t\bar{t}$ were found to be negligible. The fake-factors derived in these data regions are firstly weighted by the multijets-to-W fraction. In order to estimate the background contribution for the signal regions, the weighted fake-factors are applied to events in regions defined with the same selection of the signal regions except that the τ_{had} candidate passes a loose τ_{had} requirement and fails the medium one (anti-ID regions). The relative multijet contribution in each anti-ID region is estimated from the

yields of events that fail the light-lepton isolation requirement multiplied by a correction factor that takes into account this requirement. The major systematic contribution for the fake background in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel originates from statitistical uncertainties in the individual fake factors, they are in a range of 10%-35% depending on the signal regions. Other minor contributions to the uncertainties come from statistical uncertainties in the anti-ID regions and uncertainties in the fractional composition of the fake background in multijets.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel the dominant contribution to this background is from QCD jets. Templates are built from data in a specific region (nOS region) defined by applying the same selection as in the signal regions but where the two $\tau_{\text{had-vis}}$ candidates are allowed to have two tracks and are required to fail the opposite charge requirement (in the SR they must have exactly one or three tracks and be of opposite charge). The contribution of events with true τ leptons from other SM processes, in particular from $Z \to \tau\tau$, is estimated with simulation and subtracted from the template. The nOS region was chosen because it provides the highest purity in selecting jets misidentified as τ. However, in order to estimate this background contribution in the signal region, the template must be reweighted to a region closer to the signal region. The reweighting is then based on scale factors dependent on the $\Delta\phi_{\tau\tau}$ between the two $\tau_{\text{had-vis}}$ candidates derived in a region where the $\tau_{\text{had-vis}}$ pair must have opposite charge and the second-highest p_{T} $\tau_{\text{had-vis}}$ must fail the tight but pass the medium identification requirement. Since the statistics of these events is relatively small, a selection with looser $\Delta\eta_{\tau\tau}$ and $\Delta R_{\tau\tau}$ is applied. A summary of all the mentioned control regions can be find in Table 5.7. The $\Delta\phi_{\tau\tau}$ is used as it was found to propagate efficiently the kinematic differences between these regions as it can be seen in Figure 5.8. Differently from the other channels, the fake background normalization in the $\tau_{\text{had}}\tau_{\text{had}}$ channel is not fixed but directly derived from the fit (one normalization factor correlated accross all $\tau_{\text{had}}\tau_{\text{had}}$ SRs). Figure 5.9 shows good agreement between data and background prediction. The distribution of $\Delta\eta_{\tau\tau}$ is here shown as it provides a good separation between $Z \to \tau\tau$ and fake backgrounds.

Uncertainties in the shape and normalization of the fake background are taken into account. The dominant contribution arises from the statistical uncertainties in the $\Delta\phi_{\tau\tau}$ scale factors and amounts to 8%. Systematic uncertainties for these scale factors are estimated by comparing them with scale factors computed in the noC region and in another region defined by requiring two $\tau_{\text{had-vis}}$ of opposite charge and that pass the loose but not the medium identification requirement. Other minor contributions comes from the extrapolation from the nOC region to the signal region and the uncertainty from the subtraction of simulated backgrounds.

Region	Charge	Tracks	Subleading τ ID
Signal region	$q(\tau_1) \times q(\tau_2) = -1$	Number of tracks $\in 1, 3$	tight
nOS control region	$q(\tau_1) \times q(\tau_2) \neq -1$	Number of tracks $\in 1, 2, 3$	tight
OS medium control region	$q(\tau_1) \times q(\tau_2) = -1$	Number of tracks $\in 1, 3$	medium not tight

Table 5.7: Summary of control regions used for the misidentified τ background estimation in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

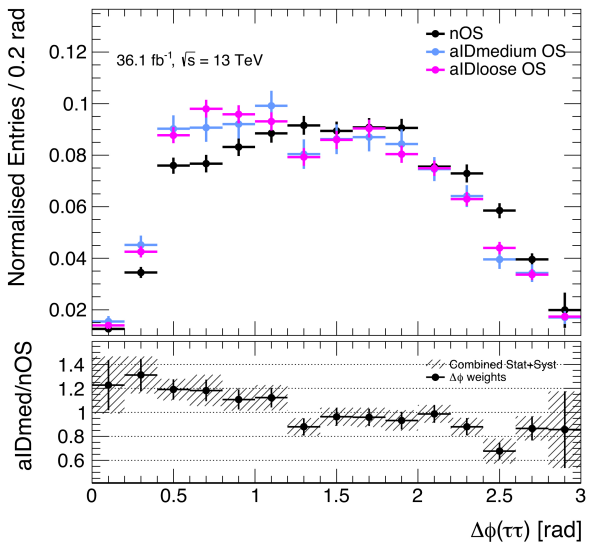


Figure 5.8: $\Delta\phi_{\tau\tau}$ distribution in the $\tau_{\text{had}}\tau_{\text{had}}$ control regions used for estimating the misidentified τ background: the nOS region where the template for this background is derived (in black), the OS medium region used for the reweighting (in blue) and alternative OS region which uses the loose ID requirement instead of the medium (in magenta). The bottom panel shows the ratio between the $\Delta\phi_{\tau\tau}$ distribution in the OS medium and nOS region, used for reweighting the template. The uncertainty band is constructed by replacing the medium not tight requirement on the second-highest p_{T} $\tau_{\text{had-vis}}$ candidate by the loose not tight requirement added in quadrature with the statistical uncertainty on the nominal weight.

5.6.5 Other backgrounds

Backgrounds that have a minor contribution are estimated using MC simulations and their normalization is fixed to their theoretical prediction. This includes diboson production and electroweak production of W/Z bosons. In addition there are contribution from $H \to WW$ decay in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel and from W bosons production in association with jets and $t\bar{t}$ production in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. These backgrounds are frequently referred to as "Others".

5.7 Systematic uncertainties

The expected signal and background yields in the different signal and control regions and the $m_{\tau\tau}^{\text{MMC}}$ shape are affected by systematic uncertainties. All the uncertainties taken into account will be described in this Section, except the ones coming from data-driven backgrounds estimation which are already described in Section 5.6.4. All the

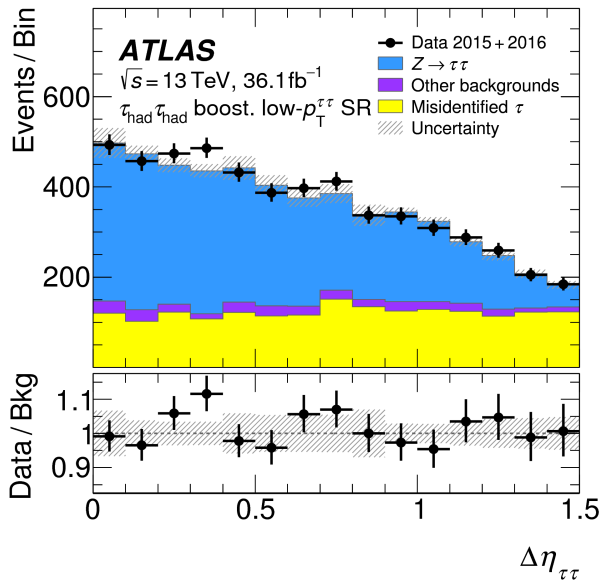


Figure 5.9: $\Delta\eta_{\tau\tau}$ distribution in the $\tau_{\text{had}}\tau_{\text{had}}$ boosted low $p_T^{\tau\tau}$ SR. Good agreement is shown between data and simulations. Due to the different shape of misidentified τ and the $Z \to \tau\tau$ backgrounds is possible to disentangle the two contributions[89].

systematic uncertainties are then included in the fit model as will be illustrated in details in Section 5.8.2.

5.7.1 Theoretical uncertainties on the signal

The procedures to estimate the theoretical uncertainties on the signal follow the recommendations of the LHC Higgs Cross Section Working Group (LHCHXSWG). All the theoretical uncertainties are factorized into uncertainties on the total cross section, on the acceptance in the SRs as well as on the shape of the $m_{\tau\tau}^{\text{MMC}}$. The uncertainty on the total cross section is treated as uncorrelated from the shape and acceptance uncertainties, while shape and acceptance uncertainties from a given source are treated as correlated. The relevant theoretical uncertainties on the SM predictions arise from three main sources:

- **QCD scale**: uncertainties due to missing higher order QCD corrections.
- **Parton Shower**: uncertainties due to the parton shower algorithm and to the non perturbative part of the simulation (underlying event and hadronization).
- **Parton Distribution Function and** α_s: uncertainties on the experimental input parameters used for the PDFs evaluation and on the α_s.

Uncertainties on the total cross section are given by the LHCXSWG [25] and reported in Table 5.8 and Table 5.9 for the QCD scale and PDF+α_s uncertainties respectively.

Production process	+ QCD scale (N^3LO)	- QCD scale (N^3LO)
ggF	+3.9 %	-3.9 %
VBF	+0.4 %	-0.3 %
WH	+0.5 %	-0.7 %
ZH	+3.8 %	-3.1 %

Table 5.8: Total cross section uncertainties due to missing higher order QCD corrections [25].

Production process	PDF	α_s	(PDF + α_s)
ggF	±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 5.9: Total cross section uncertainties due to PDFs and α_s [25].

In addition to the uncertainties that are used in the fit, additional studies were made at truth level, namely using samples where only the matrix element and the parton shower are simulated and not the full chain which includes simulation of the interaction with the detector and reconstruction of the physics objects. Altough the uncertainties evaluated at truth level are not used in the fit, they constituted the first evaluation of signal theory uncertainties and were used as a cross-check.

5.7.2 Truth level studies

Truth level studies have been performed using truth level samples produced with the same generators described in Section 5.2 and using Herwig7 as alternative parton shower algorithm for evaluating UEPS uncertainties. The uncertainties are evaluated in VBF-like and boosted-like regions as much as possible closer to the SRs used in the analysis.

VBF and VH scale uncertainties

As explained in Section 1.2, the expression for the QCD running coupling constant can be written as:

$$\alpha_s(\mu_R^2) = \frac{12\pi}{(11N_C - 2N_f)\ln(\frac{\mu_R^2}{\Lambda_{QCD}^2})} \tag{5.3}$$

Then a process at a certain order can be computed as a LO process with scale $\alpha_s(\mu_R^2)$ where all the perturbative series contributions are absorbed in $\alpha_s(\mu_R^2)$. Varying the renormalization scale μ_R gives an estimate on the contribution of missing higher order QCD corrections. In the same way the variation of the factorization scale in the DGLAP equations (Equation 1.17) quantifies the contribution of missing higher order in the PDFs.

The most conservative method to evaluate the impact of truncation of the perturbative series is then to vary the renormalization and factorization scales, by factors of 2 and 1/2 with the constraint $\frac{1}{2} \leq \frac{\mu_R}{\mu_f} \leq 2$ (the diagonal values are excluded since they represent corrections of higher order) and using $\mu_R = \mu_F = m_W$ as the nominal scale. These uncertainties have been evaluated using samples with event weights for each scale variation. The largest uncertainty is then kept as the final scale uncertainty. Table 5.10 shows the acceptance uncertainties values. VH scale variations are evaluated with the same method.

channel	boosted region	VBF region
$\tau_{\text{lep}}\tau_{\text{lep}}$	+0.604 %	+0.608 %
	-0.759 %	-0.790 %
$\tau_{\text{lep}}\tau_{\text{had}}$	+0.594 %	+0.490 %
	-0.746 %	-0.559 %
$\tau_{\text{had}}\tau_{\text{had}}$	+0.573 %	+0.407 %
	-0.583 %	-0.383 %

Table 5.10: Summary of scale uncertainties, evaluated at truth level, for the VBF process in the boosted and VBF signal regions for the three decay channels.

ggF scale uncertainties

The uncertainties due to missing higher order QCD corrections in the case of ggF process are evaluated in principle in the same way as for VBF and VH. However, experimental cuts defining exclusive regions in terms of jet multiplicity, generally introduces large logarithmic contributions, $\log p_{\text{T}}^{\text{cut}}/Q$ (Sudakov logarithms), where $Q \sim m_H$ for ggF, which lead to large uncertainties, absent in the inclusive cross section. Due to accidental cancellations, this uncertainty is often underestimated when varying the factorization and renormalization scales. Therefore jet bin migration uncertainties must be taken into account.
In this analysis, jet bin migration uncertainties have been estimated at truth level using the Stewart-Tackmann method [115]. The nominal ggF Higgs production samples are generated with dynamical factorization and renormalization scales both determined according to $\sqrt{(p_T^H)^2 + m_H^2}$. Jet bin uncertainties have been evaluated using the MCFM [116, 117, 118] program, which is a matrix element Monte Carlo generator. MCFM is able to calculate cross sections for the processes H+1jet and H+2jet both at LO and NLO, but it is not able to generate events.
When applying a kinematic cut that imposes a binning in the number of jets, an inclusive N-jet region is divided into an exclusive N-jet region and an inclusive N+1-jet region:

$$\sigma_{\geq N} = \sigma_N + \sigma_{\geq N+1} \tag{5.4}$$

where $\sigma_{\geq N}$ is the inclusive N-jet region, σ_N is the exclusive N-jet region and $\sigma_{\geq N+1}$

is the inclusive N+1-jet region.
In addition to the uncertainties on the absolute yield, an uncertainty on the jet bin migration applies leading to the following expressions for the total uncertainties in the N-jet exclusive region and in the N+1 jet inclusive region:

$$\begin{aligned} \Delta_N^2 &= (\Delta_N^Y)^2 + \Delta_{\text{cut}}^2 \\ \Delta_{\geq N+1}^2 &= (\Delta_{\geq N+1}^Y)^2 + \Delta_{\text{cut}}^2 \end{aligned} \tag{5.5}$$

where the Δ^Y are the uncertainties on the yield and Δ_{cut} are the uncertainties introduced by the cut applied. An uncertainty on the total cross section can be derived considering that the yield uncertainties are correlated across the two regions while the uncertainties introduced by the cut are totally anticorrelated and therefore they vanish when the two regions are added up:

$$\Delta_{\geq N}^2 = (\Delta_N^Y)^2 - (\Delta_{\geq N+1}^Y)^2 \tag{5.6}$$

It is then possible to derive the following cross section, since we are interested in the N-jet exclusive region,

$$\sigma_N = \sigma_{\geq N} - \sigma_{\geq N+1} \tag{5.7}$$

and its uncertainty,

$$\Delta_N^2 = \Delta_{\geq N}^2 + \Delta_{\geq N+1}^2 \tag{5.8}$$

In the analysis no explicit veto in the number of jets is imposed but an exclusive cut is applied requiring that the events that lie in the boosted category must fail the VBF selection, therefore removing events with two or more jets that have certain kinematic features. Figure 5.10 illustrates the regions necessary for the computation of jet bin uncertainties.

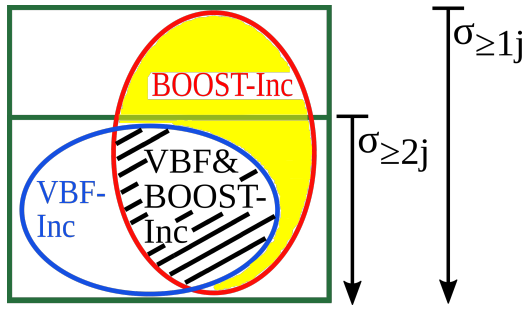


Figure 5.10: Illustration of jet multiplicity bins. "BOOST-Inc" refers to the inclusively defined boosted category, while BOOST&VBF-Inc is the overlap between the Boosted and VBF categories, which is inclusive in jet multiplicity. "VBF-Inc" is the inclusive VBF category. The boosted exclusive region is the set of events belonging to "BOOST-Inc" but not to "VBF-Inc", colored in yellow.

Similarly to equation 5.7 and 5.8 it is possible to define the exclusive boosted category as,

$$\sigma_{BOOST-Excl} = \sigma_{\geq 1j}^{\text{BOOST}-\text{Inc}} - \sigma_{\geq 2j}^{\text{BOOST\&VBF}-\text{Inc}} \tag{5.9}$$

and its uncertainty,

$$\Delta^2_{BOOST-Excl} = (\Delta_{\geq 1j}^{\text{BOOST}-\text{Inc}})^2 + (\Delta_{\geq 2j}^{\text{BOOST\&VBF}-\text{Inc}})^2 \tag{5.10}$$

The impact of uncertainties in these regions were simulated using the H+1jet process at NLO for the "BOOST-Inc" category and the H+2jet process at LO (for preserving the same order of α_s) for the "BOOST&VBF-Inc" category. The selections applied at truth level are reported in Table 5.11.

Channel	H+1j, BOOST-Inc, H+2j	BOOST&VBF-Inc
$\tau_{\text{lep}}\tau_{\text{lep}}$	$p_T(H) > 100$ GeV, ≥ 1 jet with $p_T(j_1) > 40$ GeV	BOOST-Inc cuts & ≥ 2 jets with $p_T(j_2) > 30$ GeV, $M(jj) > 400$ GeV, $\lvert\Delta\eta(jj) > 3\rvert$
$\tau_{\text{lep}}\tau_{\text{had}}$	$p_T(H) > 100$ GeV	BOOST-Inc cuts & ≥ 2 jets with $p_T(j_2) > 30$ GeV, $M(jj) > 300$ GeV, $\lvert\Delta\eta(jj) > 3\rvert$
$\tau_{\text{had}}\tau_{\text{had}}$	$p_T(H) > 100$ GeV, ≥ 1 jet with $p_T(j_1) > 70$ GeV	BOOST-Inc cuts & ≥ 2 jets with $p_T(j_2) > 30$ GeV, $M(jj) > 300$ GeV, $\lvert\Delta\eta(jj) > 3\rvert$

Table 5.11: Selection cuts applied for jet bin migration uncertainty at truth level.

The resulted uncertainties in the two regions, which are uncertainties on the cross section and not on the acceptance, are reported in Table 5.12.

Channel	BOOST-Inc	BOOST&VBF-Inc
$\tau_{\text{lep}}\tau_{\text{lep}}$	24.3%	59.7%
$\tau_{\text{lep}}\tau_{\text{had}}$	24.1%	59.6%
$\tau_{\text{had}}\tau_{\text{had}}$	23.3%	59.6%

Table 5.12: Jet bin migration uncertainties on the cross section evaluated with the Stewart-Tackmann method.

The uncertainties are large in particular for the Boosted&VBF-Inc category because the process is evaluated at LO. The two uncertainties are not simply added in quadrature but in principle they are associated to different parameters in the fit (see Section 5.8.2 for systematic uncertainties treatment in the fit). Considering the two parameters as log-normal distributed and considering the relations between the jet bins, they can be

expressed by the following formulas:

$$\begin{aligned} &\exp\left[{\epsilon_{BOOST-Inc}}^{(1/f_{BOOST-Excl})}\right] \\ &\exp\left[{\epsilon_{BOOST\&VBF-Inc}}^{(-f_{BOOST\&VBF-Inc}/f_{BOOST-Excl})}\right] \end{aligned} \tag{5.11}$$

where ϵ are the gaussian distributed uncertainties and f indicates the fraction of events in a certain category. Thus, even though the uncertainties in the BOOST&VBF-Inc are large, their impact is small because the fraction of events is small. These uncertainties and the related parameters in the fit are however not used in the analysis since they have been replaced by the method described in Section 5.7.3.

Parton Shower uncertainties

Parton shower programs simulate the process from the hard scattering to the scale where perturbative QCD breaks down and hadronization happens ($Q \sim 1$ GeV). Therefore different sources of uncertainties are related to the parton shower program used, from the matching to the matrix element, to the simulation of the showering and to the non perturbative effects like simulation of hadronization and underlying event.
The most conservative way to estimate this uncertainty is to compare different parton shower algorithms. In the analysis the default Pythia8 showering software was replaced by Herwig7. The showering is simulated as a Markovian process where the parton splitting is ordered according to an kinematic variable. In the case of Pythia the ordering is based on the transverse momentum of the daughter partons relative to the direction of the mother while Herwig uses an angular ordered shower where the angle between the daughter partons is considered. In addition, different hadronization algorithms are employed: Pythia uses the Lund String Model while Herwig uses the Cluster Model [119]. Finally, different underlying event modeling are embedded in the two generators.
The result for this truth level comparison are reported in Table 5.13 and Table 5.14 for the VBF and ggF samples respectively.

channel	boosted region	VBF region
$\tau_{\text{lep}}\tau_{\text{lep}}$	3.36%	-8.38%
$\tau_{\text{lep}}\tau_{\text{had}}$	1.51%	-8.31%
$\tau_{\text{had}}\tau_{\text{had}}$	-0.002%	-9.40%

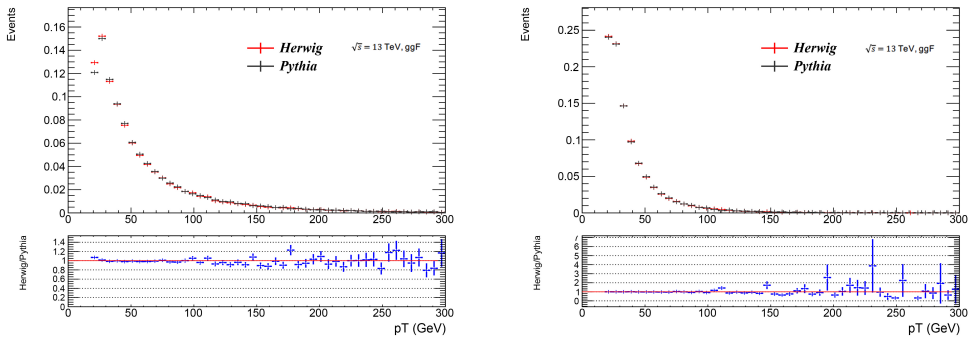
Table 5.13: Parton shower uncertainties derived from Herwig to Pythia comparison at truth level for the VBF process.

The differences between Herwig and Pythia can be seen also in the distribution of the jet related kinematic variables as it is shown in Figure 5.11 for the ggF sample.

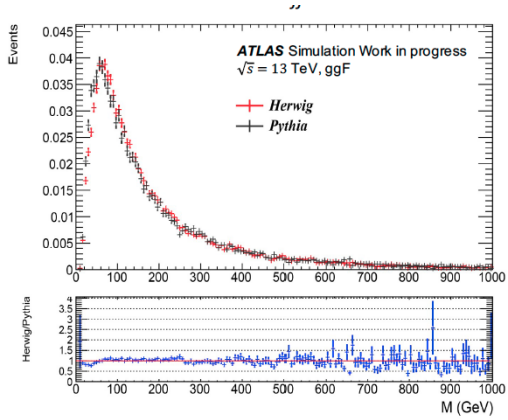
In addition to the comparison of different shower algorithms, which is rather conservative and can overestimate the uncertainty, ATLAS has developed tunes which allows to vary internal Pythia parameters. The AZNLO tune [100] has been used to estimate uncertainties related to internal Pythia parameters variations. The parameters refer to

channel	boosted region	VBF region
$\tau_{\text{lep}}\tau_{\text{lep}}$	-15.28%	18.49%
$\tau_{\text{lep}}\tau_{\text{had}}$	-14.80%	14.56%
$\tau_{\text{had}}\tau_{\text{had}}$	-15.48%	14.10%

Table 5.14: Parton shower uncertainties derived from HERWIG to PYTHIA comparison at truth level for the *gg*F process.



(a) (b)



(c)

Figure 5.11: Transverse momentum distribution of the highest-p_{T} jet (a), the second highest-p_{T} jet (b) and the mass of the two jets system (c) simulated with the two different parton shower programs: HERWIG and PYTHIA [120].

multi parton interaction, to final state radiation scale variations and to the primordial k_T of initiators of hard-scattering which covers initial state radiation uncertainties. The resulted uncertainties obtained by summing in quadrature each individual parameter variation contribution are shown in Table 5.15 and Table 5.16 for *gg*F and VBF processes

respectively.

Channel	boosted region	VBF region
$\tau_{\text{lep}}\tau_{\text{lep}}$	2.83%	11.10%
$\tau_{\text{lep}}\tau_{\text{had}}$	2.87%	12.05%
$\tau_{\text{had}}\tau_{\text{had}}$	2.27%	13.54%

Table 5.15: Summary of the total parton shower uncertainties, obtained with the AZNLO tune by summing the individual contributions in quadrature, for the ggF process.

Channel	boosted region	VBF region
$\tau_{\text{lep}}\tau_{\text{lep}}$	1.40%	1.34%
$\tau_{\text{lep}}\tau_{\text{had}}$	1.79%	2.32%
$\tau_{\text{had}}\tau_{\text{had}}$	2.23%	2.05%

Table 5.16: Summary of the total parton shower uncertainties, obtained with the AZNLO tune by summing the individual contributions in quadrature, for the VBF process.

Since also the central sample has an impact on the uncertainty, an alternative tune (A14 [121]) has been used to evaluate the uncertainties but since they are smaller than the other parton shower uncertainties, they were used only as a cross-check. Table 5.17 and Table 5.18 shows the results for the ggF and VBF processes respectively.

Channel	boosted region	VBF region
$\tau_{\text{lep}}\tau_{\text{lep}}$	0.10%	0.76%
$\tau_{\text{lep}}\tau_{\text{had}}$	0.22%	5.17%
$\tau_{\text{had}}\tau_{\text{had}}$	1.95%	5.67%

Table 5.17: Parton shower uncertainties from the AZNLO/A14 comparison for the ggF process.

Channel	boosted region	VBF region
$\tau_{\text{lep}}\tau_{\text{lep}}$	0.79%	1.58%
$\tau_{\text{lep}}\tau_{\text{had}}$	0.70%	0.005%
$\tau_{\text{had}}\tau_{\text{had}}$	0.75%	1.17%

Table 5.18: Parton shower uncertainties from the AZNLO/A14 comparison for the VBF process.

PDF uncertainties

In this Section the method, used for initially evaluating the PDF uncertainties, is described. Later, due to the availability of dedicated MC samples, it was possible to use a different recommendation illustrated in Section 5.7.3. However this initial approach is explained here because of its pedagogical value and because its applicable to all MC generators.
The PDF uncertainties at truth level have been estimated using the the internal errors available with the set used for the nominal samples, the CT10 [122] and using alternative sets, namely MSTW2008 [123] and NNPDF 2.3 [103].
The CT10 error band is calculated using the Hessian method. Each of the 26 eigenvectors provided by the CT10 PDF set, can be varied up and down. The total up variation is then determined as:

$$\Delta X^{+}_{\max} = \sqrt{\sum_{i=1}^{N} [\max\left(X_i^{+} - X_0, X_i^{-} - X_0, 0\right)]^2} \tag{5.12}$$

and the total down variation as:

$$\Delta X^{-}_{\max} = \sqrt{\sum_{i=1}^{N} [\max\left(X_0 - X_i^{+}, X_0 - X_i^{-}, 0\right)]^2}, \tag{5.13}$$

where X_0 is the nominal value of a variable, while X_i^{+} is its value using the PDF corresponding to the up variation of the eigenvector i and X_i^{-} is its value using the PDF corresponding to the down variation of the eigenvector i. Thus, $\Delta X^{+}_{\max}$ considers the errors that increase the nominal value while $\Delta X^{-}_{\max}$ considers those that decrease the nominal value. The final PDF uncertainty is compute taking the envelope between the CT10 error band and the difference with the alternative PDF sets. The final PDF uncertainty are listed in Table 5.19 and Table 5.20 for the ggF and VBF processes respectively.

Channel	boosted region		VBF region	
$\tau_{\text{lep}}\tau_{\text{lep}}$	+1.83%	-1.48%	+1.34%	-1.11%
$\tau_{\text{lep}}\tau_{\text{had}}$	+2.03%	-1.68%	+1.36%	-1.10%
$\tau_{\text{had}}\tau_{\text{had}}$	+1.87%	-1.55%	+2.73%	-2.32%

Table 5.19: PDF uncertainties, evaluated at truth level, for the ggF process.

5.7.3 Uncertainties at reconstruction level

The uncertainties calculated at truth level have been replaced by reconstruction level uncertainties, using full simulated samples and evaluating them in the actual SRs of the analysis. This has been possible due to the availability of new MC samples which generally contain specific event weights which allow to propagate the uncertainties and

Channel	boosted region		VBF region	
$\tau_{\text{lep}}\tau_{\text{lep}}$	+1.02%	-1.19%	+1.56%	-1.60%
$\tau_{\text{lep}}\tau_{\text{had}}$	+1.11%	-1.19%	+1.91%	-1.62%
$\tau_{\text{had}}\tau_{\text{had}}$	+0.89%	-0.85%	+2.15%	-1.85%

Table 5.20: PDF uncertainties, evaluated at truth level, for the VBF process.

to estimate them in each SR employed in the analysis. How they have been computed will be described in this Section.

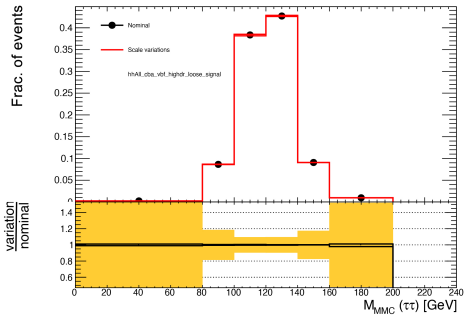
VBF and VH scale uncertainties

VBF and VH scale variations have been evaluated with the same procedure as for truth level uncertainties, described in Section 5.7.2, with the only difference that the event weights are embedded in the full simulated samples. The resulted uncertainties, obtained by varying the renormalization and factorization scales, are reported in Table 5.21 for the VBF process.

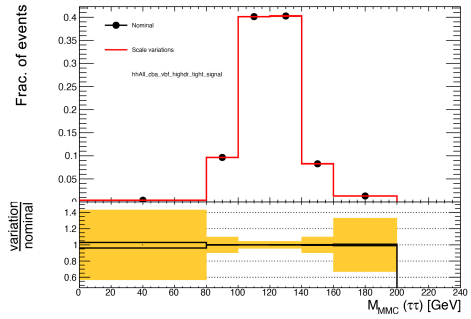
Channel	Signal region	maximum up variation	maximum down variation
$\tau_{\text{lep}}\tau_{\text{lep}}$	VBF loose	0.77 % ± 0.01 %	-0.84 % ± 0.02 %
	VBF tight	1.21,% ± 0.02 %	-1.60 % ± 0.02 %
	Boosted low-$p_T^{\tau\tau}$	0.554 % ± 0.009 %	-0.65 % ± 0.01 %
	Boosted high-$p_T^{\tau\tau}$	1.42 % ± 0.03 %	-0.80 % ± 0.01 %
$\tau_{\text{lep}}\tau_{\text{had}}$	μh VBF loose	0.351 % ± 0.008 %	-0.221 % ± 0.005 %
	eh VBF loose	0.67 % ± 0.02 %	-0.702 % ± 0.02 %
	μh VBF tight	1.02 % ± 0.01 %	-1.04 % ± 0.01 %
	eh VBF tight	1.08 % ± 0.01 %	-1.21 % ± 0.02 %
	μh Boosted low-$p_T^{\tau\tau}$	0.63 % ± 0.01 %	-0.62 % ± 0.01 %
	eh Boosted low-$p_T^{\tau\tau}$	0.65 % ± 0.01 %	-0.65 % ± 0.01 %
	μh Boosted high-$p_T^{\tau\tau}$	1.48 % ± 0.03 %	-0.81 % ± 0.01 %
	eh Boosted high-$p_T^{\tau\tau}$	1.35 % ± 0.03 %	-1.04 % ± 0.02 %
$\tau_{\text{had}}\tau_{\text{had}}$	VBF high-$p_T^{\tau\tau}$	0.7372 % ± 0.0004 %	-1.366 % ± 0.0004 %
	VBF loose	1.75 % ± 0.07 %	-2.22 % ± 0.09 %
	VBF tight	0.423 % ± 0.008 %	-0.56 % ± 0.01 %
	Boosted low-$p_T^{\tau\tau}$	0.83 % ± 0.01 %	-1.02 % ± 0.02,%
	Boosted high-$p_T^{\tau\tau}$	1.59 % ± 0.03 %	-2.11 % ± 0.04 %

Table 5.21: QCD scale uncertainties in the various signal regions for the VBF process.

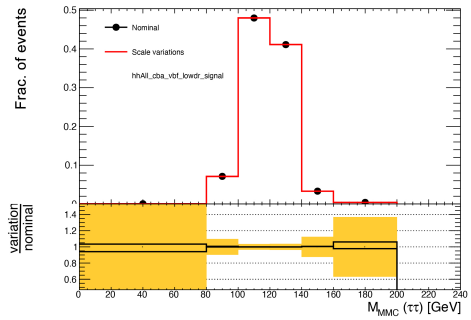
In addition to the acceptance uncertainties, shape uncertainties on the fit variable $m_{\tau\tau}^{\text{MMC}}$ are taken into account. In the case of VBF and VH scale variations they are all negligible. Figure 5.12 shows an example for VBF regions in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.



(a) VBF loose



(b) VBF tight



(c) VBF high-$p_T^{\tau\tau}$

Figure 5.12: Effect of QCD scale uncertainties on the $m_{\tau\tau}^{\mathrm{MMC}}$ for the VBF process in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ VBF signal regions: loose (a), tight (b) and high-$p_T^{\tau\tau}$ (c).The yellow band indicates Monte Carlo statistical uncertainties.

ggF scale uncertainties

The ggF scale uncertainties computed with the Stewart Tackmann method at truth level and using the MCFM program, have been replaced by a method recommended by the LHCHXSWG [25]. The uncertainties due to the QCD higher order corrections are split into nine sources of uncertainties considered as uncorrelated. These sources of uncertainties, which take into account also the effects of jet bin migration, are the following:

- Scale variations:
  - $\Delta\mu$: factorization and renormalization scale variations.
  - $\Delta\phi$: resummation scale variation.
  - $\Delta_{\mathrm{cut}}^{0/1}(\Delta_{\mathrm{cut}}^{1/2})$: jet bin migration uncertainty between the 0 and 1-jet bins (and the 1 and 2-jet bins), introduced by the binning cut.
- VBF topology uncertainties:

  - Variation of the VBF phase space.
  - Third jet veto.

- Higgs p_{T}-shape uncertainties:
  - Higgs p_{T} shape dependence within $0 - 60/60 - \infty$ GeV.
  - Higgs p_{T} shape dependence within $0 - 120/120 - \infty$ GeV.
- Top mass dependence.

The scale variations are established using the method described by Booughezal, Liu, Petriello, Stewart and Tackmann in Yellow Report 4 [25]. This method utilizes the framework of soft-collinear effective theory [124] for jet-veto resummation which allows to control different sources of yield and migration uncertainties. It allows to separate uncertainties due to higher fixed-order corrections and uncertainties due to higher order p_{T} logarithms arising from jet binning. The two VBF topology uncertainties are derived using Yellow Report 3 methods [125] based on the Stewart-Tackmann method. The Higgs p_{T} shape variations are taken from QCD scale variations of Powheg NNLOPS. Finally, the last uncertainty related to the top mass is derived from difference between LO and NLO rescaling.
In order to be determined at reconstruction level, the variations are propagated as event weights to the ggF samples and the uncertainties are then evaluated in each SR of the analysis. In contrast to other signal theory uncertainties, no factorization in total cross section uncertainty and shape/acceptance uncertainty is foreseen in this case. The results for these uncertainties are reported in Tables 5.22 5.23 5.24 for the various channels while in Figure 5.13 are visualized the effects of uncertainties on the $m_{\tau\tau}^{\mathrm{MMC}}$ distribution for the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel (as example). All the systematics variations are rather small compared to the statistical uncertainties. The uncertainties on the total cross section are disentangled from that on the acceptance/shape with a dedicated procedure, before performing the fit.

Variation	VBF tight	VBF loose	Boosted high-$p_T^{\tau\tau}$	Boosted low-$p_T^{\tau\tau}$
$\Delta\mu$	(4.3 ± 0.2) %	(3.6 ± 0.1) %	(7.40 ± 0.08) %	(6.84 ± 0.08) %
$\Delta\phi$	(4.3 ± 0.2) %	(3.5 ± 0.1) %	(7.15 ± 0.08) %	(6.48 ± 0.07) %
$\Delta_{\text{cut}}^{0/1}$	(2.6 ± 0.1) %	(2.5 ± 0.1) %	(5.59 ± 0.06) %	(6.06 ± 0.07) %
$\Delta_{\text{cut}}^{1/2}$	(7.4 ± 0.3) %	(5.3 ± 0.2) %	(8.9 ± 0.1) %	(5.26 ± 0.06) %
VBF phase space	(9.6 ± 0.4) %	(10.8 ± 0.4) %	(0.399 ± 0.005) %	(0.493 ± 0.02) %
VBF w/ 3rd jet veto	(0.45 ± 0.02) %	(-0.83 ± 0.03) %	(0.0757 ± 0.0009) %	(0.0297 ± 0.0003) %
Higgs p_{T} 0-60/60-∞ GeV	(7.9 ± 0.4) %	(7.3 ± 0.3) %	(9.7 ± 0.1) %	(7.36 ± 0.08) %
Higgs p_{T} 0-120/120-∞ GeV	(11.0 ± 0.5) %	(10.0 ± 0.4) %	(13.8 ± 0.2) %	(7.34 ± 0.08) %
top mass dependence	(6.4 ± 0.3) %	(3.1 ± 0.1) %	(7.46 ± 0.08) %	(0.335 ± 0.004) %

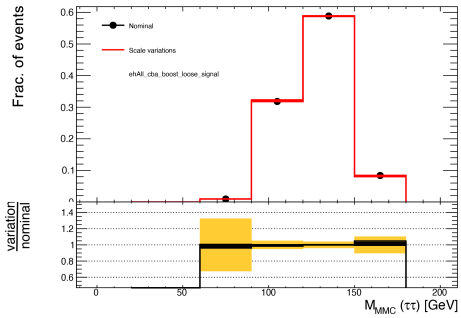
Table 5.22: QCD scale uncertainties in the $\tau_{\text{lep}}\tau_{\text{lep}}$ SRs. Each of the nine sources of uncertainties is shown.

Variation	$\mu\tau_{\text{had}}$ VBF tight	$\mu\tau_{\text{had}}$ VBF loose	μh Boosted high-$p_T^{\tau\tau}$	μh Boosted low-$p_T^{\tau\tau}$
$\Delta\mu$	(3.8 ± 0.2) %	(2.8 ± 0.2) %	(7.5 ±0.2) %	(6.8 ± 0.1) %
$\Delta\phi$	(3.7 ± 0.2) %	(2.6 ± 0.2) %	(7.2 ± 0.2) %	(6.4 ± 0.1) %
$\Delta_{\text{cut}}^{0/1}$	(2.6 ± 0.2) %	(2.6 ± 0.2) %	(5.6 ± 0.1) %	(6.1 ± 0.1) %
$\Delta_{\text{cut}}^{1/2}$	(5.6 ± 0.4) %	(1.8 ± 0.2) %	(9.1 ± 0.2) %	(5.1 ± 0.1) %
VBF phase space	(10.4 ± 0.7) %	(11 ± 1) %	(0.222 ± 0.005) %	(0.55 ± 0.01) %
VBF w/ 3rd jet veto	(-0.27 ± 0.02) %	(-2.7 ± 0.2) %	(0.069 ± 0.002) %	(0.0311 ± 0.0007) %
Higgs p_{T} 0-60/60-∞ GeV	(7.5 ± 0.5) %	(3.1± 0.3) %	(9.7 ± 0.2) %	(7.16 ± 0.2) %
Higgs p_{T} 0-120/120-∞ GeV	(10.7 ± 0.7) %	(3.4 ± 0.3) %	(13.8 ± 0.3) %	(6.8 ± 0.1) %
top mass dependence	(5.4 ± 0.4) %	(1.02 ± 0.09) %	(8.2 ± 0.2) %	(0.73 ± 0.02) %
Variation	eh VBF tight	$e\tau_{\text{had}}$ VBF loose	$e\tau_{\text{had}}$ Boosted high-$p_T^{\tau\tau}$	μh Boosted low-$p_T^{\tau\tau}$
$\Delta\mu$	(4.9 ± 0.3) %	(2.9 ± 0.3) %	(7.5 ±0.2) %	(6.8 ± 0.1) %
$\Delta\phi$	(4.8 ± 0.3) %	(2.7 ± 0.3) %	(7.2 ± 0.2) %	(6.4 ± 0.1) %
$\Delta_{\text{cut}}^{0/1}$	(3.1 ± 0.5) %	(2.8 ± 0.3) %	(5.6 ± 0.1) %	(6.1 ± 0.1) %
$\Delta_{\text{cut}}^{1/2}$	(8.1 ± 0.5) %	(1.4 ± 0.1) %	(9.2 ± 0.2) %	(4.8 ± 0.1) %
VBF phase space	(8.0 ± 0.5) %	(11 ± 1) %	(0.259 ± 0.006) %	(0.51 ± 0.01) %
VBF w/ 3rd jet veto	(1.37 ± 0.08) %	(0.25 ± 0.03) %	(0.192 ± 0.004) %	(0.0144 ± 0.0003) %
Higgs p_{T} 0-60/60-∞ GeV	(8.2 ± 0.5) %	(3.0 ± 0.3) %	(9.7 ± 0.2) %	(7.2 ± 0.2) %
Higgs p_{T} 0-120/120-∞ GeV	(11.1 ± 0.7) %	(3.4 ± 0.3) %	(13.8 ± 0.3) %	(6.9 ± 0.1) %
top mass dependence	(6.9 ± 0.4) % ± 0.0 %	(0.85 ± 0.09) %	(7.9 ± 0.2) %	(0.94 ± 0.02) %

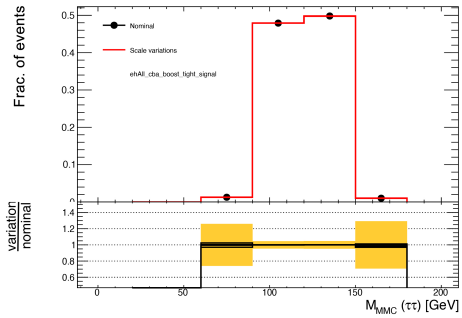
Table 5.23: QCD scale uncertainties in the $\tau_{\text{lep}}\tau_{\text{had}}$ SRs. Each of the nine sources of uncertainties is shown.

Variation	VBF loose	VBF tight	VBF high-$p_T^{\tau\tau}$	Boosted high-$p_T^{\tau\tau}$	Boosted low-$p_T^{\tau\tau}$
$\Delta\mu$	(2.2 ± 0.2) %	(1.3 ± 0.1) %	(5.2 ± 0.3) %	(6.6 ± 0.1) %	(7.4 ± 0.1) %
$\Delta\phi$	(2.1 ± 0.2) %	(1.3 ± 0.1) %	(5.2 ± 0.3) %	(6.2 ± 0.1) %	(7.1 ± 0.1) %
$\Delta_{\text{cut}}^{0/1}$	(1.8 ± 0.2) %	(1.3 ± 0.1) %	(2.8 ± 0.2) %	(6.3 ± 0.1) %	(5.6 ± 0.1) %
$\Delta_{\text{cut}}^{1/2}$	(2.2 ± 0.3) %	(0.57 ± 0.06) %	(10.1 ± 0.6) %	(3.75 ± 0.07) %	(8.7 ± 0.2) %
VBF phase space	(14 ± 2) %	(16 ± 2) %	(8.0 ± 0.5) %	(0.55 ± 0.01) %	(0.343 ± 0.006) %
VBF w/ 3rd jet veto	(-4.6 ± 0.5) %	(-0.35 ± 0.04) %	(0.47 ± 0.03) %	(0.157 ± 0.003) %	(0.101 ± 0.002) %
Higgs p_{T} 0-60/60-∞ GeV	(4.2 ± 0.5) %	(2.7 ± 0.3) %	(9.5 ± 0.6) %	(7.2 ± 0.1) %	(9.7 ± 0.2) %
Higgs p_{T} 0-120/120-∞ GeV	(6.0 ± 0.7) %	(3.6 ± 0.4) %	(13.8 ± 0.8) %	(6.2 ± 0.1) %	(13.8 ± 0.2) %
top mass dependence	(0.10 ± 0.01) %	(0.10 ± 0.01) %	(6.1 ± 0.4) %	(0.098 ± 0.002) %	(5.8 ± 0.1) %

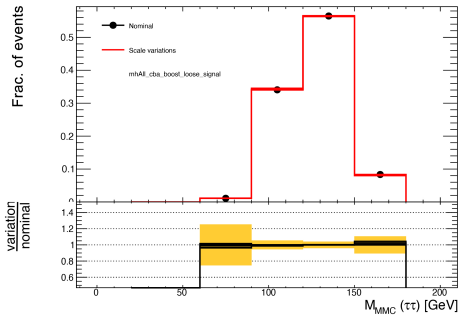
Table 5.24: QCD scale uncertainties in the $\tau_{\text{lep}}\tau_{\text{had}}$ SRs. Each of the nine sources of uncertainties is shown.



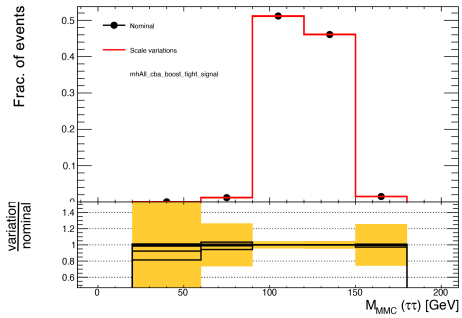
(a) $e\tau_{\text{had}}$ boosted low-$p_T^{\tau\tau}$



(b) $e\tau_{\text{had}}$ boosted high-$p_T^{\tau\tau}$



(c) $\mu\tau_{\text{had}}$ boosted low-$p_T^{\tau\tau}$



(d) $\mu\tau_{\text{had}}$ boosted high-$p_T^{\tau\tau}$

Figure 5.13: Effect of QCD scale uncertainties on the $m_{\tau\tau}^{\text{MMC}}$ for the ggF process in the $\tau_{\text{lep}}\tau_{\text{had}}$ boosted SRs (splitted by lepton flavour): $e\tau_{\text{had}}$ boosted low-$p_T^{\tau\tau}$ (a), $e\tau_{\text{had}}$ boosted high-$p_T^{\tau\tau}$ (b), $\mu\tau_{\text{had}}$ boosted low-$p_T^{\tau\tau}$ (c) and $\mu\tau_{\text{had}}$ boosted high-$p_T^{\tau\tau}$ (d). The yellow band indicates the Monte Carlo statistical uncertainties.

Parton shower uncertainties

Parton shower uncertainties at reconstruction level are evaluated only taking into account the difference between two different parton shower algorithms, namely Pythia8 and Herwig7, since, as truth level studies have shown, this is the largest uncertainty and overlaps with the uncertainties derived from the AZNLO tune. The comparison has been performed using fast simulation samples [126] for both Pythia and Herwig. Powheg is able to generate events at NNLO but since the NNLO weights are not propagated to Herwig, the comparison has been done at NLO for being consistent. The uncertainties are then evaluated in each SR as the difference between the two MC programs and then applied to the nominal Pythia sample. The results for acceptance uncertainties are shown in Table 5.25 and 5.26 for ggF and VBF production respectively. Uncertainties are quite large in particular for ggF samples in VBF regions and for VBF samples in boosted regions. Figure 5.14 5.15 shows also the impact of uncertainties on the $m_{\tau\tau}^{\text{MMC}}$

Channel	Signal region	Variation
$\tau_{\text{lep}}\tau_{\text{lep}}$	VBF loose	23 % ± 4 %
	VBF tight	26 % ± 4 %
	Boosted low-$p_T^{\tau\tau}$	-8.8 % ± 0.4 %
	Boosted high-$p_T^{\tau\tau}$	-4.8 % ± 0.2 %
$\tau_{\text{lep}}\tau_{\text{had}}$	VBF loose	18 % ± 3 %
	VBF tight	-13 % ± 2 %
	Boosted low-$p_T^{\tau\tau}$	-5.1 % ± 0.3 %
	Boosted high-$p_T^{\tau\tau}$	-24 % ± 1 %
$\tau_{\text{had}}\tau_{\text{had}}$	VBF high-$p_T^{\tau\tau}$	1.8 % ± 0.4 %
	VBF loose	9 % ± 3 %
	VBF tight	9 % ± 2 %
	Boosted low-$p_T^{\tau\tau}$	9.2 % ± 0.7 %
	Boosted high-$p_T^{\tau\tau}$	4.7 % ± 0.3 %

Table 5.25: Parton shower acceptance uncertainties for ggF production in the various SRs.

distribution. The observed differences are quite large in some cases but almost always within the statistical uncertainties (they suffer from lack of Monte Carlo statistics).

PDF uncertainties

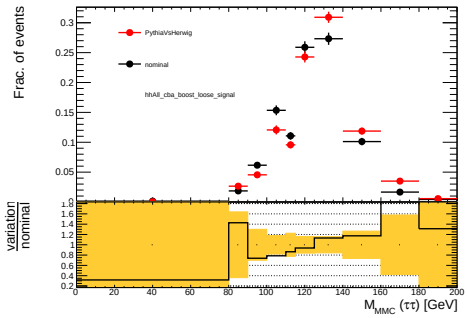
PDF uncertainties evaluated at truth level have been replaced by the PDF uncertainties evaluated at reconstruction level with the most recent reccomendations from the PDF4LHC collaboration [99]. A single PDF set which is a combination of a suitable subset of the available PDF sets is used along with 30 eigenvector variations and 2 variations of α_s. All the PDF and α_s uncertainties are treated in the fit model as uncorrelated nuisance parameters. In Tables 5.27 and 5.28 are quoted only the maximum up and down variation for VBF and ggF production respectively.

Channel	Signal region	Variation
$\tau_{\text{lep}}\tau_{\text{lep}}$	VBF loose	1.32 % ± 0.06 %
	VBF tight	0.65 % ± 0.02 %
	Boosted low-$p_T^{\tau\tau}$	4.3,% ± 0.2 %
	Boosted high-$p_T^{\tau\tau}$	9.9 % ± 0.3 %
$\tau_{\text{lep}}\tau_{\text{had}}$	VBF loose	5.5 % ± 0.3 %
	VBF tight	10.7 % ± 0.3 %
	Boosted low-$p_T^{\tau\tau}$	7.9 % ± 0.3 %
	Boosted high-$p_T^{\tau\tau}$	5.4 % ± 0.2 %
$\tau_{\text{had}}\tau_{\text{had}}$	VBF high-$p_T^{\tau\tau}$	9.2 % ± 0.7 %
	VBF loose	-10 % ± 2 %
	VBF tight	2.7 % ± 0.3%
	Boosted low-$p_T^{\tau\tau}$	18 % ± 2 %
	Boosted high-$p_T^{\tau\tau}$	2.2 % ± 0.2 %

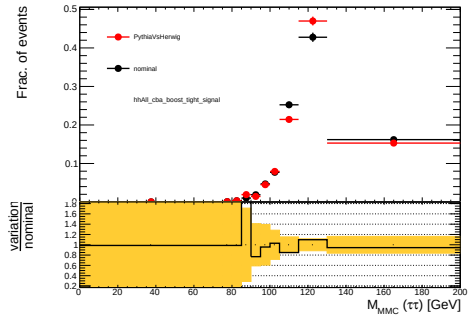
Table 5.26: Parton shower acceptance uncertainties for VBF production in the various SRs.

Channel	Signal region	maximum up variation	maximum down variation
$\tau_{\text{lep}}\tau_{\text{lep}}$	VBF loose	0.117 % ± 0.002 %	-0.297 % ± 0.006 %
	VBF tight	0.386,% ± 0.005 %	-0.491 % ± 0.007 %
	Boosted low-$p_T^{\tau\tau}$	0.116 % ± 0.002 %	-0.271 % ± 0.004 %
	Boosted high-$p_T^{\tau\tau}$	0.301 % ± 0.006 %	-0.339 % ± 0.006 %
$\tau_{\text{lep}}\tau_{\text{had}}$	$\mu\tau_{\text{had}}$ VBF loose	0.132 % ± 0.003 %	-0.229 % ± 0.006 %
	$e\tau_{\text{had}}$ VBF loose	0.145 % ± 0.004 %	-0.149 % ± 0.004 %
	$\mu\tau_{\text{had}}$ VBF tight	0.308 % ± 0.004 %	-0.393 % ± 0.005 %
	$e\tau_{\text{had}}$ VBF tight	0.294 % ± 0.004 %	-0.356 % ± 0.005 %
	$\mu\tau_{\text{had}}$ Boosted loose	0.163 % ± 0.003 %	-0.183 % ± 0.003 %
	$e\tau_{\text{had}}$ Boosted low-$p_T^{\tau\tau}$	0.086 % ± 0.002,%	-0.247 % ± 0.005 %
	$\mu\tau_{\text{had}}$ Boosted tight	0.275 % ± 0.005 %	-0.487 % ± 0.009 %
	$e\tau_{\text{had}}$ Boosted high-$p_T^{\tau\tau}$	0.320 % ± 0.006 %	-0.358 % ± 0.007 %
$\tau_{\text{had}}\tau_{\text{had}}$	VBF high-$p_T^{\tau\tau}$	0.91 % ± 0.02 %	-0.93 % ± 0.02 %
	VBF loose	1.33 % ± 0.06 %	-1.26 % ± 0.06 %
	VBF tight	1.13 % ± 0.02 %	-1.12 % ± 0.02,%
	Boosted low-$p_T^{\tau\tau}$	1.44 % ± 0.03 %	-1.01 % ± 0.02 %
	Boosted high-$p_T^{\tau\tau}$	0.97 % ± 0.02 %	-0.98 % ± 0.02 %

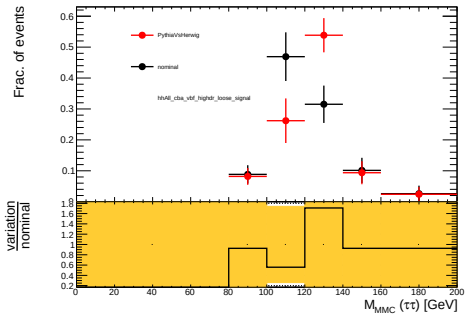
Table 5.27: PDF and α_s uncertainties for VBF production.



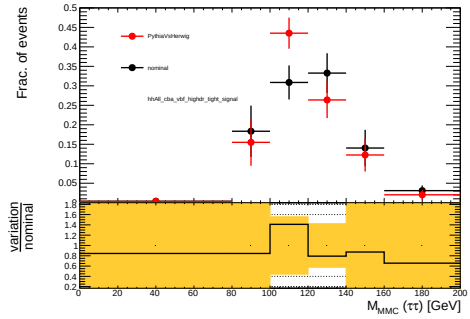
(a) Boosted low-$p_T^{\tau\tau}$



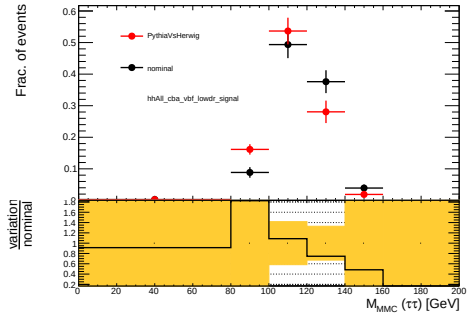
(b) Boosted high-$p_T^{\tau\tau}$



(c) VBF loose

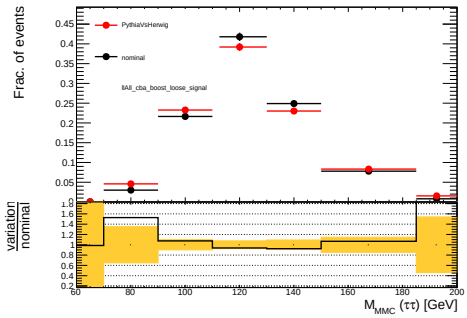


(d) VBF tight

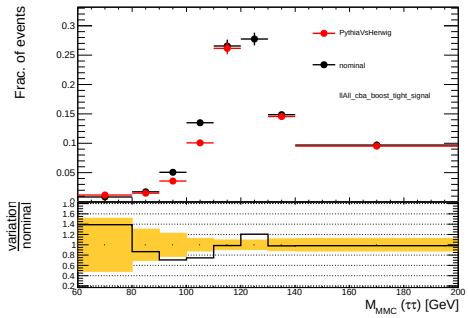


(e) VBF high-$p_T^{\tau\tau}$

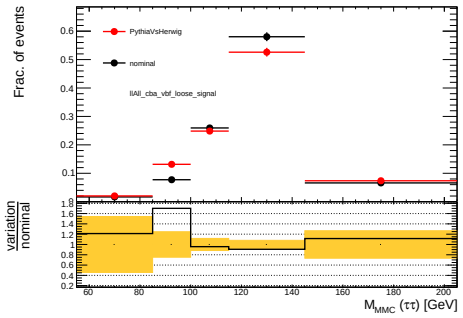
Figure 5.14: Dependence of parton shower uncertainties for ggF production on the $m_{\tau\tau}^{\text{MMC}}$ in the $\tau_{\text{had}}\tau_{\text{had}}$ SRs: boosted low-$p_T^{\tau\tau}$ (a), boosted high-$p_T^{\tau\tau}$ (b), VBF loose (c), VBF tight (d), VBF high-$p_T^{\tau\tau}$ (e). The red markers represent the uncertainty derived from the ratio Pythia/Herwig and applied to the nominal sample, and the black ones the nominal sample (Pythia). The yellow band indicates Monte Carlo statistical uncertainties of the nPythia sample produced in Fast Simulation and used for the evaluation of uncertainties.



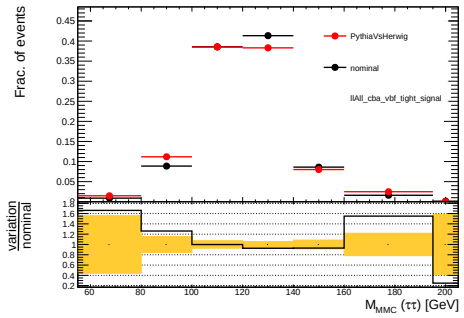
(a) Boosted low-$p_T^{\tau\tau}$



(b) Boosted high-$p_T^{\tau\tau}$



(c) VBF loose



(d) VBF tight

Figure 5.15: Dependence of parton shower uncertainties for VBF production on the $m_{\tau\tau}^{\mathrm{MMC}}$ in the $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ SRs: boosted low-$p_T^{\tau\tau}$ (a), boosted high-$p_T^{\tau\tau}$ (b), VBF loose (c), VBF tight (d). The red markers represent the uncertainty derived from the ratio PYTHIA/HERWIG and applied to the nominal sample, and the black ones the nominal sample (PYTHIA).The yellow band indicates Monte Carlo statistical uncertainties of the nPYTHIA sample produced in Fast Simulation and used for the evaluation of uncertainties.

5.7.4 Theoretical uncertainties on the backgrounds

Theoretical uncertainties are also considered for the simulated dominant $Z \to \tau\tau$ background. The uncertainties taken into account are due to missing higher order QCD corrections, PDF parametrization and parton shower and underlying event modelling. In particular the following sources of uncertainties have been evaluated:

- $\mu_{\mathbf{R}}/\mu_{\mathbf{F}}$: variations of renormalization and factorization scales. It has been performed using event weights provided by SHERPA.

- **CKKW**: uncertainty due to the matching between matrix element and parton shower. It has been evaluated using truth level parametrisation as a function of jet multiplicity and transverse momentum of the Z boson.

Channel	Signal region	maximum up variation	maximum down variation
$\tau_{\text{lep}}\tau_{\text{lep}}$	VBF loose	1.47 % ± 0.06 %	-1.32 % ± 0.05 %
	VBF tight	1.64% ± 0.08 %	-2.00 % ± 0.09 %
	Boosted low-$p_T^{\tau\tau}$	1.54 % ± 0.02,%	-1.36 % ± 0.02 %
	Boosted high-$p_T^{\tau\tau}$	1.49 % ± 0.02 %	-1.88% ± 0.02 %
$\tau_{\text{lep}}\tau_{\text{had}}$	$\mu\tau_{\text{had}}$ VBF loose	1.5 % ± 0.1 %	-1.3 % ± 0.1 %
	$e\tau_{\text{had}}$ VBF loose	1.3 % ± 0.1 %	-1.2 % ± 0.1 %
	$\mu\tau_{\text{had}}$ VBF tight	1.7 % ± 0.1 %	-1.9 % ± 0.1 %
	$e\tau_{\text{had}}$ VBF tight	1.6 % ± 0.1 %	-1.8 % ± 0.1 %
	$\mu\tau_{\text{had}}$ Boosted low-$p_T^{\tau\tau}$	1.52 % ± 0.03 %	-1.381 % ± 0.03 %
	$e\tau_{\text{had}}$ Boosted low-$p_T^{\tau\tau}$	1.46 % ± 0.03 %	-1.33 % ± 0.03 %
	$\mu\tau_{\text{had}}$ Boosted high-$p_T^{\tau\tau}$	1.49 % ± 0.03 %	-1.91 % ± 0.04 %
	$e\tau_{\text{had}}$ Boosted high-$p_T^{\tau\tau}$	1.45 % ± 0.03 %	-1.79% ± 0.04 %
$\tau_{\text{had}}\tau_{\text{had}}$			
	VBF high-$p_T^{\tau\tau}$	5.5 % ± 0.3 %	-5.1 % ± 0.3 %
	VBF loose	5.6 % ± 0.6 %	-5.2 % ± 0.6%
	VBF tight	5.5 % ± 0.6 %	-5.1 % ± 0.5%
	Boosted low-$p_T^{\tau\tau}$	5.5 % ± 0.1 %	-5.1 % ± 0.1 %
	Boosted high-$p_T^{\tau\tau}$	5.5 % ± 0.1 %	-5.01 % ± 0.09 %

Table 5.28: PDF and α_s uncertainties for ggF production.

- **QSF**: resummation scale uncertainty. It has been evaluated using truth level parametrisation as a function of jet multiplicity and transverse momentum of the Z boson.
- **PDF**: uncertainty due to PDF parametrization. It has been evaluated using event weight provided by SHERPA.
- **Parton Shower**: this uncertainty has been evaluated comparing two different MC generators, namely comparing the nominal sample, generated with SHERPA to an alternative sample, generated with MADGRAPH5_aMC@NLO.
- **Underlying event**: in addition to the parton shower uncertainty, specific underlying event uncertainties have been evaluated using different setups of multi parton interactions but they have found to be negligible.

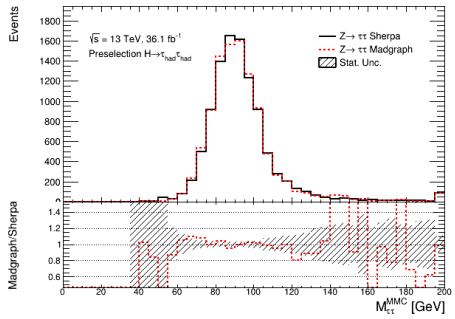
The normalization of $Z \to \tau\tau$ background is directly retrieved from fit to data. Two different parameters are used in order to constrain separately the normalization in the VBF and boosted SRs respectively. The $Z \to \tau\tau$ modelling systematics are therefore not allowed to change the overall yield within an inclusive region. Shape uncertainties on the $m_{\tau\tau}^{\text{MMC}}$ are considered as well as event migration between subregions: this is obtained by scaling the uncertainty in a specific region by the impact of the variation in the respective inclusive region. Additional uncertainties are considered for the variation in acceptance

between the analysis channels and this is evaluated as the impact of the uncertainty on the relative normalization of each channel with respect to the total expectation regardless of the decay channel.
The parton shower uncertainties derived by comparing SHERPA and MADGRAPH5_aMC@NLO partly overlap with the othe considered systematics, such as the ones due to the matching between matrix element and parton shower and the resummation scale uncertainties. However, this is the most conservative available way to evaluate them. Due to the limited statistics, the $Z \to \tau\tau$ parton shower uncertainty is evaluated in the preselection region and then extrapolated to the SRs. This is done by computing the bin by bin ratio from MADGRAPH to SHERPA comparison and then applying it to the nominal SHERPA distribution in each SR. Figure 5.16 shows the MADGRAPH and SHERPA distributions of $Z \to \tau\tau$ in the preselection, VBF inclusive and boosted inclusive regions for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The $m_{\tau\tau}^{\text{MMC}}$ shape variation is similar between the preselection and boosted region while this is a bit harder statement to made for the VBF region. Here, the statistics decreases significantly for the MADGRAPH5_aMC@NLO sample leading to large fluctuations. Anyway, considering also the MADGRAPH statistics, it has been decided to keep the variations resulting from the preselection region.

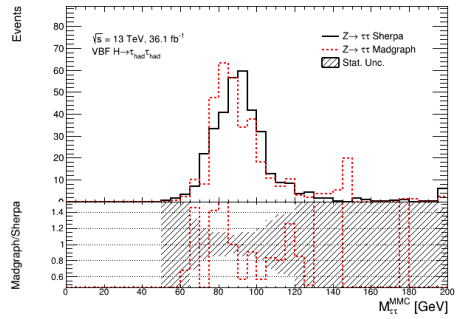
For all simulated backgrounds, other than $Z \to \tau\tau$, no theoretical uncertainties are considered as they were found to be rather small compared to the uncertainties of the dominant backgrounds. In addition, all the other major sources of background as top-quark in the $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels and $Z \to ll$ in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel, are constrained by CRs yields while the other minor backgrounds are normalized to their theoretical prediction.

5.7.5 Experimental uncertainties

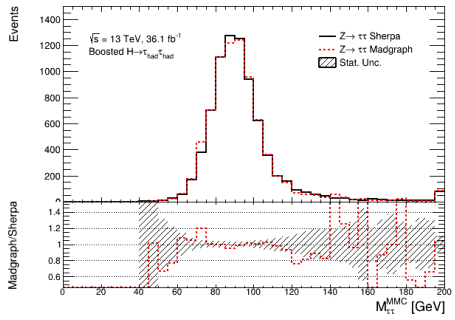
Various experimental systematic uncertainties are considered in the analysis. They are related to trigger, object reconstruction, identification and also energy scale and resolution of the phyisics objects used: electrons, muons, $\tau_{\text{had-vis}}$, jets and $E_{\text{T}}^{\text{miss}}$. In addition, uncertainties on the luminosity and pile-up are taken into account.
Electrons and muons related uncertainties are rather small. In the case of electrons, the uncertainties on the energy scale are obtained from variation in the scale of electron momentum. Trigger, identification, reconstruction and isolation efficiencies uncertainties are evaluated varying the errors on the trigger, identification, reconstruction and isolation scale factors respectively. An uncertainty on the electron momentum resolution is considered as well.
For what concerns the muons, an uncertainty on the muon energy scale is considered and calculated from variations in the scale of the muon momentum. Uncertainties on the muon momentum resolution are evaluated considering variations of the ID and MS tracks. Finally, uncertainties on the trigger, identification, isolation and track to vertex association efficiency are calculated using variations on the errors of the respective scale factors.
Uncertainties related to the $\tau_{\text{had-vis}}$ are significant in this analysis and range from 2% to 14% depending on the uncertainty source. The uncertainties on the $\tau_{\text{had-vis}}$ energy scale



(a) Preselection



(b) VBF inclusive



(c) Boosted inclusive

Figure 5.16: $m_{\tau\tau}^{\mathrm{MMC}}$ distribution for $Z \to \tau\tau$ process in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ preselection (a), VBF inclusive (b) and boosted inclusive (c) regions, simulated with SHERPA and MADGRAPH5_aMC@NLO (they are normalized to the same area). The bottom panel shows the ratio of MADGRAPH5_aMC@NLO to SHERPA, where the dashed band is the statistical uncertainty of the nominal sample (SHERPA).

are determined by fitting the Z boson mass using $Z \to \tau\tau$ events, reconstructed using the visible τ decay products. Three indipendent sources of uncertainties are considered: uncertainty related to the modeling of the detector geometry, due to the measurement in a tag-and-probe analysis and due to the Monte Carlo simulation (GEANT4). In addition, trigger, identification and reconstruction efficiency uncertainties are considered as well as an uncertainty related to the overlap removal with the electrons which affects the identification efficiency. These uncertainties have been evaluated using parameter variations in the MC simulations both related to the theoretical simulation and to the interaction with the detector and calorimeter response. Since most systematics uncertainties are determined by the studies of $Z \to \tau\tau$ events, whose cross section drops significantly above the Z pole, in some cases additional uncertainties referred to high p_{T} $\tau_{\text{had-vis}}$ ($p_{\mathrm{T}} > 50$ GeV) are considered.

Jet related uncertainties are another important source of systematic uncertainties in the

analysis. They are determined by comparing data with MC simulations and they are dependent on the transverse momentum as well as on the pseudorapidity of the reconstructed jet. Eleven independent sources of uncertainties are considered for the jet energy resolution systematic (they correspond to different η regions). Jet energy scale uncertainties are taken into account as well and factorized into eight principal components. Additional uncertainties, considered in the analysis, are related to pile-up corrections, b-tagging efficiencies, which are measured in dedicated calibration analysis, η dependence and high-p_{T} jets. Finally, an uncertainty associated to the jet vertex tagger algorithm is considered. These uncertainties amount to 1-6% where the major contributions are from some components of the jet energy resolution and scale uncertainties.
Uncertainties related to the missing transverse momentum are due to the soft term in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, because the uncertainties on the hard terms are already included in the systematics related to the other physics objects mentioned. Uncertainties on the missing transverse momentum scale and resolution are derived by comparing data to Monte Carlo simulations.
Uncertainties on the luminosity affect the number of predicted signal and background events. The uncertainty amounts to 2.1% for the 2015+2016 dataset and it is derived using both the LUCID-2 [127] detector measurements and the track-counting methodology described in Ref. [44].
Monte Carlo simulations are performed using a generalized profile distribution of the number of interactions per bunch crossing. Therefore, the simulated events need to be reweighted to match the observed distribution in 2015 and 2016 data. For the best agreement between data and MC a correction factor of 1.16 need to be applied. A systematic uncertainty associated to this correction factor is then defined.
Examples on how experimental uncertainties affect the $m_{\tau\tau}^{\mathrm{MMC}}$ shape as well as the normalization in the SRs are shown in Figure 5.17 for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. The choice of uncertainties, regions and processes is just intended to be an example among the most relevant sources of uncertainty.

5.8 Statistical analysis

The analysis results (and this applies to the measurement that will be illustrated in Chapter 6 as well) are obtained through likelihood based statistical tests. Therefore, before going into the specific details of this measurement (Section 5.8.2), a more general introduction to the statistical methods employed, is given in Section 5.8.1.

5.8.1 Statistical framework

Likelihood-based statistical test

The search for new phenomena in high energy physics is generally performed in the context of frequentist statistical tests [128]. The general approach is to define a null hypothesis, H_0, describing only the known processes and test it against the alternative signal hypothesis, H_1, which includes both signal and background. The agreement between the observed data with a given hypothesisis is then quantified by computing a p-value,

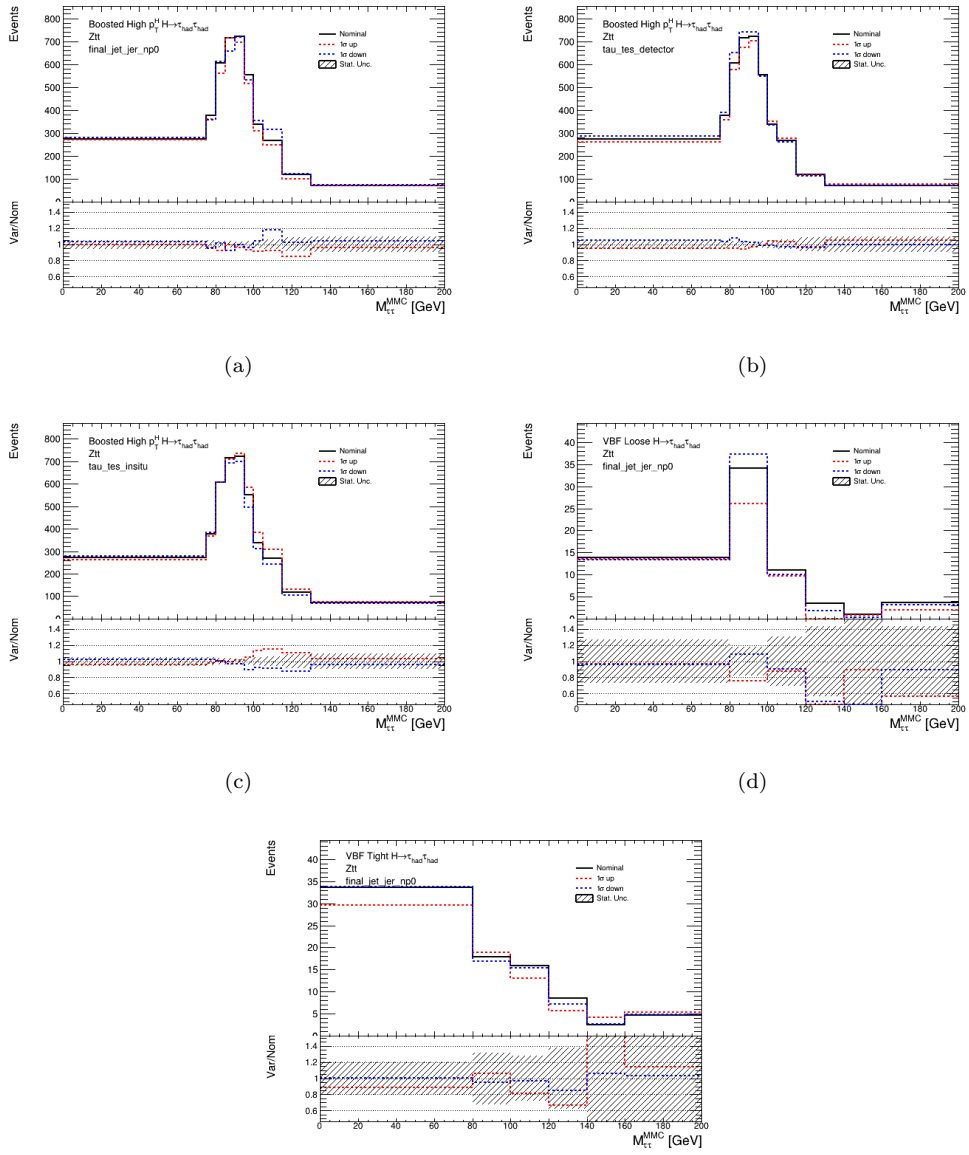


Figure 5.17: Example of experimental systematic uncertainties in the $\tau_{\text{had}}\tau_{\text{had}}$ channel: the leading jet energy resolution uncertainty component for $Z \to \tau\tau$ process in the boosted high-$p_{\text{T}}^{\tau\tau}$ region (a), tau energy scale uncertainty related to the modeling of detector geometry for $Z \to \tau\tau$ process in the boosted high-$p_{\text{T}}^{\tau\tau}$ region (b), tau energy scale uncertainty related to the tag and probe measurement in the boosted high-$p_{\text{T}}^{\tau\tau}$ (c), one of the jet energy resolution uncertainty components for $Z \to \tau\tau$ process in the VBF loose category (d) and the leading jet energy resolution uncertainty component for the misidentified τ process in the VBF category (e).

namely the probability, when the null hypothesis is true, to obtain a result of equal or greater incompatibility with respect to what is observed. For example, this measure of incompatibility, can be based on the number of events found in specific regions for certain distributions. The significance is often used instead of the p-value. It is defined as the number of equivalent standard deviations of a gaussian distribution when the integral of the upper tail is equal to the p-value:

$$\begin{aligned} Z &= \Phi^{-1}(1-p) \\ \Phi(x) &= \int_{-\infty}^{x} dx' \frac{1}{\sqrt{2\pi}} e^{-(x'^2/2)} \end{aligned} \tag{5.14}$$

where Φ is the cumulative distribution of the standard gaussian and Φ^{-1} is the quantile, namely the inverse of the cumulative distribution of the standard gaussian.
Usually, in high energy physics, an evidence is claimed with $Z = 3$ corresponding to a p-value, $p = 1.35 \times 10^{-3}$, while a discovery is claimed only when reaching at least $Z = 5$ which corresponds to $p = 2.87 \times 10^{-7}$ (considering one-sided interval as will be discussed later in this Section).
The procedure often used as a test statistic in high energy physics, is the profile likelihood ratio test and according to the Neyman-Pearson lemma [129], this is the test statistic providing the best discriminating power. In particle physics, the events are usually classified in different channels and for each channel the observed number of events (n_i) is compared to the expected number of signal (s_i) and background events (b_i). For instance, consider an experiment where a certain observable has been measured and the resulted data can be represented as histograms. Suppose for each event one measures a variable x in order to build the histogram $\mathbf{n} = (n_1, ..., n_N)$. For each histrogram bin the expectation value can be written as:

$$E[n_i] = \mu s_i + b_i \tag{5.15}$$

where s_i and b_i are the mean number of entries of signal and background respectively and they can be expressed as:

$$\begin{aligned} s_i &= s_{\text{tot}} \int_i f_s(x; \vec{\theta}_s) dx \\ b_i &= b_{\text{tot}} \int_i f_b(x; \vec{\theta}_s) dx \end{aligned} \tag{5.16}$$

The parameter μ in Equation 5.15 represents the strength of the signal process: $\mu = 0$ for the background-only hypothesis and $\mu = 1$ for the expected signal hypothesis. The functions f_s and f_b in Equation 5.16 are the probability density functions (pdfs) of the variable x for the signal and background respectively. $\vec{\theta}_s$ and $\vec{\theta}_b$ are further parameters that can change the shape of the pdfs. Finally, s_{tot} and b_{tot} are the total mean numbers of signal and background events.
The μ parameter is generally refered to as the "parameter of interest" because it is the parameter that we want to measure and can be retrieved from the fit to data. In addition,

systematic uncertainties must always be taken into account in an experimental measurement and therefore they are included in the set of θ parameters, which are generally a function of $\vec{\theta} = (\vec{\theta}_s, \vec{\theta}_b, b_{\text{tot}})$ and are known as nuisance parameters (NPs). The values of NPs are not simply known a priori but they can be derived from the fit to data as well. However, systematic uncertainties are generally constrainted by subsidiary measurements (or they are derived from calculations such as in the case of theoretical uncertainties).
The likelihood function is then the product of Poisson probabilities for all the bins describing the observed number of events and the pdfs (f_θ) representing the systematic uncertainties:

$$L(\mu, \theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} f_{\theta,k}(\tilde{\theta}_k, \theta_k) \tag{5.17}$$

One can say that the likelihood is the joint pdf for the observed data, even though the data are fixed and the likelihood is a function of μ and θ. The $\tilde{\theta}$ variable in Equation 5.17 is the value provided by the subsidiary measurement around which the θ parameter can be varied. Different distributions can be used for the term accounting for systematic uncertainties and the specific case of this measurement will be described in Section 5.8.2. The additional flexibility given by adding more degrees of freedom introduced for parametrizing systematic uncertainties, results, as expected, in a loss in sensitivity.
In order to test an hypothesized value of μ we consider the following profile likelihood ratio:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})} \tag{5.18}$$

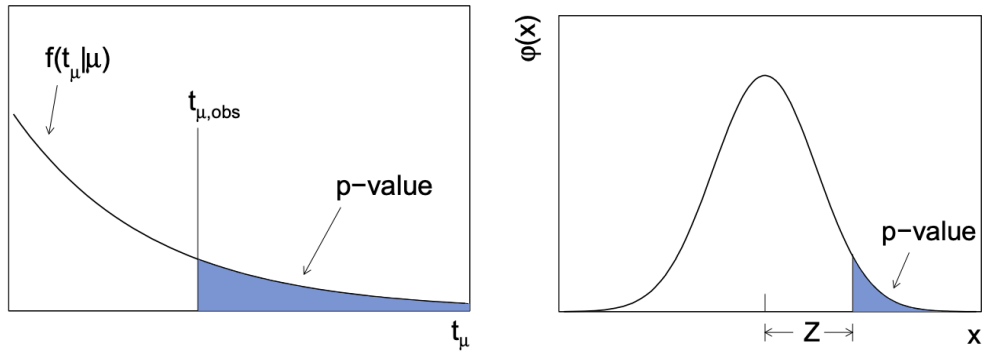
Here $\hat{\hat{\theta}}$ in the numerator represents the value of θ that maximizes the likelihood for the value of μ that is being tested, namely it is the conditional maximum likelihood (ML) estimator of θ. In the denominator, both μ and θ are varied and $\hat{\mu}$ and $\hat{\theta}$ represent the values that maximize (unconditionally) the likelihood function, therefore they are the ML estimators of μ and θ. The presence of nuisance parameters broadens the profile likelihood, thus reflecting the loss of information due to systematic uncertainties.
The value of $\lambda(\mu)$ is in the range $0 \leq \lambda \leq 1$ as it can be seen from Equation 5.18 where the closer the value is to 1, the higher is the agreement between data and the hypothesized value of μ. Instead of using $\lambda(\mu)$ it is convenient to use the statistic:

$$t_\mu = -2 \ln \lambda(\mu) \tag{5.19}$$

The maximization of the profile likelihood is then converted into a minimization and the logarithm transforms the product into sum of terms. t_μ can be used directly to measure the discrepancy between data and the hypothesis, considering that higher values of t_μ correspond to higher disagreement. This disagreement can be quantified by the p-value:

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu) dt_\mu \tag{5.20}$$

where t_μ is the value of the test statistic obtained from the data and $f(t_\mu|\mu)$ is the pdf of t_μ assuming the hypothesized value for μ. This p-value represents then the probability for the test statistic to be larger (or lower) than the observed value, under the assumption of a certain μ value (signal plus background hypothesis). The p-value is maximum for $\hat{\mu} = \mu$, and decreases as μ takes values larger or smaller than $\hat{\mu}$. Thus, one may obtain a two-sided interval for μ and if interested in only one of the directions, a one-sided interval can be determined. Figure 5.18 shows a general illustration of the relation between test statistic and p-value and significance and p-value.



(a) (b)

Figure 5.18: Relation between the p-value obtained from the observed value of the test statistic (a) and a gaussian distribution ($\phi(x)$) showing the relation between the significance (Z) and the p-value.

If the signal can only increase the event rate beyond what is expected for the background, the test statistic is further modified. This implies having $\mu \geq 0$ and in case a value of $\hat{\mu} < 0$ is found, this means that the best agreement between data and μ is the signal free hypothesis, $\mu = 0$. The test statistic can then be rewritten as:

$$\tilde{t_\mu} = -2\ln\tilde{\lambda}(\mu) = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & \hat{\mu} > 0 \end{cases} \tag{5.21}$$

The disagreement between data and the hypothesized μ value can be computed in the same way as for the case aforementioned (using the p-value from Equation 5.20). The estimated $\hat{\mu}$ value can be greater or less than the μ value being tested in this case as well, thus obtaining either a two-sided or one-sided interval.
Another important test statistic employed in high energy physics, assumes $\mu \geq 0$ but it is used to test the $\mu = 0$ hypothesis, hence the purpose is to reject the background-only hypothesis. The Equation 5.21 can then be rewritten for the case $\mu = 0$:

$$q_0 = \begin{cases} -2\ln\lambda\mu = 0 & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \tag{5.22}$$

In this particular test statistic the disagreement between data and background-only hypothesis is considered only when $\hat{\mu} = 0$. In case the expected data do not show any deviation from the background prediction or they are even smaller, the test statistics gives $q_0 = 0$ corresponding to the case where the data are described by the background prediction. As the event yield increase over the expected background, the disagreement between data and the $\mu = 0$ hypothesis increases. In order to quantify the disagreement a p-value for this case can be defined:

$$p_0 = \int_{q_0,obs}^{\infty} f(q_0|0)dq_0 \tag{5.23}$$

where $f(q_0|0)$ is the pdf of the test statistic q_0 under the assumption of the background-only hypothesis ($\mu = 0$).

Limits setting

In the case of no significant excess is found, limits on the signal stregth can be derived, using the definition of CLs [130]:

$$\mathrm{CL}_s(\mu) = \frac{p_\mu}{1 - p_b} \tag{5.24}$$

If a $\mathrm{CL}_s \leq \alpha$ is found, assuming the signal plus background hypothesis, then this hypothesis is excluded with $(1 - \alpha)$ confidence level (C.L.). The upper limit on μ, determined with $(1-\alpha)$ C.L., correspond to the μ value that leads to $\mathrm{CL}_s = \alpha$. Conventionally, the signal plus background hypothesis can be excluded with $\mathrm{CL}_s < 0.05$, therefore with 95% C.L..

Asimov data set and expected significance

In order to investigate the potential and to optimise the sensitivity of an analysis, an expected significance is defined. Furthermore analysis are often "blinded" before the whole analysis strategy is finalized, meaning that it is not possible to look at data distributions in order not to be biased by them.
For quantifiying the expected sensitivity, the idea is to build an artificial dataset called the Asimov dataset. This dataset is defined such that, when using it for evaluating the estimators of the parameters, one obtains their "true" values. In particular, the ML estimators of μ and θ parameters assume the values predicted by the signal plus background hypothesis, namely $\mu = 1$ and the input values for the nuisance parameters. The Asimov dataset represents then the model prediction assuming the presence of a signal.
From the Asimov dataset one can extract the expected significance, defined as the median of the inverse cumulative gaussian distribution, and, under the asymptotic approximation, its expression is [128]:

$$Z_{\mathrm{exp}} = \sqrt{-2ln\lambda(\hat{\mu}_0, \hat{\theta}_0)} \tag{5.25}$$

where $\hat{\mu}_0$ and $\hat{\theta}_0$ are the ML estimators of μ and the θ for the Asimov dataset.

5.8.2 $H \to \tau\tau$ fit model

The $H \to \tau\tau$ fit model is based on the statistical framework already described in Section 5.8.1. The specific details of this analysis will be illustrated in this section.

Likelihood function and statistical test

The statistical analysis employs a binned likelihood function built as the product of the Poisson probability terms for each histogram bin in each signal and control regions and the term accounting for the systematic uncertainties. Therefore, the likelihood is the same as Equation 5.17 where the product also runs over all the regions. The SRs and CRs used in the fit are the same described in Section 5.5.2. In summary, the following regions are employed: one high-$p_{\mathrm{T}}^{\tau\tau}$ and one low-$p_{\mathrm{T}}^{\tau\tau}$ boosted region for each channel, one VBF tight and one VBF loose region for each channel and one VBF high-$p_{\mathrm{T}}^{\tau\tau}$ region for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel and in addition one VBF and one boosted $Z \to ll$ control regions for $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ and one VBF and one boosted top control region for $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$. In the SRs the binned $m_{\tau\tau}^{\mathrm{MMC}}$ distribution is employed in the fit (Figure 5.3 and Figure 5.4 shows the exact binning used in the fit for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel) while CRs are only used to constrain the normalization of backgrounds, therefore they are represented by one-bin histogram containing the event yield. All the main Higgs production modes, ggF, VBF, VH and $t\bar{t}H$, are considered as signal and the relative systematic uncertainties are used separately in the fit. The most significant backgrounds, $Z \to \tau\tau$ and the misidentified τ, are also considered separately. In addition, $Z \to ll$ and top-quark backgrounds are considered as additional separate backgrounds in the $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels. Minor backgrounds as diboson production, W/Z bosons electroweak production are considered together in addition to $Z \to ll$ and top-quark for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel.
All the other parameters beside the parameter of interest are included in the fit as nuisance parameters. These can refer to systematic uncertainties, normalization factors for backgrounds, which are left free to float in the fit, as well as bin by bin statistical uncertainties which are also encoded as nuisance parameters. In this analysis, the pdf $f_\theta(\tilde{\theta}, \theta)$ in Equation 5.17 assumes different expressions according to the type of nuisance parameter:

- A gaussian distribution is the default choice for parametrizing systematic uncertainties. The θ parameter is varied around $\tilde{\theta}$ in the likelihood maximization. The pdf is hence given as:

$$f_\theta(\tilde{\theta}, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta - \tilde{\theta})^2}{2\sigma^2}\right) \tag{5.26}$$

- The gaussian distribution is not suitable for the nuisance parameters that have to be positive. Therefore, for all the normalization factors the log-normal distribution is used instead:

$$f_\theta(\tilde{\theta}, \theta) = \frac{1}{\sqrt{2\pi}\ln(k)} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln k)^2}\right) \frac{1}{\theta} \tag{5.27}$$

where k is the width of the log-normal pdf, meaning that the observable has 68% probability to be in the range $\hat{\theta}/k < \theta < \hat{\theta}k$. For small uncertainties, the Gaussian with a relative uncertainty ϵ and the log-normal with $k = 1 + \epsilon$ are asimptotically identical.

- The gamma distribution is used for bin-by-bin statistical uncertainties:

$$f_\gamma(n) = \alpha \frac{(n\alpha)^N}{N!} \exp(-n\alpha) \tag{5.28}$$

where the event yield N is decomposed in the nominal yield n and a factor α parametrizing the relative statistical uncertainty. The parameter α can be further decomposed into a variable which accounts for the statistical uncertainty δ and a parameter γ which reflects the fact that the predicted yield n can be varied by a certain amount. The γ parameter is then extracted from the fit. In the analysis only one nuisance parameter per bin is associated to the total estimate and the total statistical uncertainty in that bin [131].

The systematic uncertainties can affect both the shape and the normalization of the observable. Usually two variations are provided for each systematic: one corresponding to the $+1\sigma$ variation of the nuisance parameter and the other one corresponding to the -1σ variation. Indeed the uncertainties are extracted by varying the nuisance parameters within $\pm\sigma$ and then repeating the fit.
Some background processes, as already mentioned, are left free to float in the fit. This concerns all the background processes associated with dedicated CRs with one normalization factor (NF) associated to each region, namely one VBF and one boosted top NF for $\tau_{\text{lep}}\tau_{\text{lep}}$, one VBF and one boosted $Z \to ll$ NF for $\tau_{\text{lep}}\tau_{\text{lep}}$, one VBF and one boosted top NF for $\tau_{\text{lep}}\tau_{\text{had}}$. In addition, the $Z \to \tau\tau$ and the misidentified τ backgrounds have also normalizations which are retrieved from the fit. Two NFs are associated to the $Z \to \tau\tau$ process, one correlated across all the VBF regions and one correlated across all the boosted regions. One NF is used instead for the misidentified τ background, correlated across all the $\tau_{\text{had}}\tau_{\text{had}}$ regions, while for the other two channels the normalization of this background is fixed before performing the fit. A summary of the $H \to \tau\tau$ fit model is shown in Figure 5.19.

The fit model described so far has a quite high complexity due to the large number of regions and especially to the very large number of nuisance parameters. Moreover, many shape systematic variations are dominated by statistical noise due to the lack of statistics of some samples in certain regions. Therefore in order to suppress the statistical noise, avoiding to remove genuine shape variations, and reducing the complexity of the fit, the following procedures have been applied:

- **Symmetrisation**: in the case of shape systematics, it may happen to have the $+1\sigma$ and -1σ variation lying in the same direction for certain histograms bins (one-sided bin). This can introduce discontinuity in the fit and the presence of double minima for the ML estimators in the likelihood minimization. For each one-sided bin the largest variation is kept and symmetrized in the opposite direction.

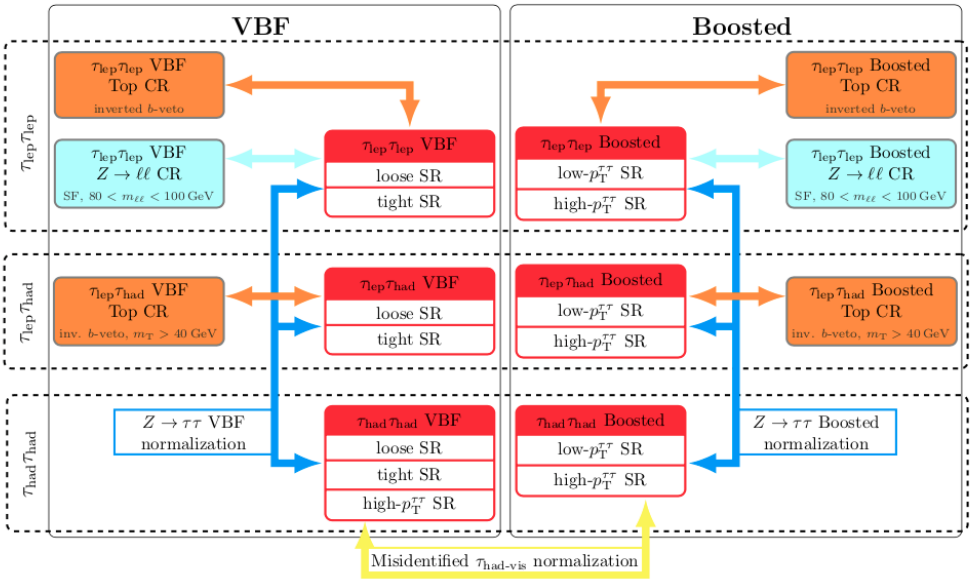
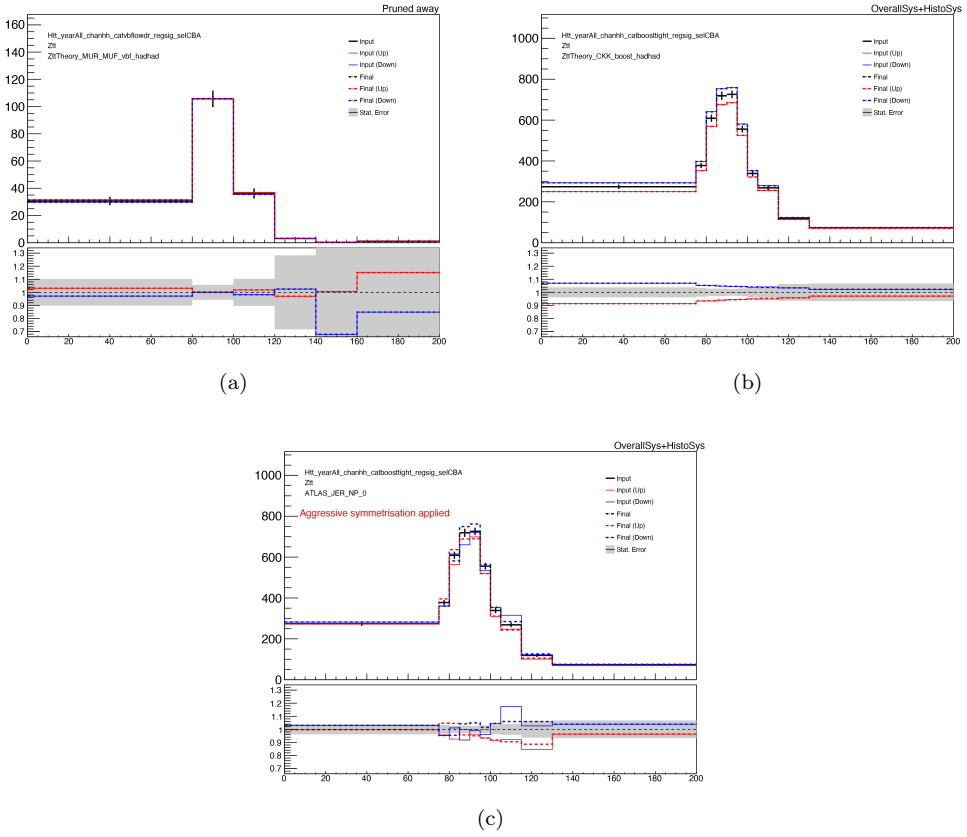


Figure 5.19: Summary of the fit model used in the analysis. All the signal and control regions used in the fit are shown. Different colors correspond to different background components with a separate normalization factor.

- **Pruning**: some nuisance parameters have a completely negligible effect on the observed distribution and can be dropped simplifying the fit without any impact on the parameter estimation. NPs are therefore considered only if they fulfill some requirements. The exact tresholds have been optimised such that only negligible uncertainties are removed.
  Firstly all normalization variations are considered if their impact is larger than 0.5% or smaller than the statistical uncertainty of the respective sample.
  Then, several criteria are applied to the shape components of the systematics. A large statistical uncertainty can cause artificial large shape variation due to statistical fluctuation. Therefore shape systematics are pruned away when the total statistical uncertainty is greater than 10%.
  A further step is performed using a χ^2 test between the $+1\sigma$ and the -1σ variations with respect to the nominal and considering as statistical uncertainty only the largest between the nominal and the variation (since they are generally correlated). If the χ^2 result is lower than a given treshold (p-value< 0.1), the uncertainty is not considered significant and it is pruned away.
  Lastly a bin-by-bin significance is calculated as $S_i = |u_i - d_i|/\sigma_i^{tot}$ where u_i (d_i) are the $+1\sigma$ and the -1σ event yield in the bin i respectively and σ_i^{tot} is the total statistical uncertainty for all background components. If the significance is lower than a given treshold ($S_i < 0.1$) for all bins, than the shape systematic is considered non-significant.

- **Smoothing**: to further reduce statistical noise a smoothing procedure is employed. This procedure relies on a quadratic extrapolation and it is applied to the ratio of the varied histogram over the nominal histograms [132]. Applying it to the ratio and not to the shape variation itself ensure that only statistical effects are treated and no physical effect, like the presence of a peak, is smoothed.

An example of the procedures aforementioned is given in Figure 5.20 where the $m_{\tau\tau}^{\mathrm{MMC}}$ distributions for some systematic uncertainties is shown before and after applying these procedures. Figure 5.20 shows different examples: a systematic uncertainty considered negligible, one accepted and one where the symmetrisation procedure has been applied.



(a) (b)

(c)

Figure 5.20: Examples of simplifications procedures applied before performing the fit: $Z \to \tau\tau$ scale uncertainty in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ VBF high-p_{T} region considered negligible (a), $Z \to \tau\tau$ CKKW uncertainty in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ boosted high-p_{T} region accepted as both normalization and shape systematic (b), one of the jet energy resolution systematic component for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ boosted high-p_{T} region where the symmetrisation has been applied (c). The solid lines represent the distributions before applying the procedures mentioned in the text while the dashed lines are the distributions after applying them.

The test statistic used in the analysis to extract the signal strength is the t_μ test statistic defined in Equation 5.21, which reduces to the q_0 test statistic (Equation 5.22) in order to quantify the significance of any excess in data, namely to test the compatibility between the background-only hypothesis and the observed data. All the statistical analysis from the likelihood construction to the profile likelihood ratio test statistic have been performed using a framework based on the RooFit and RooStats packages [133].

5.9 Results

The results of this analysis will be summarised in this section with particular emphasys on the $\tau_{\text{had}}\tau_{\text{had}}$ channel. When the results are referred to the $\tau_{\text{had}}\tau_{\text{had}}$ channel only this means that the $\tau_{\text{had}}\tau_{\text{had}}$ standalone fit has been performed, namely a fit which is reduced to only the $\tau_{\text{had}}\tau_{\text{had}}$ regions and the two $Z \to \tau\tau$ normalization factors are retrieved using only the VBF and boosted $\tau_{\text{had}}\tau_{\text{had}}$ regions respectively. Furthermore, some results may be referred to the VBF only or to the boosted only regions, meaning that the fit has been performed considering only the boosted or the VBF subregions.
In addition to the observed results, expected results will be presented as well, obtained using an Asimov dataset as described in Section 5.8.1, where all the observed data are replaced by the SM prediction and the nuisance parameters are set to their nominal value.

5.9.1 Expected results

A maximum likelihood fit using Asimov data has been firstly performed as this can be used to validate the fit model. In addition, it gives useful insights on the expected significance which is used to optimise the analysis.

$H \to \tau_{\mathbf{had}}\tau_{\mathbf{had}}$

In this section, results referred to the $\tau_{\text{had}}\tau_{\text{had}}$ standalone maximum likelihood fit are discussed.
The default parameter of interest used in the fit is the signal strength μ, i.e. the ratio of measured signal yield to its SM prediction, $\mu = \frac{\sigma_{\text{obs}}}{\sigma_{\text{SM}}}$. Therefore, if not stated otherwise, we will refer to the signal strength as the parameter of interest.
The compatibility between the background-only hypothesis and the observed data is established using the q_0 test statistic as explained in Section 5.8.1. The compatibility is then quantified by the p-value and the equivalent expected significance, which are reported in Table 5.29. In addition, a fit where all the nuisance parameters are kept fixed to their nominal value, meaning that the systematic uncertainties do not affect the result and only the statistical uncertainties have an impact. These significances are also quoted in Table 5.29 where the large impact of systematic uncertainties, resulting in a $\sim 22\%$ drop in sensitivity, is shown.

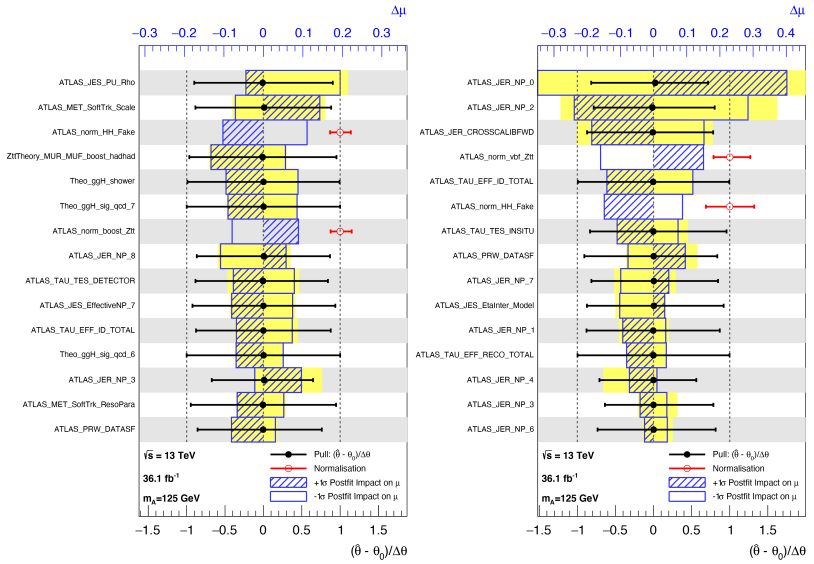
The behaviour of the NPs is highly important for validating the fit model. The impact of each NP on the measured parameter of interest can be derived by shifting a given NP to its $\pm 1\sigma$ uncertainty and then repeating the fit for the other NPs. The difference between the resulted $\hat{\mu}$ and the unconditional ML estimator μ determines the impact on the signal

	$Z_{\text{exp-stat+syst}}[\sigma]$	p-value (stat+syst)	$Z_{\text{exp-stat}}[\sigma]$	p-value(stat)
VBF	1.84	0.033	2.45	7.2×10^{-3}
Boosted	1.85	0.032	2.67	3.8×10^{-3}
Combined	2.84	2.3×10^{-3}	3.63	$1.4 \times 10^{-}4$

Table 5.29: Expected significances and p-values in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The significances and p-values are reported both for the case where all the NPs are included as free floating parameters in the fit and both for the case where they are fixed to their nominal value.

strength. Furthermore the differences between pre-fit and post-fit nominal values and uncertainties of each NP are crucial for monitoring the fit behaviour. Therefore it is useful to display the difference between the extracted value from the fit of each NP and its nominal input value, normalized to the NP pre-fit uncertainty (the so called "pulls"). The fit can prefer different θ values with respect to the nominal ones and then it is actually able to modify the background prediction. In addition, the $\pm 1\sigma$ uncertainties can be equal or shifted from their pre-fit values, more precisely the uncertainties can be smaller than the pre-fit ones resulting in a constraint of the NP. A constrained NP must always be investigated since this could indicate a bad behaviour of the fit, otherwise if the constraint is considered real this means that an higher precision for the NP can be achieved with the actual data and the pre-fit estimation may be too much conservative. All the NPs features described so far are represented in Figure 5.21 where the NPs are listed in decreasing order of their impact on the signal strength. Both the uncertainties and the pulls can be visualised in this plot.

All the NPs in Figure 5.21 are not pulled since the results shown in this Section are the expected results obtained by using an Asimov dataset where the NPs are set to their input values. Therefore, by construction we expect all the NPs to have their best-fit values equal to their input values. Experimental systematic uncertainties are dominant in the case of VBF region while some theoretical uncertainties appear in the highly ranked systematics for the boosted region, specifically systematics referring to ggF parton shower and to ggF and $Z \to \tau\tau$ QCD scale. In the combined $\tau_{\text{had}}\tau_{\text{had}}$ fit, uncertainties on the normalization of $Z \to \tau\tau$ have a large impact (from both the VBF and the boosted region). In addition, some jet energy resolution uncertainties are between those highly ranked as well as τ related uncertainties and signal and $Z \to \tau\tau$ theory uncertainties, where the latter are particularly driven by the boosted region. One of the components of jet energy resolution uncertainties ("jer_np_0") shows a significant constraint as it can be also seen in Figure 5.22 where the pulls and constraints for all the NPs are shown. The reason for the behaviour of this NP, which has shown similar constraints in the other channels, has been investigated in the analysis. In the case of real τ events the $m_{\tau\tau}^{\text{MMC}}$ distribution is sensitive to the jet related uncertainties beacuse the di-τ events are characterised by jets that recoil against the two τ. Therefore, jet related uncertainties cause a shift in the mean jet p_T that directly translates into a shift of the $E_{\text{T}}^{\text{miss}}$, that in turns affects the $m_{\tau\tau}^{\text{MMC}}$ reconstruction (shifted to higher values as demonstrated by Figure 5.17). The constraint originates from the measurement of the Z peak position which is apparently



(a) $\tau_{\text{had}}\tau_{\text{had}}$ boosted

(b) $\tau_{\text{had}}\tau_{\text{had}}$ VBF

(c) $\tau_{\text{had}}\tau_{\text{had}}$ combined

Figure 5.21: Expected impact of nuisance parameters on the signal strength for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit in the case of boosted region (a), VBF region (b) or combining all the regions (c). The uncertainties are listed in decreasing order of their impact on μ and only the 15 most highly ranked uncertainties are shown. The hatched blue and the open blue boxes represent the $+1\sigma$ and the -1σ uncertainties respectively while the yellow box represent the pre-fit ones. Their fractional impact is shown referring to the top x-axis. The pulls and constraints of the uncertainties are indicated by the black points and lines, referring to the bottom x-axis. If the uncertainty is not pulled nor constrainted the pull is 0 and the lines lie between $+1$ and -1.

more sensitive in the analysis phase space than in the subsidiary measurement done by ATLAS [76].

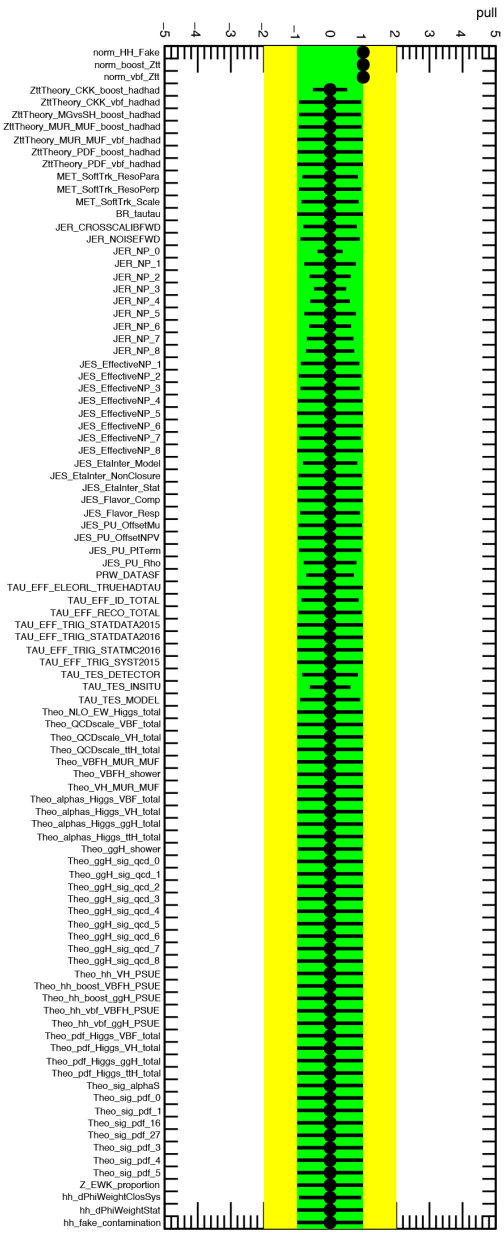


Figure 5.22: Expected nuisance parameters pulls for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit.

Other constraints can be observed for the uncertainties related to the τ reconstruction

and identification. The $\tau_{\text{had}}\tau_{\text{had}}$ channel probes very specific phase space with respect to the ATLAS performance measurement [81, 82], mostly based on $Z \to \tau_{\text{lep}}\tau_{\text{had}}$ decays, therefore the constraints are somehow expected.
Another useful check consists in monitoring the correlation between NPs built by the fit. Figure 5.23 shows the correlations between NPs which are higher than 0.25 in the $\tau_{\text{had}}\tau_{\text{had}}$ channel (signal strength and normalization factors are considered as well). A clear correlation between the VBF and boosted $Z \to \tau\tau$ NF is observed and can be expected as the normalization of the $Z \to \tau\tau$ affects the two categories similarly. Beside that, an anti-correlation between $Z \to \tau\tau$ NFs and the uncertainty on the τ identification efficiency can also be noticed.

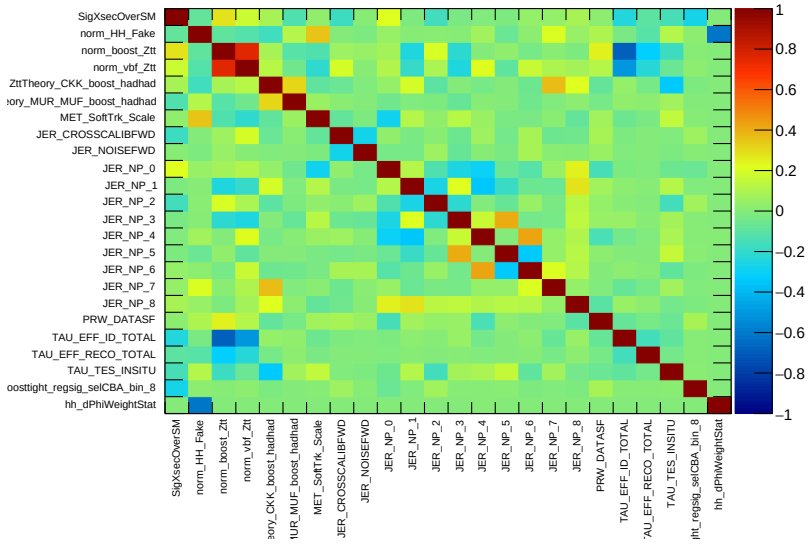


Figure 5.23: Expected nuisance parameters correlation for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit. Only the nuisance parameters with absolute correlation higher than 0.25 are shown.

An unconditional maximum likelihood fit is then performed in order to extract the parameter of interest μ. The resulting μ for the $\tau_{\text{had}}\tau_{\text{had}}$ standalone fit is reported in Table 5.30 together with its statistical and systematic uncertainties. Furthermore, the impact of different groups of systematic uncertainties is also quoted. This is obtained as for the nuisance parameters ranking, by fixing the NPs corresponding to a certain category and repeating the fit. As expected, the resulting value for μ is 1 since the fit has been conducted using an Asimov dataset. From Table 5.30 it is possible to notice how the $\tau_{\text{had}}\tau_{\text{had}}$ channel (but this is valid for the combined result as well) is dominated by systematic uncertainties rather than the statistical ones, arising particularly from jet and missing transverse energy related uncertainties, from signal theory systematics as well as from Monte Carlo statistics.

POI μ	Central Value 1
Set of nuisance parameters	Impact on uncertainty
Total	+0.44 / -0.38
Data statistics	+0.24 / -0.23
Total systematic unc.	+0.37 / -0.30
$Z \to \tau\tau$ normalization	+0.16 / -0.08
All normalizations	+0.16 / -0.09
All but normalizations	+0.36 / -0.29
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	+0.20 / -0.15
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.00 / -0.00
Taus	+0.17 / -0.11
Pileup reweighting	+0.02 / -0.01
Misedintified τ	+0.04 / -0.03
Luminosity	+0.02 / -0.01
Signal theory unc.	+0.16 / -0.08
$Z \to \tau\tau$ theory unc.	+0.08 / -0.08
MC statistics	+0.17 / -0.16

Set of nuisance parameters	Fractional impact on uncertainty (square of fraction of total)
Data statistics	+0.28 / -0.37
Total systematic unc.	+0.72 / -0.63
$Z \to \tau\tau$ normalization	+0.13 / -0.05
All normalizations	+0.13 / -0.06
All but normalizations	+0.67 / -0.58
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	+0.19 / -0.16
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.00 / -0.00
Taus	+0.2 / -0.08
Pileup reweighting	+0.002 / -2.8e-05
Misedintified τ	+0.007 / -0.006
Luminosity	+0.003 / -0.001
Signal theory unc.	+0.13 / -0.04
$Z \to \tau\tau$ theory unc.	+0.03 / -0.04
MC statistics	+0.15 / -0.18

Table 5.30: Expected impact of uncertainties, grouped in categories, on the signal strength for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel standalone fit. On the left absolute values are shown while on the right the fractional contribution is given.

Combined

The results obtained in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel are combined with those of the $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels, using a combined likelihood fit performed as described in Section 5.8.2. The expected significances reached in each channel and for the combined result are presented in Table 5.31.

	Combined	$\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$	$\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	$\tau_{\mathrm{had}}\tau_{\mathrm{had}}$
$Z_{\mathrm{exp}}[\sigma]$	4.48	1.20	2.81	2.84

Table 5.31: Expected significances in each channel and for the combined result.

The sensitivity is mainly driven by the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ and the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channels, with the sensitivity reached in $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ being slightly higher. The NPs ranked in order of decreasing impact on the signal strength are shown in Figure 5.24. It is interesting to notice that many signal theory uncertainties have a large impact, particularly almost all the components of QCD scale uncertainties for the ggF process and parton shower uncertainties, with the highest ranked being the Higgs transverse momentum shape dependence. In addition, uncertainty on the $Z \to \tau\tau$ normalization and jet energy scale and resolution uncertainties are highly ranked as well. The most constrained nuisance parameters are shown in Figure 5.25. The same statements made for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel are valid for the combined result as well, either because the constraints are driven by the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel or beacuse the NPs present a similar behaviour in the other two channels. The

correlations among NPs higher than 25% are shown in Figure 5.26. Beside the correlations between $Z \to \tau\tau$ normalization in the boosted and vbf region, already discussed for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, other major anticorrelations involve the two normalization factors of the $Z \to ll$ background in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel with a jet energy scale component uncertainty for this specific process, decorrelated from the rest of the uncertainty. In addition, anticorrelations between the Top background normalization in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel and the uncertainty related to the statistics of the misidentified τ template in the same channel can be noticed.
Finally the extracted value from the fit of the signal strength is given in Table 5.32 together with its uncertainties. The contribution of systematic uncertainties is significant here as well, as already illustrated for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The main uncertainties arise from jet and missing transverse energy related systematics, signal theory systematics as well as Monte Carlo statistics.

POI μ	Central Value 1
Set of nuisance parameters	Impact on uncertainty
Total	+0.28 / -0.25
Data statistics	+0.16 / -0.15
Total systematic unc.	+0.23 / -0.20
$Z \to \tau\tau$ normalization	+0.06 / -0.04
All normalizations	+0.07 / -0.05
All but normalizations	+0.22 / -0.19
Jets and $E_{\text{T}}^{\text{miss}}$	+0.11 / -0.09
Flavor tagging	+0.03 / -0.02
Electrons and muons	+0.01 / -0.01
Taus	+0.06 / -0.04
Pileup reweighting	+0.01 / -0.01
Misedintified τ	+0.04 / -0.03
Luminosity	+0.03 / -0.02
Signal theory unc.	+0.13 / -0.09
$Z \to \tau\tau$ theory unc.	+0.05 / -0.04
MC statistics	+0.10 / -0.10

Set of nuisance parameters	Fractional impact on uncertainty (square of fraction of total)
Data statistics	+0.31 / -0.38
Total systematic unc.	+0.69 / -0.62
$Z \to \tau\tau$ normalization	+0.05 / -0.02
All normalizations	+0.06 / -0.03
All but normalizations	+0.63 / -0.55
Jets $E_{\text{T}}^{\text{miss}}$	+0.14 / -0.12
Flavor tagging	+0.01 / -0.008
Electrons and muons	+0.002 / -0.001
Taus	+0.04 / -0.03
Pileup reweighting	+0.002 / -0.0004
Misedintified τ	+0.02 / -0.02
Luminosity	+0.01 / -0.008
Signal theory unc.	+0.23 / -0.12
$Z \to \tau\tau$ theory unc.	+0.03 / -0.03
MC statistics	+0.14 / -0.15

Table 5.32: Expected impact of uncertainties, grouped in categories, on the signal strength for the combined $H \to \tau\tau$ fit. On the left absolute values are shown while on the right the fractional contribution is given.

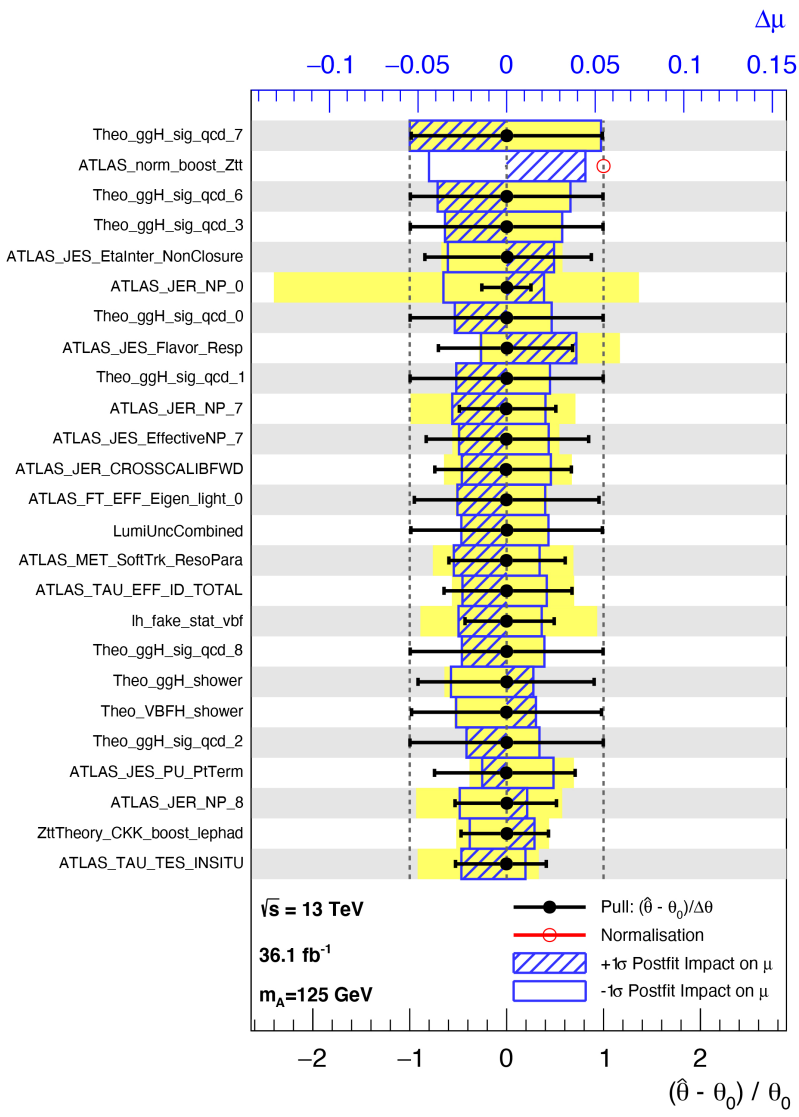


Figure 5.24: Expected impact of nuisance parameters on the signal strength for the combined $H \to \tau\tau$ fit. The uncertainties are listed in decreasing order of their impact on μ and only the 25 most highly ranked uncertainties are shown. The hatched blue and the open blue boxes represent the $+1\sigma$ and the -1σ uncertainties respectively while the yellow box represent the pre-fit ones. Their fractional impact is shown referring to the top x-axis. The pulls and constraints of the uncertainties are indicated by the black points and lines, referring to the bottom x-axis. If the uncertainty is not pulled nor constrainted the pull is 0 and the lines lie between $+1$ and -1.

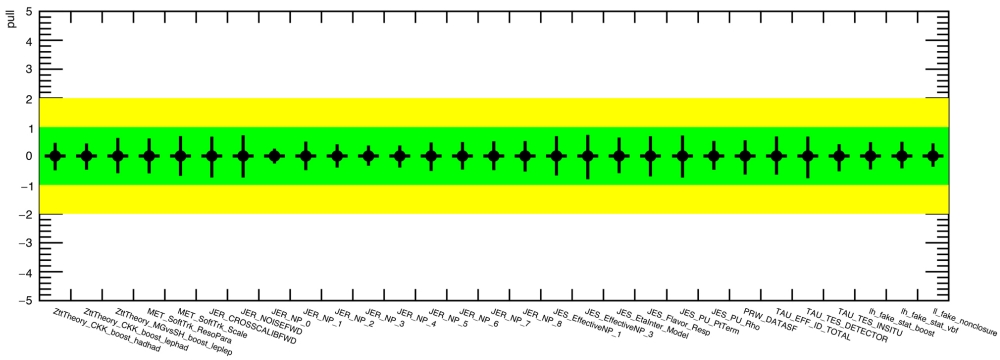


Figure 5.25: Expected nuisance parameters pulls for the combined $H \to \tau\tau$ fit. Only the parameters for which the post-fit uncertainty is less than 75% of the pre-fit uncertainty are shown.

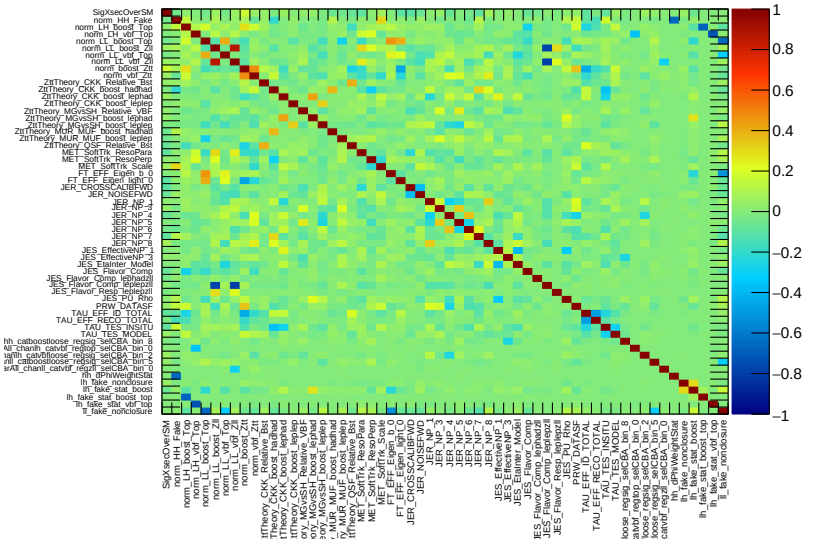


Figure 5.26: Expected nuisance parameters correlation for the combined $H \to \tau\tau$ fit. Only the nuisance parameters with absolute correlation higher than 0.25 are shown.

5.9.2 Observed results

After the analysis optimization, done performing a fit to Asimov data, the observed results are finally derived through a maximum likelihood fit where the data are used in all the signal and control regions.

$H \to \tau_{\mathbf{had}}\tau_{\mathbf{had}}$

Before performing the actual fit to data, a fit, which still employs an Asimov data but built using NPs and NFs evaluated in a data fit, has been performed. This allows to have a more realistic prediction for the expected results since the NPs and NFs central values are set to the values determined in a fit to data. Significances for this kind of fit are given in Table 5.33.

	$Z_{\text{exp}}[\sigma]$	p-value
VBF	1.78	0.038
Boosted	1.80	0.036
Combined	2.85	2.2×10^{-3}

Table 5.33: Expected significances and p-values in the $\tau_{\text{had}}\tau_{\text{had}}$ channel calculated in a fit where the Asimov dataset is built using nuisance parameters and normalization factors evaluated in a fit to data.

The observed significances are instead presented in Table 5.34. We observe a decrease in sensitivity for the $\tau_{\text{had}}\tau_{\text{had}}$ channel with respect to the expected one (similarly in the VBF and Boosted regions).

	$Z_{\text{obs}}[\sigma]$	p-value
VBF	1.40	0.08
Boosted	1.24	0.11
Combined	2.21	0.014

Table 5.34: Observed significances and p-values in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

The observed impact of NPs on the signal strenght can be seen in Figure 5.27. The NPs show a similar behaviour as for the expected results, similar ranking and similar constraints are observed, as demonstrated also by Figure 5.28. Jet related uncertainties, uncertainty on the $Z \to \tau\tau$ normalization and tau related uncertainties have a significant impact. In addition, jet related uncertainties are in some cases strongly constrained and also shifted from their nominal value. The same reasons for the constraints, explained for the expected results, apply to the observed results as well, while some more details on the pulls of NPs will be discussed for the combined $H \to \tau\tau$ results as these pulls affect the other channels as well. Finally, Figure 5.26 shows the correlation among NPs which exhibits no particular unforeseen behaviour with respect to the expected results.

The observed μ extracted from the fit is:

$$\mu = 0.73^{+0.41}_{-0.35} \tag{5.29}$$

which is also given in Table 5.35 together with the impact of the various types of uncertainties. The observed μ indicates an observed yield for signal compatible the SM prediction. The uncertainties exhibit a slightly higher impact than expected (Table 5.30) but their relative weight follows expectations. Therefore, the most significant contributions come from jet and missing transverse energy related uncertainties, signal theory uncertainties, tau related uncertainties as well as Monte Carlo statistics. Moreover, in Table 5.36 are reported all the normalization factors as determined from the fit.

POI μ	Central Value 0.73
Set of nuisance parameters	Impact on uncertainty
Total	+0.41 / -0.35
Data statistics	+0.23 / -0.22
Total systematic unc.	+0.34 / -0.27
$Z \to \tau\tau$ normalization	+0.13 / -0.06
All normalizations	+0.13 / -0.07
All but normalizations	+0.33 / -0.25
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	+0.19 / -0.14
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.00 / -0.00
Taus	+0.13 / -0.06
Pileup reweighting	+0.01 / -0.00
Misedintified τ	+0.05 / -0.04
Luminosity	+0.02 / -0.01
Signal theory unc.	+0.16 / -0.07
$Z \to \tau\tau$ theory unc.	+0.09 / -0.07
MC statistics	+0.16 / -0.16

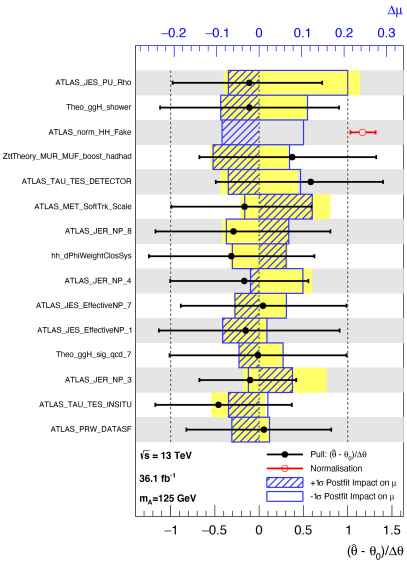
Set of nuisance parameters	Fractional impact on uncertainty (square of fraction of total)
Data statistics	+0.31 / -0.40
Total systematic unc.	+0.69 / -0.60
$Z \to \tau\tau$ normalization	+0.10 / -0.03
All normalizations	+0.10 / -0.04
All but normalizations	+0.64 / -0.54
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	+0.22 / -0.16
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.00 / -0.00
Taus	+0.1 / -0.03
Pileup reweighting	+1.7e-05 / -3.3e-05
Misedintified τ	+0.01 / -0.01
Luminosity	+0.002 / -0.0009
Signal theory unc.	+0.15 / -0.04
$Z \to \tau\tau$ theory unc.	+0.05 / -0.05
MC statistics	+0.16 / -0.20

Table 5.35: Observed impact of uncertainties, grouped in categories, on the signal strength for the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel standalone fit. On the left absolute values are shown while on the right the fractional contribution is given.

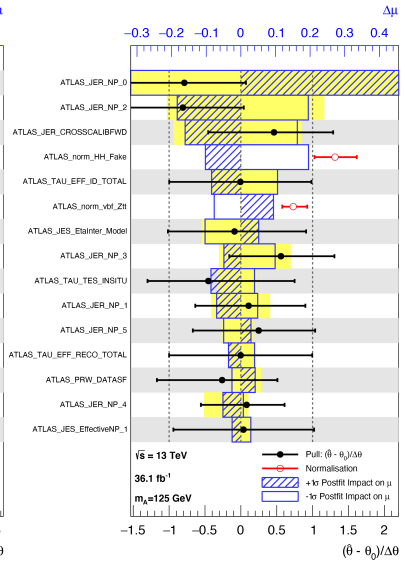
Normalization factor	Value
$Z \to \tau\tau$ (VBF)	0.82 ± 0.16
$Z \to \tau\tau$ (Boosted)	0.90 ± 0.12
Misidentified τ	1.15 ± 0.14

Table 5.36: Normalization factors for the $Z \to \tau\tau$ and the misidentifed τ backgrounds derived from the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ standalone fit.

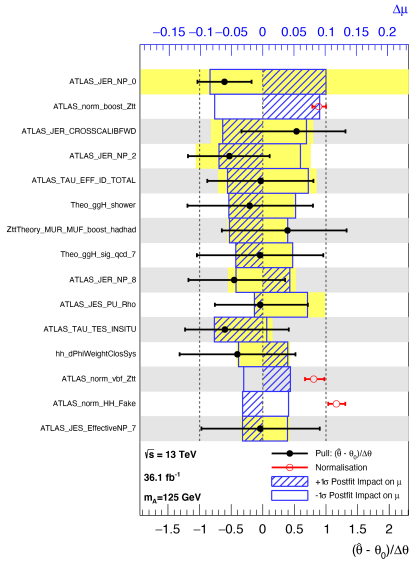
The result on the signal strength can be reinterpreted in terms of cross section. Indeed, from the definition of signal strength, $\mu = \frac{[\sigma \times B.R.(H \to \tau\tau)]_{\mathrm{obs}}}{[\sigma \times B.R.(H \to \tau\tau)]_{\mathrm{SM}}}$, it is possible to derive the



(a) $\tau_{\text{had}}\tau_{\text{had}}$ boosted



(b) $\tau_{\text{had}}\tau_{\text{had}}$ VBF



(c) $\tau_{\text{had}}\tau_{\text{had}}$ combined

Figure 5.27: Observed impact of nuisance parameters on the signal strength for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit in the case of boosted region (a), VBF region (b) or combining all the regions (c). The uncertainties are listed in decreasing order of their impact on μ and only the 15 most highly ranked uncertainties are shown. The hatched blue and the open blue boxes represent the $+1\sigma$ and the -1σ uncertainties respectively while the yellow box represent the pre-fit ones. Their fractional impact is shown referring to the top x-axis. The pulls and constraints of the uncertainties are indicated by the black points and lines, referring to the bottom x-axis. If the uncertainty is not pulled nor constrainted the pull is 0 and the lines lie between $+1$ and -1.

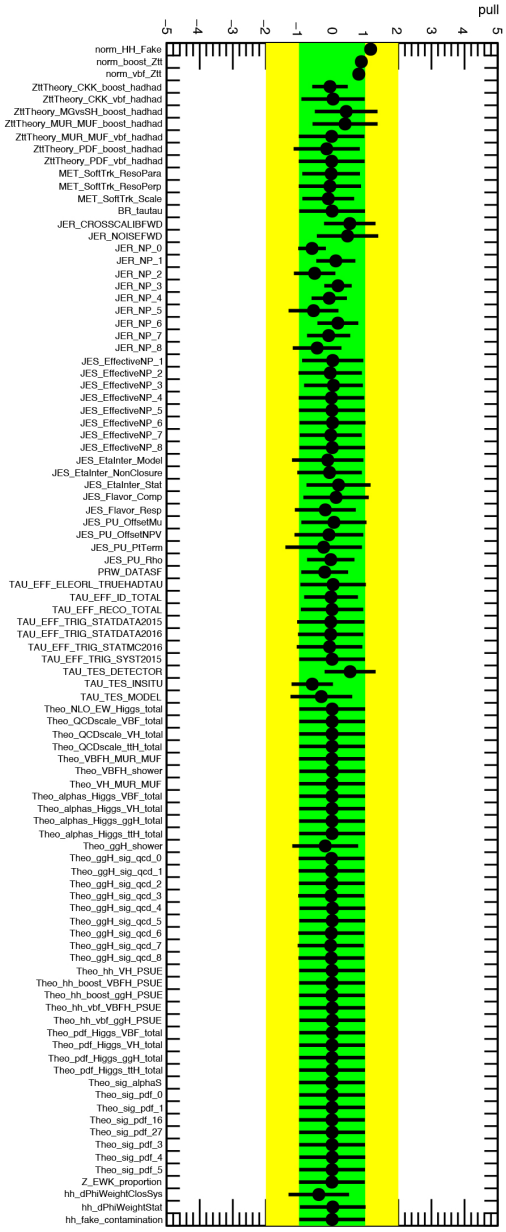


Figure 5.28: Observed nuisance parameters pulls for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit.

observed cross section times the $H \to \tau\tau$ branching ratio as:

$$\sigma_{H\to\tau\tau,\text{obs}} = \mu[\sigma \times B.R.(H \to \tau\tau)]_{\text{SM}} \tag{5.30}$$

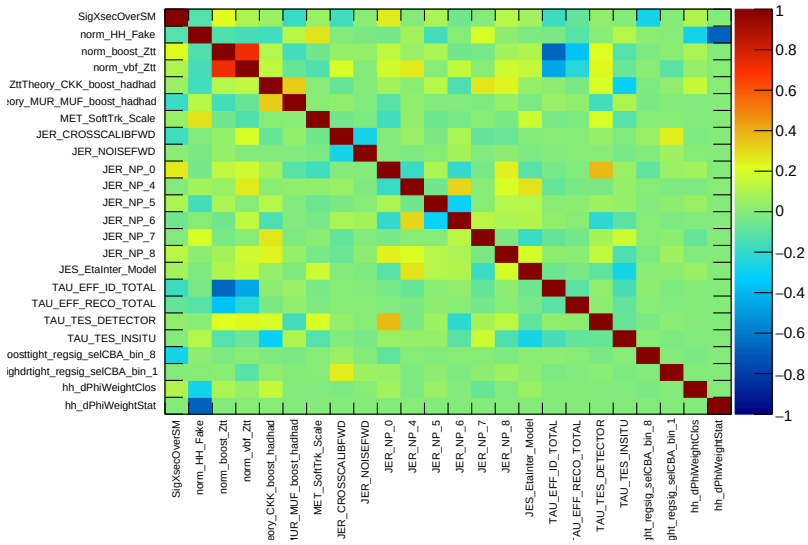


Figure 5.29: Observed nuisance parameters correlation for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit. Only the nuisance parameters with absolute correlation higher than 0.25 are shown.

A fit using $\sigma_{H\to\tau\tau,\text{obs}}$ as parameter of interest has then been performed. Considering all the Higgs main production processes at the LHC (ggF, VBF, VH and $t\bar{t}H$) with the cross sections listed in Table 5.37 (already scaled by the B.R.$(H \to \tau\tau) = 0.06256$), the resulted cross section from the $\tau_{\text{had}}\tau_{\text{had}}$ standalone fit is:

$$\sigma_{H\to\tau\tau} = 2.49^{+1.38}_{-1.17} \text{ pb} \tag{5.31}$$

Production mode	$\sigma^{\text{SM}}_{H\to\tau\tau}$ [pb]
ggF	3.035
VBF	0.236
VH	0.141
$t\bar{t}H$	0.053
total	3.465

Table 5.37: Cross sections of the Higgs production modes times the branching of Higgs decaying into a pair of tau leptons, as predicted by the SM.

Combined

A combined fit to data, which exploits all the signal and control regions of the three $H \to \tau\tau$ subchannels, leads to an observed (expected) significance of the signal excess relative to the background only hypothesis of 4.4 (4.1) standard deviations. The expected

signficance is calculated, as for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, using an Asimov dataset built with NPs and NFs evaluated from a fit to data. The expected significances for each channel and for the combined fit can be found in Table 5.38 while the observed ones are reported in Table 5.39.

	Combined	$\tau_{\text{lep}}\tau_{\text{lep}}$	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
$Z_{\text{exp}}[\sigma]$	4.13	1.20	2.63	2.84

Table 5.38: Expected significances in each channel and for the combined result using nuisance parameters and normalization factors obtained from a fit to data.

	Combined	$\tau_{\text{lep}}\tau_{\text{lep}}$	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
$Z_{\text{obs}}[\sigma]$	4.36	2.20	2.43	2.21

Table 5.39: Observed significances in each channel and for the combined result using nuisance parameters and normalization factors obtained from a fit to data.

The expected significance is mainly driven by the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels with the latter showing a slightly higher significance while for the observed data similar significances across the channels are observed with a drop in sensitivity for the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels compensated by the increased sensitivity in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel. However, it is not straightforward to estimate the combined expected significance from the three standalone significances since in the combined fit the signal is correlated across the channels. This result constitutes an evidence for the $H \to \tau\tau$ decay and it is an improvement of the result performed using Run 1 data at $\sqrt(s) = 7$ and 8 TeV center of mass energy [88]. In addition, a fit combining this measurement with the Run 1 measurement, has been carried out and has led to a combined observed (expected) significance of 6.4 (5.4) standard deviations, establishing the first ATLAS observation of the Higgs decaying into a pair of tau leptons.
The signal strength extracted from the combined fit is:

$$\mu = 1.09^{+0.18}_{-0.17}\ (\text{stat.})\ ^{+0.26}_{-0.22}\ (\text{syst.})\ ^{+0.16}_{-0.11}\ (\text{theory syst.}) \tag{5.32}$$

and it is shown in Table 5.40 together with the list of uncertainties impact and summarised in Figure 5.30. The value is compatible within uncertainties with 1, i.e. with the SM prediction for a Higgs with mass $m_H = 125$ GeV. As already foreseen by the expected results, the high impact of systematic uncertainties can be observed, greater than the effect of statistical uncertainties, and the large contribution of theoretical systematics on the signal as well. Among the systematics uncertainties the jet and missing transerve energy related uncertainties and the Monte Carlo statistics are quite significant. In addition, the values of NFs obtained in the combined fit to data are given in Table 5.41. The two $Z \to \tau\tau$ normalization factors are different from the values found in the $\tau_{\text{had}}\tau_{\text{had}}$ standalone fit since these parameters are now correlated across the channels.

The differences in normalization factors are absorbed in the large pull of the systematic related to identification efficiency and energy scale for taus (particularly "tau_eff_id_total" and "tau_tes_detector") since these are anticorrelated with the $Z \to \tau\tau$ NFs. As can be seen in Figure 5.31, another significant pull appears for the tau energy scale uncertainty related to the tag and probe measurement ("tau_tes_insitu"). The "tau_tes_detector" and "tau_tes_insitu" uncertainties pulls originate from the $\tau_{\text{had}}\tau_{\text{had}}$ channel even if their shift is more pronounced in the combined fit while the "tau_eff_id_total" exhibits no pulls in the individual channels fit. Some other parameters related to the jet energy resolution uncertainties show large pulls and triggered some further investigations in the analysis. In order to monitor the behaviour of such NPs and to exclude the presence of multiple minima, a profile likelihood scan has been performed for each parameter. Figure 5.32 shows an example for leading component of jet energy resolution uncertainty. Here, the nominal value of a given parameter is fixed to a value within $\pm 2.5\ \sigma$ of its input uncertainty. For each point of the grid the negative log-likelihood is evaluated and the difference with the minimum likelihood value found is determined and represented in the plot. This likelihood scan, as many others checked in the analysis, does not exhibit any particular bad behaviour of the fit.

POI μ	Central Value 1.09
Set of nuisance parameters	Impact on uncertainty
Total	+0.32 / -0.28
Data statistics	+0.18 / -0.17
Total systematic unc.	+0.26 / -0.22
$Z \to \tau\tau$ normalization	+0.06 / -0.04
All normalizations	+0.07 / -0.05
All but normalizations	+0.25 / -0.21
Jets and $E_{\text{T}}^{\text{miss}}$	+0.13 / -0.10
Flavor tagging	+0.04 / -0.03
Electrons and muons	+0.01 / -0.01
Taus	+0.05 / -0.03
Pileup reweighting	+0.02 / -0.01
Misidentified τ	+0.05 / -0.05
Luminosity	+0.04 / -0.03
Signal theory unc.	+0.16 / -0.11
$Z \to \tau\tau$ theory unc.	+0.05 / -0.04
MC statistics	+0.12 / -0.11

Set of nuisance parameters	Fractional impact on uncertainty (square of fraction of total)
Data statistics	+0.30 / -0.37
Total systematic unc.	+0.70 / -0.63
$Z \to \tau\tau$ normalization	+0.04 / -0.02
All normalizations	+0.05 / -0.03
All but normalizations	+0.63 / -0.55
Jets and $E_{\text{T}}^{\text{miss}}$	+0.16 / -0.13
Flavor tagging	+0.01 / -0.01
Electrons and muons	+0.002 / -0.002
Taus	+0.03 / -0.02
Pileup reweighting	+0.003 / -0.0005
Misidentified τ	+0.03 / -0.03
Luminosity	+0.01 / -0.008
Signal theory unc.	+0.26 / -0.14
$Z \to \tau\tau$ theory unc.	+0.03 / -0.02
MC statistics	+0.15 / -0.15

Table 5.40: Observed impact of uncertainties, grouped in categories, on the signal strength for the combined $H \to \tau\tau$ fit. On the left absolute values are shown while on the right the fractional contribution is given.

The likelihood fit is able to adjust signal and background predictions, either the event yield and the $m_{\tau\tau}^{\text{MMC}}$ shape. Figure 5.33 and Figure 5.34 shows the $m_{\tau\tau}^{\text{MMC}}$ distributions in all the VBF and boosted SRs respectively after performing the fit, where the particular binning shown is the one used in the fit. The compatibility of data with the signal plus background hypothesis can be observed in each plot of Figure 5.33 and Figure 5.34. The

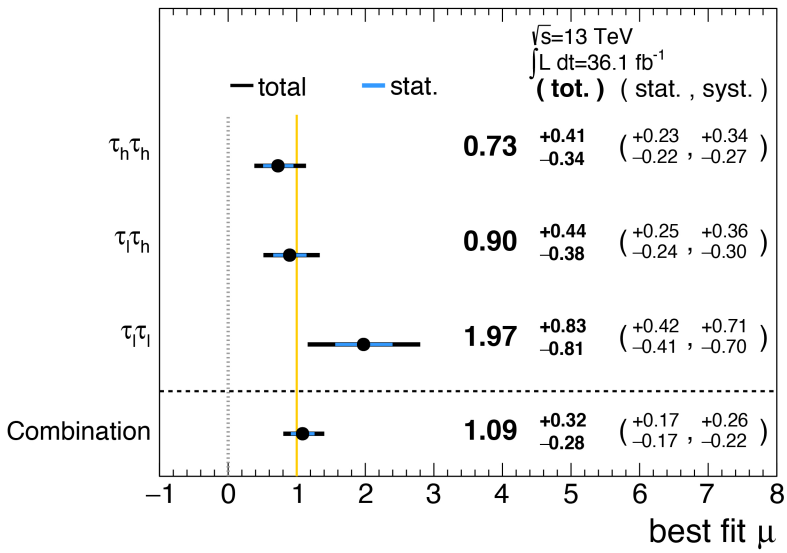


Figure 5.30: Signal strenght measurement for each channel and for the combined result. The theoretical prediction is shown in yellow.

Normalization Factor	Channel	Value	
		VBF	Boosted
$Z \to ll$	$\tau_{\text{lep}}\tau_{\text{lep}}$	$0.88^{+0.34}_{-0.30}$	$1.27^{+0.30}_{-0.25}$
Top	$\tau_{\text{lep}}\tau_{\text{lep}}$	1.19 ± 0.09	1.07 ± 0.05
Top	$\tau_{\text{lep}}\tau_{\text{had}}$	$1.53^{+0.30}_{-0.27}$	1.13 ± 0.07
Misidentified τ	$\tau_{\text{had}}\tau_{\text{had}}$	1.12 ± 0.12	
$Z \to \tau\tau$	$\tau_{\text{lep}}\tau_{\text{lep}}$ $\tau_{\text{lep}}\tau_{\text{had}}$ $\tau_{\text{had}}\tau_{\text{had}}$	$1.04^{+0.10}_{-0.09}$	1.11 ± 0.05

Table 5.41: Normalization factors for backgrounds that have their normalization retrieved from the fit to data. The values are extracted from the combined $H \to \tau\tau$ fit to data.

same statetement applies to the event yield prediction as shown in Tables 5.42, 5.43 and 5.44.

The fit is also repeated using the cross section times branching ratio as parameter of interest, as explained for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The measured value for the cross section in the combined fit is then:

$$\sigma_{H\to\tau\tau} = 3.77^{+0.60}_{-0.59}\ (\text{stat.})\ ^{+0.87}_{-0.74}\ (\text{syst.})\ \text{pb} \tag{5.33}$$

consistent with the SM prediction, $\sigma^{\text{SM}}_{H\to\tau\tau} = 3.46 \pm 0.13$ pb. The uncertainty on the SM cross section prediction refers to the signal theoretical uncertainties on the total cross section. In principle, signal theory systematics should not affect the measured total cross section however they still need to be considered due to the selection which defines the

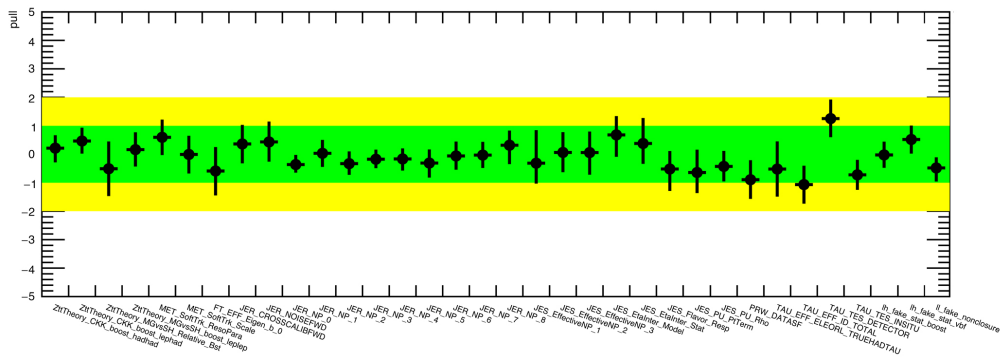


Figure 5.31: Observed nuisance parameters pulls for the combined $H \to \tau\tau$ fit. Only the parameters for which the postfit uncertainty is less than 75% of the pre-fit uncertainty and/or the pull is higher than 0.5 in absolute value are shown.

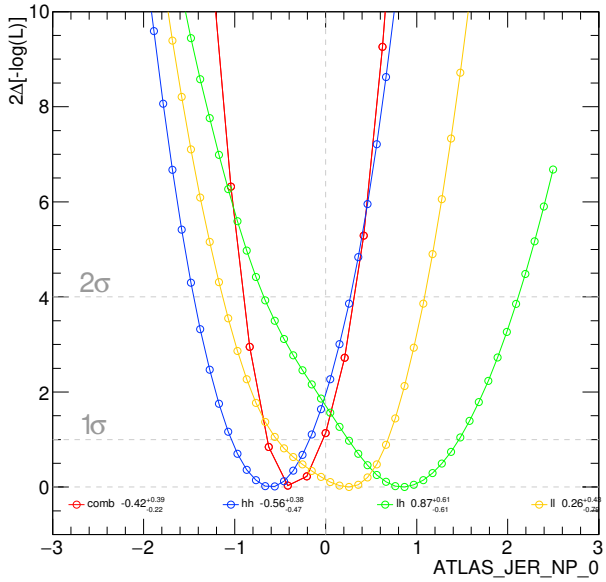


Figure 5.32: Likelihood scan for one of the components of the jet energy resolution systematic for each channel and in the combined $H \to \tau\tau$ fit.

specific fiducial region used for the measurement.
Figure 5.35 shows the systematic uncertainties with the largest impact on $\sigma_{H\to\tau\tau}$. As predicted from the expected fit, signal theory uncertainties, specifically QCD scale uncertainties, jet energy resolution and scale and normalization of the backgrounds ($Z \to \tau\tau$ and $Z \to ll$) have a significant impact (the same behaviour is expected for the signal strength and for the $\sigma_{H\to\tau\tau}$ because of how the cross section parameter of interest is constructed). In addition, Table 5.45 gives a summary of the observed and expected impact of uncertainties from different sources. As mentioned, the impact of jet and missing

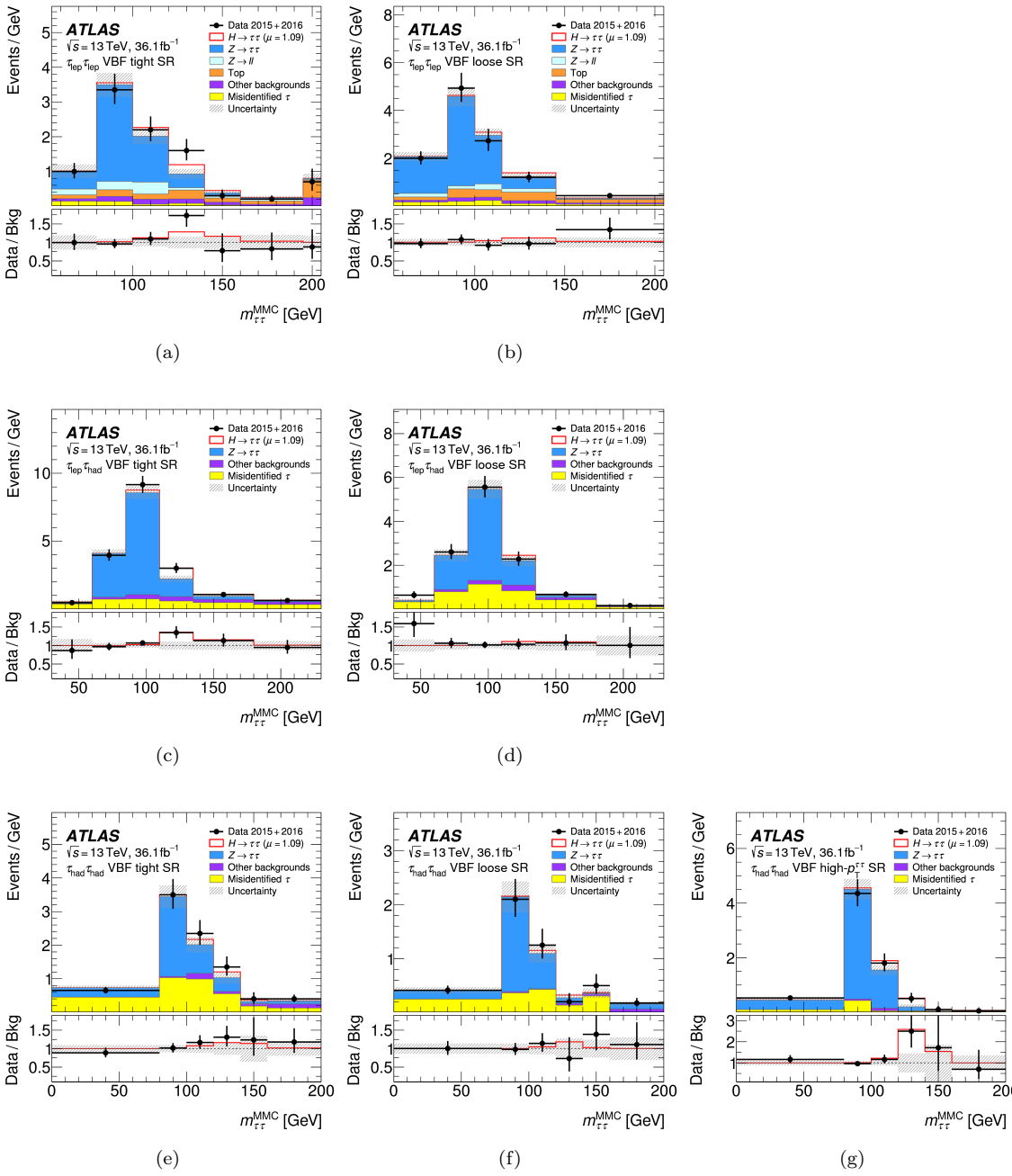
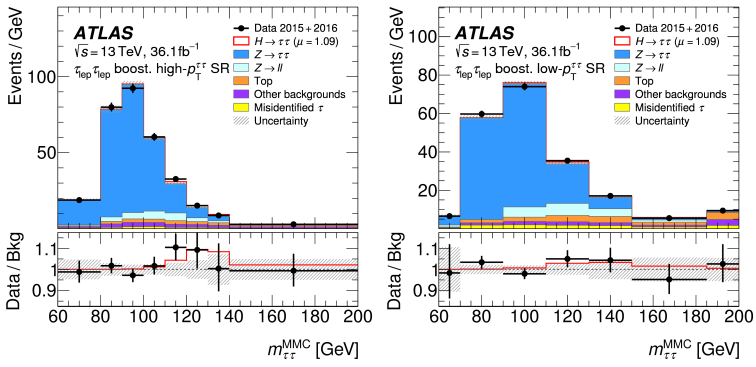
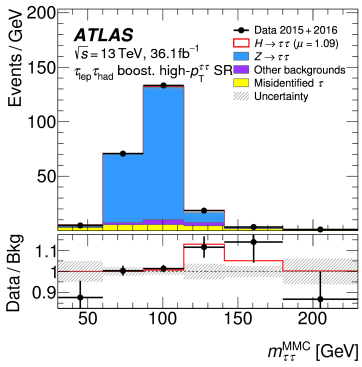


Figure 5.33: $m_{\tau\tau}^{\mathrm{MMC}}$ distribution in all the VBF signal regions of the three decay channels ($\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ at the top, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ in the midlle and $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ at the bottom) after performing the fit. The observed Higgs signal is shown with a red line.

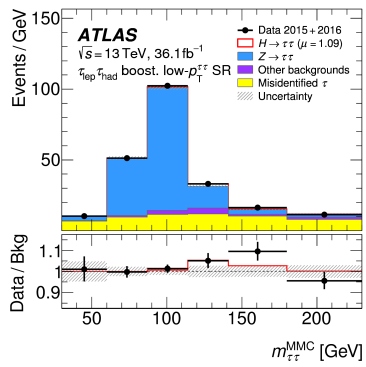


(a)

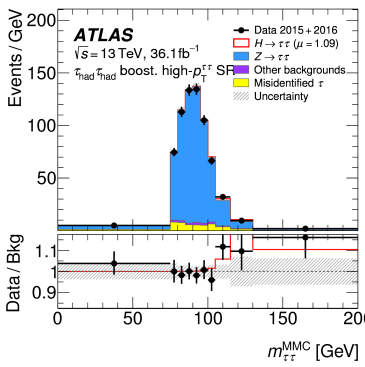
(b)



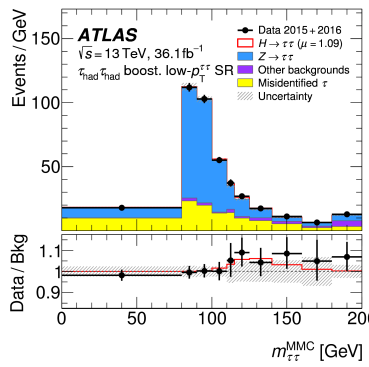
(c)



(d)



(e)



(f)

Figure 5.34: $m_{\tau\tau}^{\mathrm{MMC}}$ distribution in all the boosted signal regions of the three decay channels ($\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ at the top, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ in the midlle and $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ at the bottom) after performing the fit. The observed Higgs signal is shown with a red line.

	$\tau_{\text{lep}}\tau_{\text{lep}}$ VBF		$\tau_{\text{lep}}\tau_{\text{lep}}$ boosted	
	Loose	Tight	Low-$p_{\text{T}}^{\tau\tau}$	High-$p_{\text{T}}^{\tau\tau}$
$Z \to \tau\tau$	151 ± 13	107 ± 12	2977 ± 90	2687 ± 64
$Z \to \ell\ell$	15.1 ± 4.9	20.3 ± 6.6	360 ± 54	236 ± 31
Top	33.0 ± 6.4	25.1 ± 4.5	321 ± 50	189 ± 29
VV	11.8 ± 2.2	10.7 ± 1.5	194.1 ± 8.5	195.3 ± 8.8
Misidentified τ	18.3 ± 9.6	9.6 ± 4.8	209 ± 92	80 ± 35
ggF, $H \to WW^*$	1.2 ± 0.2	1.4 ± 0.3	11.8 ± 2.6	16.4 ± 1.7
VBF, $H \to WW^*$	1.7 ± 0.2	4.1 ± 0.5	2.9 ± 0.3	2.9 ± 0.3
ggF, $H \to \tau\tau$	2.6 ± 0.9	1.8 ± 0.9	34.4 ± 9.2	33.8 ± 9.5
VBF, $H \to \tau\tau$	5.3 ± 1.5	11.3 ± 3.0	7.7 ± 2.1	8.2 ± 2.3
WH, $H \to \tau\tau$	< 0.1	< 0.1	2.5 ± 0.7	3.1 ± 0.9
ZH, $H \to \tau\tau$	< 0.1	< 0.1	1.3 ± 0.4	1.6 ± 0.4
$t\bar{t}H$, $H \to \tau\tau$	< 0.1	0.1 ± 0.1	1.5 ± 0.5	1.2 ± 0.4
Total background	232 ± 13	178 ± 12	4075 ± 61	3408 ± 54
Total signal	8.0 ± 2.2	13.2 ± 3.5	47 ± 12	48 ± 12
Data	237	188	4124	3444

Table 5.42: Observed and predicted event yield in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel as calculated by the fit.

	$\tau_{\text{lep}}\tau_{\text{had}}$ VBF		$\tau_{\text{lep}}\tau_{\text{had}}$ boosted	
	Loose	Tight	Low-$p_{\text{T}}^{\tau\tau}$	High-$p_{\text{T}}^{\tau\tau}$
$Z \to \tau\tau$	178 ± 18	323 ± 21	4187 ± 92	5347 ± 82
$Z \to \ell\ell$	10.0 ± 3.0	12.7 ± 3.1	130 ± 37	115 ± 16
Top	5.8 ± 1.6	17.9 ± 4.6	121 ± 20	57 ± 10
Misidentified τ	103 ± 16	101 ± 15	1895 ± 80	605 ± 29
Other backgrounds	4.0 ± 1.6	9.3 ± 1.9	115.0 ± 7.8	129.0 ± 8.8
ggF, $H \to \tau\tau$	3.8 ± 1.1	7.1 ± 1.9	62 ± 16	66 ± 22
VBF, $H \to \tau\tau$	7.6 ± 2.2	24.7 ± 6.8	11.9 ± 3.4	14.0 ± 4.0
WH, $H \to \tau\tau$	< 0.1	0.1 ± 0.0	3.9 ± 1.1	5.4 ± 1.4
ZH, $H \to \tau\tau$	< 0.1	< 0.1	1.8 ± 0.5	2.8 ± 0.7
$t\bar{t}H$, $H \to \tau\tau$	< 0.1	< 0.1	0.1 ± 0.0	0.2 ± 0.1
Total background	301 ± 17	463 ± 21	6448 ± 81	6253 ± 80
Total signal	11.5 ± 3.2	32.0 ± 8.2	80 ± 20	89 ± 26
Data	318	496	6556	6347

Table 5.43: Observed and predicted event yield in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel as calculated by the fit.

transverse energy related uncertainties, of Monte Carlo statistics and of signal theory uncertainties is quite large for this measurement.

The $\sigma_{H\to\tau\tau}$ measured cross section is also shown in Figure 5.36, either for each standalone

	$\tau_{\text{had}}\tau_{\text{had}}$ VBF			$\tau_{\text{had}}\tau_{\text{had}}$ boosted	
	Loose	Tight	High-$p_{\text{T}}^{\tau\tau}$	Low-$p_{\text{T}}^{\tau\tau}$	High-$p_{\text{T}}^{\tau\tau}$
$Z \to \tau\tau$	67.3 ± 9.2	100 ± 12	141 ± 12	3250 ± 130	3582 ± 82
Misidentified τ	45.0 ± 5.4	96.4 ± 9.2	20.0 ± 2.9	1870 ± 140	364 ± 53
Other backgrounds	4.4 ± 1.4	11.6 ± 1.7	4.4 ± 0.7	281 ± 21	109.9 ± 9.2
ggF, $H \to \tau\tau$	1.1 ± 0.4	2.0 ± 0.7	3.5 ± 1.0	41 ± 11	48 ± 14
VBF, $H \to \tau\tau$	1.4 ± 0.5	6.4 ± 1.8	11.2 ± 3.0	9.0 ± 3.4	10.7 ± 2.9
WH, $H \to \tau\tau$	< 0.1	< 0.1	< 0.1	3.3 ± 0.9	4.4 ± 1.2
ZH, $H \to \tau\tau$	< 0.1	< 0.1	< 0.1	2.4 ± 0.7	2.9 ± 0.8
$t\bar{t}H$, $H \to \tau\tau$	< 0.1	< 0.1	< 0.1	1.6 ± 0.5	1.9 ± 0.5
Total background	116.7 ± 9.4	208 ± 12	165 ± 12	5401 ± 78	4057 ± 64
Total signal	2.6 ± 0.8	8.6 ± 2.4	14.9 ± 3.8	57 ± 15	68 ± 18
Data	121	220	179	5455	4103

Table 5.44: Observed and predicted event yield in the $\tau_{\text{had}}\tau_{\text{had}}$ channel as calculated by the fit.

channel fit and for the combined one. The results are compatible with the SM prediction within uncertainties, in particular the $\tau_{\text{had}}\tau_{\text{had}}$ channel exhibit the lowest uncertainty on $\sigma_{H\to\tau\tau}$.

Source of uncertainty	Impact $\Delta\sigma/\sigma_{H\to\tau\tau}$ [%]	
	Observed	Expected
Theoretical uncert. in signal	+13.4 / −8.7	+12.0 / −7.8
Background statistics	+10.8 / −9.9	+10.1 / −9.7
Jets and $E_{\text{T}}^{\text{miss}}$	+11.2 / −9.1	+10.4 / −8.4
Background normalization	+6.3 / −4.4	+6.3 / −4.4
Misidentified τ	+4.5 / −4.2	+3.4 / −3.2
Theoretical uncert. in background	+4.6 / −3.6	+5.0 / −4.0
Hadronic τ decays	+4.4 / −2.9	+5.5 / −4.0
Flavor tagging	+3.4 / −3.4	+3.0 / −2.3
Luminosity	+3.3 / −2.4	+3.1 / −2.2
Electrons and muons	+1.2 / −0.9	+1.1 / −0.8
Total systematic uncert.	+23 / −20	+22 / −19
Data statistics	±16	±15
Total	+28 / −25	+27 / −24

Table 5.45: Observed and expected fractional impact of uncertainties on $\sigma_{H\to\tau\tau}$. The uncertainties are listed in decreasing order of impact.

Due to the good sensitivity of the VBF and boosted category to the VBF and ggF production processes respectively, a two parameter fit has been performed to measure separately the VBF and ggF cross section. Two parameters of interest have been used in this case, $\sigma_{H\to\tau\tau}^{\text{VBF}}$ and $\sigma_{H\to\tau\tau}^{\text{ggF}}$, in order to distinguish between the two processes, while

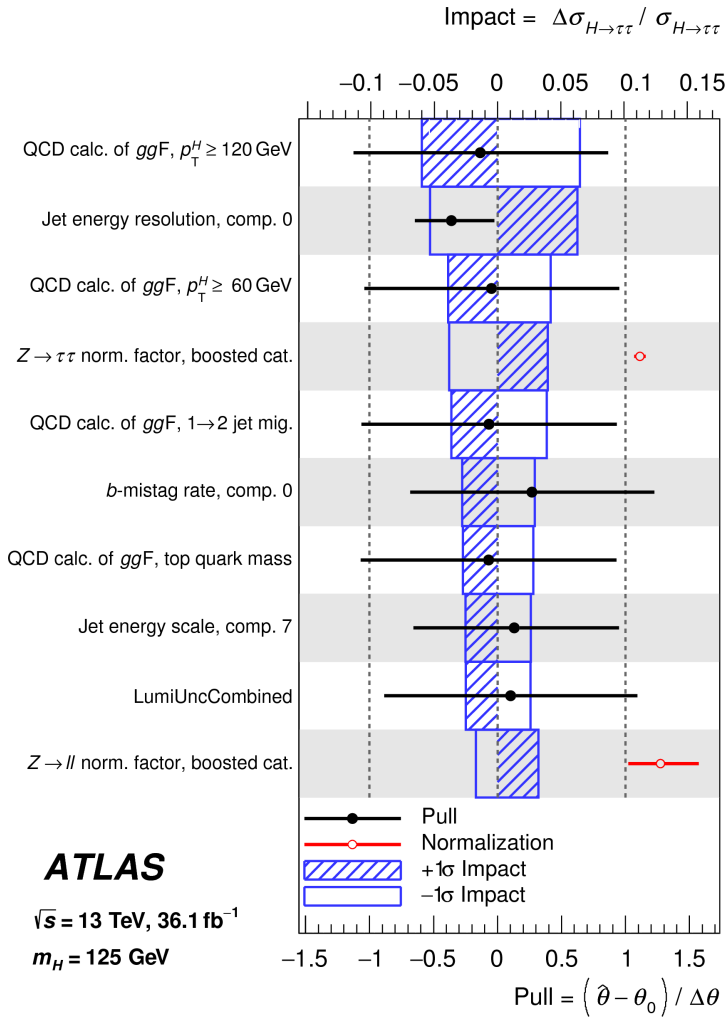


Figure 5.35: Observed impact of nuisance parameters on $\sigma_{H\to\tau\tau}$ for the combined $H \to \tau\tau$ fit. The uncertainties are listed in decreasing order of their impact on $\sigma_{H\to\tau\tau}$ and only the 10 highly ranked uncertainties are shown. The hatched blue and the open blue boxes represent the $+1\sigma$ and the -1σ uncertainties respectively. Their fractional impact is shown referring to the top x-axis. The pulls and constraints of the uncertainties are indicated by the black points and lines, referring to the bottom x-axis.

the other Higgs production mode are set to their SM prediction. Figure 5.37 shows the 68% and 95% confidence level contours in the plane of $\sigma^{\text{VBF}}_{H\to\tau\tau}$ and $\sigma^{\text{VBF}}_{H\to\tau\tau}$. The measured values for the cross sections are $\sigma^{\text{VBF}}_{H\to\tau\tau} = 0.28 \pm 0.09$ (stat.) $^{+0.11}_{-0.09}$ (syst.) pb and $\sigma^{\text{ggF}}_{H\to\tau\tau} = 3.1 \pm 1.$ (stat.) $^{+1.6}_{-1.3}$ (syst.) pb. Both the cross sections are in agreement with the SM predictions: $\sigma^{\text{VBF, SM}}_{H\to\tau\tau} = 0.237 \pm 0.006$ pb and $\sigma^{\text{ggF, SM}}_{H\to\tau\tau} = 3.05 \pm 0.13$ pb.
Lastly, a fit with three parameters of interest has been performed, considering two mutually exclusive regions in the ggF phase space defined according to different Higgs p_{T}

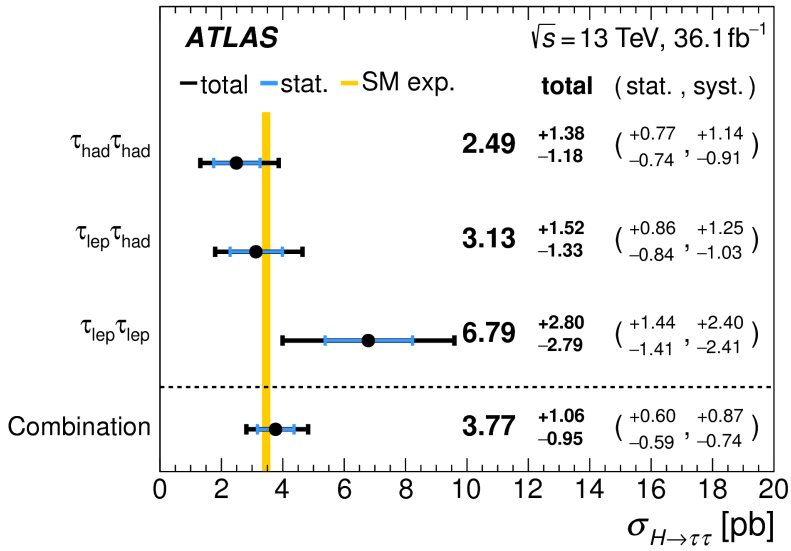


Figure 5.36: Measured values of $\sigma_{H\to\tau\tau}$ both for the individual channels and for the combined fit. The theory prediction together with its uncertainty is shown in yellow.

ranges as shown in Table 5.46 and one inclusive VBF region. The choice of selections is based on the simplified template cross section framework [134]. This framework has been developed in order to maximize the sensitivity of the measurements and to allow cross section measurements in mutually exclusive regions of phase space. Indeed, the phase space is split in different regions according to some kinematic properties of the events: for instance, number of jets and Higgs transverse momentum are used for defining different regions enriched in ggF events and kinematic properties of the jets are used for defining regions which target VBF events.
Table 5.46 shows the results for the measured cross section in the three exclusive phase space regions together with their SM predictions. This constitutes the first attempt of simplified template cross section measurements in the $H \to \tau\tau$ channel.

Process	Particle-level selection	σ [pb]	σ^{SM} [pb]
ggF	$N_{\mathrm{jets}} \geq 1$, $60 < p_{\mathrm{T}}^{H} < 120\,\mathrm{GeV}$, $\lvert y_H \rvert < 2.5$	$1.79 \pm 0.53\,(\mathrm{stat.}) \pm 0.74\,(\mathrm{syst.})$	0.40 ± 0.05
ggF	$N_{\mathrm{jets}} \geq 1$, $p_{\mathrm{T}}^{H} > 120\,\mathrm{GeV}$, $\lvert y_H \rvert < 2.5$	$0.12 \pm 0.05\,(\mathrm{stat.}) \pm 0.05\,(\mathrm{syst.})$	0.14 ± 0.03
VBF	$\lvert y_H \rvert < 2.5$	$0.25 \pm 0.08\,(\mathrm{stat.}) \pm 0.08\,(\mathrm{syst.})$	0.22 ± 0.01

Table 5.46: Measurements of VBF and ggF cross sections in three mutually exclusive regions, which follow the simplified template cross section framework. The measured cross section are shown along with the SM predictions.

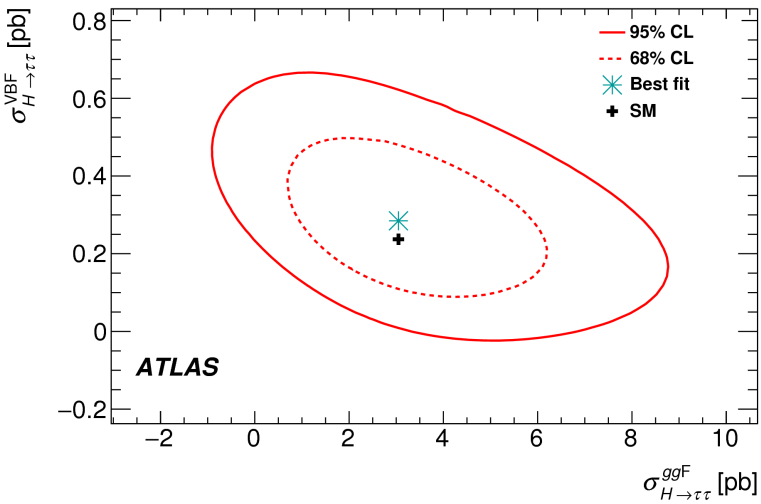


Figure 5.37: 68% and 95% likelihood contours in the ($\sigma^{\mathrm{VBF}}_{H\to\tau\tau}$, $\sigma^{\mathrm{ggF}}_{H\to\tau\tau}$). The best-fit values for the two cross sections are shown as a star while the black point represent the SM expectation.

5.10 Conclusion

The cross section measurements performed in the $H \to \tau\tau$ decay channel, using data collected by ATLAS at $\sqrt{13}$ TeV and corresponding to 36.1 fb^{-1} of integrated luminosity, have been illustrated in this Chapter with particular emphasys on the fully hadronic τ decay subchannel. The signal excess for the $H \to \tau\tau$ process over the expected background has been established with an observed (expected) significance of 4.4 (4.1) standard deviations. This result, combined with the one obtained using data taken at $\sqrt{s} = 7$ and 8 TeV during Run 1 of the LHC, reaches an observed (expected) significance of 6.4 (5.4) standard deviations and this constitutes the first $H \to \tau\tau$ ATLAS observation.
The measured signal strength is $\mu = 1.09^{+0.18}_{-0.17}$ (stat.) $^{+0.26}_{-0.22}$ (syst.) $^{+0.16}_{-0.11}$ (theory syst.), which is fully compatible with the Standard Model expectations. The measured total cross section was found to be $\sigma_{H\to\tau\tau} = 3.77^{+0.60}_{-0.59}$ (stat.) $^{+0.87}_{-0.74}$ (syst.) pb which is also in agreement with the SM prediction of $\sigma^{\mathrm{SM}}_{H\to\tau\tau} = 3.46 \pm 0.13$ pb.
In addition, measurements of the VBF and ggF cross sections separately and cross section measurements, based on the simplified template cross section, have been carried out and they also have found to be consistent with the SM expectations.
For all these measurements, the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ decay channel contribute significantly providing high sensitivity and also being the one with the smallest impact of systematic and statistical uncertainties.

Chapter 6

CP violation measurement in VBF production

This Chapter illustrates the search for CP violation performed in the $H \to \tau\tau$ channel. This search exploits many of the selections and techniques developed in the cross section measurement described in Chapter 5. It differs mainly because of concentrating on the VBF category only and by making use of a multivariate event selection to improve the purity of the VBF sample.

6.1 Introduction

After the Higgs boson discovery, as mentioned in Chapter 5, the studies have been concentrated on measuring its properties. Beside the coupling strengths, it is fundamental to measure the spin and CP properties of the Higgs as well, in order to determine wether the observed Higgs boson has the properties predicted by the SM, namely it is a CP-even scalar particle.
Moreover, the investigation of the Higgs sector at the LHC offers a new opportunity to search for CP violation. As explained in Chapter 1, one of the unsolved issue of physics today is the observed baryon asymmetry of the Universe. The presence of CP violation is one of the three Sakharov [7] conditions for explaining the generation of baryon asymmetry. In the SM, the CP violation is introduced via the complex phase in the CKM matrix. However the magnitude of CP violation in the SM is not sufficient to explain the observed value of baryon asymmetry [32] and new sources of CP violation must be introduced.
It is then interesting to look for CP violation in the Higgs sector and, at the same time, since no such effect is expected from the SM, this would be an unequivocal sign of Beyond the Standard Model (BSM) physics.
The analysis described in this Chapter, aims at studying the coupling structure of the vertex where the Higgs couples to two electroweak gauge bosons (HVV), through the vector boson fusion production mode. The VBF production is characterized by events with well defined experimental signatures: at least two high-p_{T} jets well separated in pseudorapidity. This particular production process is promising for measuring the Higgs CP properties in the HVV vertex. The measurement is in principle independent of the specific Higgs decay channel since it is focused on the Higgs production vertex. However, it is useful to exploit the $H \to \tau\tau$ decay channel since it allows to select a sizeable amount of VBF events with a good signal to background ratio.

Measurements performed so far at the LHC indicate that the Higgs boson is a scalar particle and that its coupling structure is CP-even and consistent with the SM expectations [135, 136]. The HVV vertex, in particular, has been investigated in the $H \to ZZ$ decay channel using Run 1 data of the LHC by both the ATLAS and CMS experiments [136, 137]. During Run 2 some other measurements have been performed and they have set limits on Wilson coefficients in the framework of effective field theories, which will be described later in Section 6.2. These analysis concern the $H \to \gamma\gamma$ and the $H \to ZZ$ channels and they have been performed in ATLAS [138, 139] using measurements of differential cross sections of CP-even observables or measurements of event rates. However, they are not sensitive to a direct probe of CP invariance as this work aims. CMS has also investigated the HVV vertex in the $H \to VV$ decay and VH production and also in the VBF production using the $H \to \tau\tau$ decay channel [140, 141, 142, 143]. In addition, a test of CP invariance in VBF production using the $H \to \tau\tau$ decay channel, has been performed by ATLAS using Run 1 data at $\sqrt{s} = 7$ and 8 TeV [144]. The analysis constrained the parameter $\tilde{d}$, which describes the strength of CP-violation, to the interval $[-0.11, 0.05]$ with 68% CL. All the mentioned studies have not seen any sign of CP violation.
The analysis presented in this thesis [145], have been performed using Run 2 data collected during 2015 and 2016 at $\sqrt{s} = 13$ TeV and corresponding to 36.1 fb^{-1}. Since the $H \to \tau\tau$ decay channel is selected, the analysis follows closely the cross section measurement performed in the same decay channel, meaning that the same samples, object selections and backgrounds estimation have been used. Differences and specific features of this analysis will be highlighted time by time.

6.2 Theoretical framework

The theoretical background is based on an Effective Field Theory (EFT) framework, useful to describe deviations from the SM in the most possible model independent way. As described in Section 1.5, the SM Lagrangian can be extended using dimension six operators, preserving in this way the $SU(2) \times U(1)_Y$ gauge symmetry. The Lagrangian can be constructed from the Higgs doublet Φ and the $U(1)_Y$ and $SU(2)$ electroweak gauge fields, B^μ and $W^{a,\mu}(a = 1, 2, 3)$. Since we are interested in the HVV vertex, all the interactions between the Higgs boson and the other SM particles (fermions and gluons) are assumed as predicted by the SM. The resulting Lagrangian is then [146, 147],

$$\begin{aligned}\mathcal{L}_{eff} = \mathcal{L}_{\mathrm{SM}} + \frac{f_{\tilde{\mathrm{W}}\mathrm{W}}}{\Lambda^2}\mathcal{O}_{\tilde{\mathrm{W}}\mathrm{W}} + \frac{f_{\tilde{\mathrm{B}}\mathrm{B}}}{\Lambda^2}\mathcal{O}_{\tilde{\mathrm{B}}\mathrm{B}} + \frac{f_{\tilde{\mathrm{B}}}}{\Lambda^2}\mathcal{O}_{\tilde{\mathrm{B}}} \\ + \frac{f_{\mathrm{WW}}}{\Lambda^2}\mathcal{O}_{\mathrm{WW}} + \frac{f_{\mathrm{BB}}}{\Lambda^2}\mathcal{O}_{\mathrm{BB}} + \frac{f_W}{\Lambda^2}\mathcal{O}_{\mathrm{W}} + \frac{f_{\mathrm{B}}}{\Lambda^2}\mathcal{O}_{\mathrm{B}}\end{aligned} \tag{6.1}$$

where the dimension six operators, $\mathcal{O}$, have the following expressions (the CP-even are shown on the left column and the CP-odd on the right),

$$\begin{aligned}
&\mathcal{O}_{\mathrm{BB}} = \Phi^\dagger B_{\mu\nu} B^{\mu\nu} \Phi && \mathcal{O}_{\tilde{\mathrm{B}}\mathrm{B}} = \Phi^\dagger \tilde{\mathrm{B}}_{\mu\nu} B^{\mu\nu} \Phi \\
&\mathcal{O}_{\mathrm{WW}} = \Phi^\dagger W_{\mu\nu} W^{\mu\nu} \Phi && \mathcal{O}_{\tilde{\mathrm{W}}\mathrm{W}} = \Phi^\dagger \tilde{\mathrm{W}}_{\mu\nu} W^{\mu\nu} \Phi \\
&\mathcal{O}_{\mathrm{B}} = (\mathcal{D}_\mu \Phi)^\dagger B^{\mu\nu} (\mathcal{D}_\nu \Phi) && \mathcal{O}_{\tilde{\mathrm{B}}} = (\mathcal{D}_\mu \Phi)^\dagger \tilde{\mathrm{B}}^{\mu\nu} (\mathcal{D}_\nu \Phi) \\
&\mathcal{O}_{\mathrm{W}} = (\mathcal{D}_\mu \Phi)^\dagger W^{\mu\nu} (\mathcal{D}_\nu \Phi)
\end{aligned} \tag{6.2}$$

In Equation 6.2 the covariant derivative $\mathcal{D}_\mu$ and the field strength tensors, $\hat{B}_{\mu\nu}$ and $\hat{W}_{\mu\nu}$, of the W and B fields and their dual ones are defined as:

$$\begin{aligned}
&\mathcal{D}_\mu = \partial_\mu + \frac{i}{2} g' B_\mu + i g \tau^a W^a_\mu \\
&\hat{B}_{\mu\nu} + \hat{W}_{\mu\nu} = i \frac{g'}{2} B_{\mu\nu} + i g \tau^a W^a_{\mu\nu} \\
&\tilde{V}_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} V^{\rho\sigma}, (V = W, B)
\end{aligned} \tag{6.3}$$

Two other CP-even operators should be taken into account, $\mathcal{O}_{\mathrm{BW}} = \Phi^\dagger B_{\mu\nu} W^{\mu\nu} \Phi$ and $\mathcal{O}_{\Phi,1} = (\mathcal{D}_\mu \Phi)^\dagger \Phi \Phi^\dagger (\mathcal{D}^\mu \Phi)$ but they have been strongly constrained by electroweak precision measurements [148, 149, 150] and they will therefore be neglected.
After the electroweak symmetry breaking, the effective Lagrangian can be rewritten in the mass basis of the Higgs boson H, the photon A_μ, the W boson and the Z boson:

$$\begin{aligned}
\mathcal{L}_{eff} = {} & g_{H\gamma\gamma} H A_{\mu\nu} A^{\mu\nu} + g^{(1)}_{HZ\gamma} A_{\mu\nu} Z^\mu \partial^\nu H + g^{(2)}_{HZ\gamma} A_{\mu\nu} Z^{\mu\nu} \\
& + g^{(1)}_{HZZ} Z_{\mu\nu} Z^\mu \partial^\nu H + g^{(2)}_{HZZ} H Z_{\mu\nu} Z^{\mu\nu} + g^{(2)}_{HWW} H W^+_{\mu\nu} W^{\mu\nu}_- \\
& + g^{(1)}_{HWW} (W^+_{\mu\nu} W^\mu_- \partial^\nu H + W^-_{\mu\nu} W^\mu_+ \partial^\nu H) + \tilde{g}_{H\gamma\gamma} H \tilde{A}_{\mu\nu} A^{\mu\nu} \\
& + \tilde{g}_{HZ\gamma} H \tilde{A}_{\mu\nu} Z^{\mu\nu} + \tilde{g}_{HZZ} H \tilde{Z}_{\mu\nu} Z^{\mu\nu} + \tilde{g}^{(2)}_{HWW} H \tilde{W}^+_{\mu\nu} W^{\mu\nu}_-
\end{aligned} \tag{6.4}$$

The g_{HVV} are the coefficients of the CP-even operators while the $\tilde{g}_{HVV}$ are the coefficients of the CP-odd ones. Not all the couplings are independent due to the $SU(2) \times U(1)_Y$ invariance. Indeed the CP-even couplings can be expressed in terms of four dimensionless

parameters, d, d_B, $\Delta^Z_{g_1}$ and Δk_γ,

$$\begin{aligned}
g_{H\gamma\gamma} &= \frac{g}{2m_W}(d\sin^2\theta_w + d_B\cos^2\theta_w{}^2) \\
g^{(1)}_{HZ\gamma} &= \frac{g}{m_W}(\Delta g_1^Z \sin 2\theta_w - \Delta k_\gamma \tan\theta_w) \\
g^{(2)}_{HZ\gamma} &= \frac{g}{2m_W}\sin 2\theta_w (d - d_B) \\
g^{(1)}_{HZZ} &= \frac{g}{m_W}(\Delta g_1^Z \cos 2\theta_w + \Delta k_\gamma \tan^2\theta_w) \\
g^{(2)}_{HZZ} &= \frac{g}{2m_W}(d\cos^2\theta_w + d_B\sin^2\theta_w) \\
g^{(1)}_{HWW} &= \frac{gm_W}{m_Z^2}\Delta g_1^Z \\
g^{(2)}_{HWW} &= \frac{g}{m_W}d
\end{aligned} \tag{6.5}$$

and the CP-odd can be written using only 3 dimensionless parameters, $\tilde{d}$, $\tilde{d}_B$ and $\tilde{k}_\gamma$,

$$\begin{aligned}
\tilde{g}_{H\gamma\gamma} &= \frac{g}{2m_W}(\tilde{d}\sin^2\theta_w + \tilde{d}_B\cos^2\theta_w) \\
\tilde{g}_{HZ\gamma} &= \frac{g}{2m_W}(-\tilde{k}_\gamma\tan\theta_w + \sin 2\theta_w(\tilde{d} - \tilde{d}_B)) \\
\tilde{g}_{HZZ} &= \frac{g}{2m_W}(\tilde{k}_\gamma\tan^2\theta_w + \tilde{d}\cos^2\theta_w + \tilde{d}_B\sin^2\theta_w) \\
\tilde{g}_{HWW} &= \frac{g}{m_W}\tilde{d}
\end{aligned} \tag{6.6}$$

where θ_w is the Weinberg angle and g is the $SU(2)$ coupling constant.
The relations between these dimensionless parameters and the coefficients f_i in Equation 6.1 are the following:

$$\begin{aligned}
d &= -\frac{m_W^2}{\Lambda^2}f_{WW} & \tilde{d} &= -\frac{m_W^2}{\Lambda^2}f_{\tilde{W}W} \\
d_B &= -\frac{m_W^2}{\Lambda^2}\frac{\sin^2\theta_w}{\cos^2\theta_w}f_{BB} & \tilde{d}_B &= -\frac{m_W^2}{\Lambda^2}\frac{\sin^2\theta_w}{\cos^2\theta_w}f_{\tilde{B}B} \\
\Delta k_\gamma &= k_\gamma - 1 = \frac{m_W^2}{2\Lambda^2}(f_B + f_W) & \tilde{k}_\gamma &= \frac{m_W^2}{2\Lambda^2}f_{\tilde{B}} \\
\Delta g_1^Z &= g_1^Z - 1 = \frac{m_Z^2}{\Lambda^2}\frac{f_W}{2}
\end{aligned} \tag{6.7}$$

Δg_1^Z, Δk_γ and $\tilde{k}_\gamma$ are the coefficients that parametrise the anomalous triple gauge boson couplings and that have been constrained by measurements at LEP [151, 152].
The effective Lagrangian yields the following Lorentz invariant structure for the HVV

vertex:

$$\begin{aligned} T^{\mu\nu}(p_1,p_2) &= a_1(p_1,p_2)g^{\mu\nu} && \text{(SM : CP even)} \\ &+ a_2(p_1,p_2)[p_1\cdot p_2 g^{\mu\nu} - p_2^{\mu}p_1^{\nu}] && \text{(CP even)} \\ &+ a_3(p_1,p_2)\epsilon^{\mu\nu\rho\sigma}p_{1\rho}p_{2\sigma} && \text{(CP odd)} \end{aligned} \tag{6.8}$$

where the a_i are momentum dependent form factors and the $p_{1,2}$ denote the momenta of the electroweak gauge bosons. The first term is CP-even and describes the SM coupling structure with $a_1 = \frac{2m_V^2}{v}$ ($V = W, Z$ and v is the vacuum expectation value) while the a_2 and the a_3 terms are CP-even and CP-odd respectively and arise from the dimension six CP-even and CP-odd operators.
The aim of the analysis described in this Chapter is to look for CP violation in the HVV vertex, therefore the CP-odd couplings are the only ones taken into account, more specifically a CP-mixed model, where a CP-odd contribution is considered in addition to the SM coupling, is investigated. In such a model, the form factors in the tensor structure of the vertex are:

$$a_1 = \frac{2m_V^2}{v}, a_2 = 0, a_3 \neq 0 \tag{6.9}$$

The g_{HVV} coefficients are then reduced to the CP-odd ones and, neglecting the $\tilde{k}_\gamma$ parameter, they depend only on $\tilde{d}$ and $\tilde{d}_B$:

$$\begin{aligned} \tilde{g}_{H\gamma\gamma} &= \frac{g}{2m_W}(\tilde{d}\sin\theta_w{}^2 + \tilde{d}_B\cos\theta_w{}^2) \\ \tilde{g}_{HZ\gamma} &= \frac{g}{2m_w}(\sin 2\theta_w(\tilde{d} - \tilde{d}_B)) \\ \tilde{g}_{HZZ} &= \frac{g}{2m_w}(\tilde{d}\cos\theta_w{}^2 + \tilde{d}_B\sin\theta_w{}^2) \\ \tilde{g}_{HWW} &= \frac{g}{m_w}\tilde{d} \end{aligned} \tag{6.10}$$

Experimentally, contributions to the VBF production from WW, ZZ, WZ, $Z\gamma$ or $\gamma\gamma$ are not distinguishable and thus the ratio between HZZ and HWW couplings has to be determined. In the analysis the choice $\tilde{d} = \tilde{d}_B$ is adopted. From the theoretical point of view this choice allows to recover the custodial symmetry of the SM [153]. Before the electroweak symmetry breaking, the Higgs potential has a global $SU(2)_L \times SU(2)_R$ symmetry which reduces to $SU(2)_V$ when the symmetry is broken. This remanent global symmetry is called the custodial symmetry. If the gauge symmetry of the electroweak model is broken by the Higgs doublet there is a custodial symmetry which protects the relation between the Z and W masses:

$$\rho \equiv \frac{M_W^2}{M_Z^2\cos^2\theta} = 1 \tag{6.11}$$

The ρ parameter is exactly equal to 1 at tree-level but anyway radiative corrections violate very softly the symmetry.

In addition, the choice $\tilde{d} = \tilde{d}_B$ allows to compare the results with that of other ATLAS measurements [136, 139] which make the same assumption altough using a different parametrisation. This assumption yields to the following expressions for the $\tilde{g}_{HVV}$ coefficients:

$$\begin{aligned} \tilde{g}_{H\gamma\gamma} &= \tilde{g}_{HZZ} = \frac{1}{2}\tilde{g}_{HWW} = \frac{g}{2m_W}\tilde{d} \\ \tilde{g}_{HZ\gamma} &= 0 \end{aligned} \tag{6.12}$$

In addition, the a_3 form factor in Equation 6.8 can be written as:

$$a_3 = \frac{2g}{M_V}\tilde{d} \quad (V = W, Z, \gamma) \tag{6.13}$$

The corresponding matrix element $\mathcal{M}$ for VBF production is then the sum of the SM term and a CP-odd contribution arising from the dimension six operators considered:

$$\mathcal{M} = \mathcal{M}_{\text{SM}} + \tilde{d} \cdot \mathcal{M}_{\text{CP-odd}} \tag{6.14}$$

where the dependence on $\tilde{d}$ has been factorized out. For the squared matrix element, one obtains:

$$|\mathcal{M}|^2 = |\mathcal{M}_{\text{SM}}|^2 + \tilde{d} \cdot 2\,\text{Re}(\mathcal{M}^*_{\text{SM}}\mathcal{M}_{\text{CP-odd}}) + \tilde{d}^2 \cdot |\mathcal{M}_{\text{CP-odd}}|^2 \tag{6.15}$$

The first term $|\mathcal{M}_{\text{SM}}|^2$ and the third term $|\mathcal{M}_{\text{CP-odd}}|^2$ are CP-even and they do not constitute a source of CP violation. The interference term $\tilde{d} \cdot 2\,\text{Re}(\mathcal{M}^*_{\text{SM}}\mathcal{M}_{\text{CP-odd}})$ is CP-odd instead and represents a possible new source of CP violation in the Higgs sector. The interference term integrated over a CP symmetric phase space vanishes and does not contribute to the total cross section while the third term is responsible for increasing the total cross section by an amount quadratic in $\tilde{d}$. Therefore the presence of a CP violating coupling influences the event yield as well but this is not exploited in this analysis as also CP conserving BSM couplings can affect it.

6.3 The Optimal Observable

In order to test CP invariance and to measure the adimensional parameter $\tilde{d}$, a variable sensitive to the presence of a CP-odd coupling must be used. If a BSM contribution is present, the effects should be visible in the distributions of the two tagging jets (the two jets involved in the VBF production, experimentally one usually choose the two jets with the highest p_{T}). Most of these distributions, like the jet transverse momentum, the jet energy or the dijet invariant mass, could depend strongly on the form factors and, therefore, are hard to predict without specifying the underlying model of new physics.
A simple CP-odd observable, not sensitive to form factor effects, is the signed difference in the azimuthal angle between the two tagging jets, defined as the azimuthal angle of the "away" jet minus the angle of the "towards" jet, where away and towards are referred to

the angular distance from the beam axis. The following expression provides the sign of $\Delta\phi_{jj}$:

$$\epsilon_{\mu\nu\rho\sigma} b_+^\mu p_+^\nu b_-^\rho p_-^\sigma = 2p_{T,+}p_{T,-}\sin(\phi_+ - \phi_-) = 2p_{T,+}p_{T,-}\sin\Delta\phi_{jj} \tag{6.16}$$

where b_+ and b_- denote the four-momenta of the two proton beams and p_+ and p_- denote the four-momenta of the two jets, in addition $p_+(p_-)$ points into the same detector hemisphere as $b_+(b_-)$. The $\Delta\phi_{jj}$ has a quasi-flat distributions in the case of SM while in the presence of a CP odd contribution it is asymmetric and more pronounced in the $0-\pi$ region as shown in Figure 6.1.

An alternative discriminating variable is the so-called Optimal Observable [8, 9]. In the past, this variable has been used in various analysis at LEP [154, 155, 156], in particular by the OPAL collaboration for studies of CP invariance in the $Z \to \tau\tau$ process [157]. The final state of the VBF Higgs production can be characterized by seven phase space variables assuming the mass of the Higgs boson, neglecting the jet masses and considering the transverse momentum conservation. The concept of Optimal Observable combines the information of this seven-dimensional phase space into a single observable. From the squared matrix element in Equation 6.15 we can define a first order and a second order Optimal Observables as follows:

$$\begin{aligned} OO_1 &:= \frac{2\,\mathrm{Re}(\mathcal{M}^*_{\mathrm{SM}}\mathcal{M}_{\text{CP-odd}})}{|\mathcal{M}_{\mathrm{SM}}|^2} \\ OO_2 &:= \frac{|\mathcal{M}_{\text{CP-odd}}|^2}{|\mathcal{M}_{\mathrm{SM}}|^2} \end{aligned} \tag{6.17}$$

The first order Optimal Observable has a mean value of zero and a symmetric distribution in the case of SM while it has a non-vanishing mean value, shifted to positive values for positive $\tilde{d}$ and to negative values for negative $\tilde{d}$, and an asymmetric distribution in the presence of CP violation, as can be seen in Figure 6.2. It is indeed a CP odd variable and can be used for a direct model independent test of CP invariance, by measuring its mean value. A non-zero mean value would indicate the presence of BSM physics. In addition, a fit to the full distribution can be used for measuring the $\tilde{d}$ parameter. The OO_1 provides the highest sensitivity for small $\tilde{d}$ values, performing better than the $\Delta\phi_{jj}$ distribution as demonstrated in Reference [144].

The second order Optimal Observable is a CP-even variable instead and it is not exploited in this analysis as the combined use of the two observables does not provide any gain in sensitivity with the available statistics.

Since the first order Optimal Observable is the only variable used in the analysis, we will henceforth referred to as the Optimal Observable, denote by the symbol OO.

Experimentally the Optimal Observable can be reconstructed from the final state particles involved. The matrix element calculation requires the four momentum of the Higgs, which can be reconstructed from the two τ four momentum, the four momentum of the jets and the Bjorken x of the initial state partons which represent the fraction of the proton momentum carried by the partons. The Bjorken x can also be retrieved from the final state kinematics using the formula $x_{1,2}^{reco} = \frac{M_{final}}{\sqrt{s}} e^{\pm y_{final}}$ where M_{final} (y_{final}) is

the invariant mass (rapidity) obtained from the vectorial sum of the four-momenta of the two tagging jets and the Higgs boson. Since the flavor of the initial and final state partons cannot be determined experimentally, the sum over all possible configuration weighted by the parton distribution functions is calculated separately for the numerator and the denominator:

$$\begin{aligned} 2\,\mathrm{Re}(\mathcal{M}^{*}_{\mathrm{SM}}\mathcal{M}_{\mathrm{CP-odd}}) &= \sum_{i,j,k,l} f_i(x_1) f_j(x_2) 2\,\mathrm{Re}((\mathcal{M}^{ij\to klH}_{\mathrm{SM}})^{*}\mathcal{M}^{ij\to klH}_{\mathrm{CP-odd}}) \\ |\mathcal{M}_{\mathrm{SM}}|^2 &= \sum_{i,j,k,l} f_i(x_1) f_j(x_2) |\mathcal{M}^{ij\to klH}_{\mathrm{SM}}|^2 \,. \end{aligned} \tag{6.18}$$

The matrix element leading order calculations are performed in the analysis with a code extracted from HAWK [158] and the PDFs used are the CT10 at leading order [122].

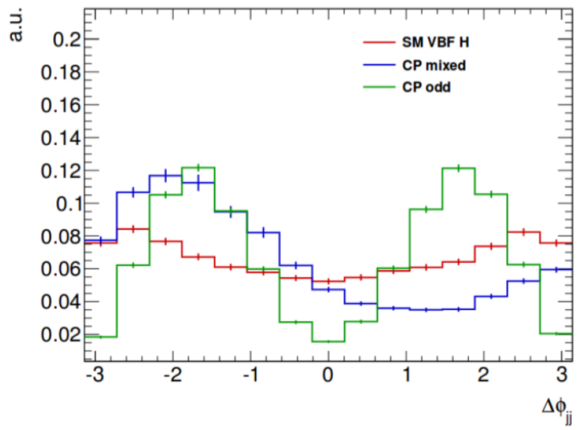


Figure 6.1: Distributions of the signed difference in the azimuthal angle between the two tagging jets. Distributions for the SM case, for a pure CP-odd scenario and for a CP-mixed scenario where a CP-odd contribution is considered in addition to the SM, are shown.

6.4 Data and Monte Carlo samples

The data used in this analysis are taken from pp collisions at $\sqrt{s} = 13$ TeV during 2015 and 2016, corresponding to an integrated luminosity of 3.2 fb^{-1} in 2015 and 32.9 fb^{-1}, for a total of 36.1 fb^{-1} integrated luminosity. Only the data which passed the data quality requirements are used. The same triggers as in the cross section measurement, summarised in Table 5.2, are used. Therefore, exactly the same dataset as in the cross section analysis is exploited.

Signal and background processes are simulated using various MC generators which do not differ from the ones reported in Table 5.1. The only difference with respect to the cross section analysis is the normalization of the electroweak $Z \to \tau\tau$ sample. Indeed, the cross

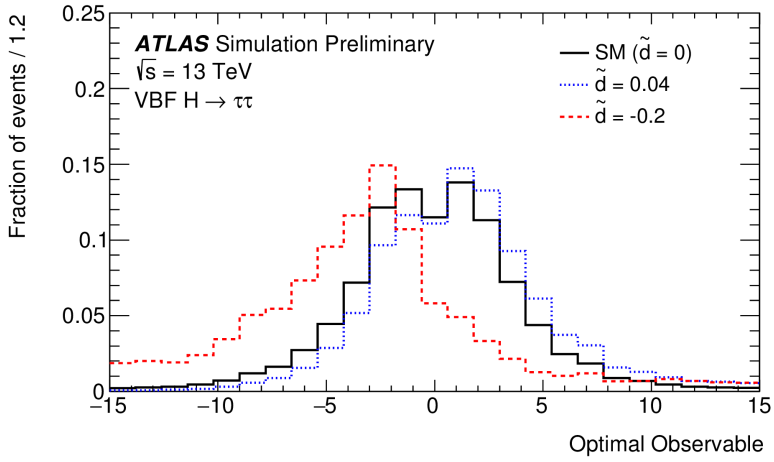


Figure 6.2: Optimal Observable distributions for three different scenarios: in the case of SM ($\tilde{d} = 0$) and for two different $\tilde{d}$ values. The distributions are obtained using the full event selection for defining the signal region of this analysis as described in Section 6.5 [145].

section of this process is underestimated by the SHERPA LO prediction and therefore it is scaled by a factor 1.7 in this analysis, in order to match the ATLAS measurement [159]. More details on this will be given in Section 6.6.1.

Differently to the coupling analysis only the VBF production process is considered as signal. In addition the $H \to WW$ is taken into account as signal in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel but the analysis is not sensitive to the possible CP violating coupling in the $H \to WW$ decay vertex. All the other Higgs production modes, ggF, VH, $t\bar{t}H$ are instead considered as backgrounds and their normalizations is fixed to their SM prediction.

The various BSM samples simulating a CP violating scenario, where a CP-odd contribution is added to the SM one, are obtained for different $\tilde{d}$ values using a reweighting technique. Indeed, the full simulation of these samples is quite expensive in terms of computational resources and time. The reweighting is applied to the VBF SM signal sample and the weights are calculated as the ratio of the squared matrix element of VBF process, considering a CP-mixed scenario as in Equation 6.15, to the squared matrix element obtained in the SM (weight $= \frac{|\mathcal{M}|^2}{|\mathcal{M}_{\text{SM}}|^2}$). At parton level three kind of events contribute to the VBF production: $qq \to qqH$, $qg \to qq\bar{q}H$ and $qq \to qqgH$. The weights are then calculated separately for the $2 \to 2 + H$ and the $2 \to 3 + H$ processes, using LO matrix elements extracted from HAWK. The matrix elements are evaluated using truth-level information for each event, namely the Bjorken x, the Higgs four-momentum, the four-momentum of the final state partons as well as the flavor of the incoming and outgoing partons. The validation of this reweighting procedure has been performed for the Run 1 measurement [144, 160]. It was achieved by comparing a reweighted VBF sample with a fully simulated NLO sample. The MADGRAPH5_aMC@NLO generator was used since it is able to simulate events with anomalous couplings at NLO. Figure 6.3

shows the Optimal Observable distribution for a generated and a reweighted sample for $\tilde{d} = 0.1$. The agreement between the two samples is good, meaning that the reweighting constitutes a good approximation of a full NLO description of the process.

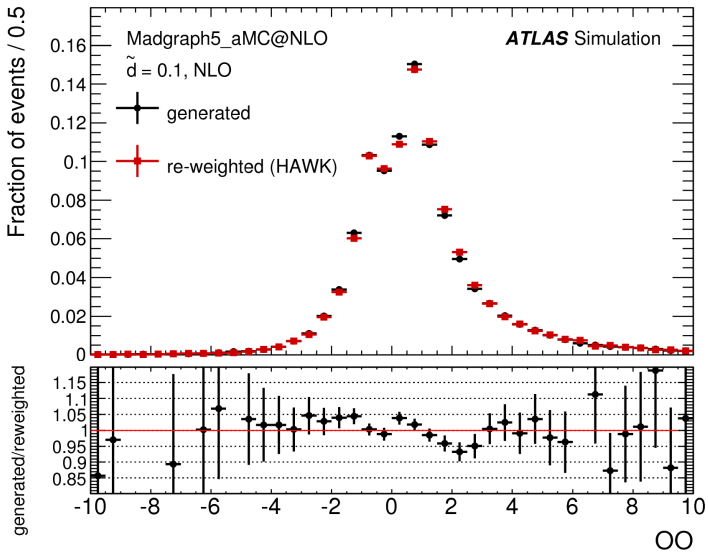


Figure 6.3: Optimal Observable distributions for a sample generated with MADGRAPH5_aMC@NLO with $\tilde{d} = 0.1$ (in black) and for a SM sample generated with MADGRAPH5_aMC@NLO after reweighting (in red) [160, 161].

6.5 Event selection

The analysis exploits all the combinations of leptonic ($\tau \to l\nu\bar{\nu}$ with $l = e, \mu$) and hadronic ($\tau \to$ hadrons ν) τ decays. Four different channels are then used: $\tau_{\text{lep}}\tau_{\text{lep}}$ with same flavor (SF) final state (ee, $\mu\mu$), $\tau_{\text{lep}}\tau_{\text{lep}}$ with different flavor (DF) final state ($e\mu$), $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$. The object definitions and the triggers correspond exactly to what has been already described in Section 5.3 and they will not be repeated here.
After the object definition and triggers, some preselection requirements are applied. They also follow very closely what has been done for the cross section measurement and they are summarised in Table 6.1. An exact number of τ decay candidates are required to avoid the overlap between the four channels, i.e. electrons, muons or $\tau_{\text{had-vis}}$. The two τ decay candidates must be of opposite charge and have to fulfill the trigger requirements. A minimal requirement on the missing transverse momentum $E_{\text{T}}^{\text{miss}} > 20$ GeV is applied in order to reject events without neutrinos. This requirement is tightened for the $\tau_{\text{lep}}\tau_{\text{lep}}$ SF channel to suppress the $Z \to ll$ background. Furthermore, a requirement on the $E_{\text{T}}^{\text{miss,hard}}$ is also introduced in this channel to reject contributions to the $E_{\text{T}}^{\text{miss}}$ not coming from physics objects. A requirement on the mass of the dilepton system is applied in $\tau_{\text{lep}}\tau_{\text{lep}}$ to further reject the $Z \to ll$ background. The di-τ mass calculated in the collinear

approximation is required to be $m_{\tau\tau}^{\text{coll}} > m_Z - 25$ GeV in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel to ensure ortoghonality between this analysis and the $H \to WW \to l\nu l\nu$ analysis. In the $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels the top-quark background is suppressed by requiring to not have b-quark jets. Low transverse mass is required in the $\tau_{\text{lep}}\tau_{\text{had}}$ channels to reject W+jets background. Finally, requirements on the angular distance between the visible decay products of the two τ and their pseudorapidity difference are applied in the $\tau_{\text{had}}\tau_{\text{had}}$ channel to reject non-resonant background.
In this analysis we are interested in selecting VBF events. A VBF inclusive category is then defined based on the same VBF selection applied in the cross section analysis. Two high-p_{T} jets are required with a large invariant mass ($m_{jj} > 300$ GeV) and a large pseudorapidity gap ($|\Delta\eta_{jj}| > 3$). This selection is also summarised in Table 6.1.

6.5.1 Multivariate technique

Differently from the cross section analysis, a multivariate technique (MVA) is used to construct signal regions enriched in VBF events. The purpose of this MVA method is to use informations from different kinematics variable, exploiting their correlations, in order to separate between signal and background events. The algorithm used in the analysis is the Boosted Decision Tree (BDT) [162]. The BDT is a supervised machine learning algorithm which uses example inputs in order to produce a desired output. A typical task in particle physics is to classify events as signal-like or background-like. In order to solve this binary classification problem, the BDT uses labeled training dataset to provide classification for another independent dataset.
The BDT is based on decision trees which are flowchart-like structures with a consecutive set of nodes (Figure 6.4). They take a set of input features, where each node represents one of those features and the data are split into further nodes accordingly. For example, in our case the nodes represent experimental cuts and the events are split accordingly, i.e. events that pass or fail a given trheshold. The final output is reached after a given maximum number of nodes. A metric that can be used to quantify how much the event is signal- or background-like is the signal purity, $p_s = \frac{s}{s+b}$. The signal purity can be evaluated at each node and used to build the decision tree. The final nodes, called "leaves", are then assigned an output value according to the purity, where $p_s > 0.5$ means that the leaf is dominated by signal events.
However a single decision tree is a weak classifier as it can be potentially unstable (a small modification of the inputs can induce large differences in the output). Boosting is a method to construct a stronger algorithm with a small misclassification rate by combining a set of decision trees. Misclassified events receive an higher weight as the series is trained, therefore the subsequent tree can focus on those events in the next training. The final result is given by averaging the output of each decision tree, weighted by its accuracy. A final classification score in the range [-1,1] is then the result of the BDT. Values closer to 1 mean that the event is likely a signal event, as the majority of the trees labeled it as signal.

In this analysis a BDT classifier as implemented in TMVA package [163] is used. Since there are four independent analysis channels, four different BDTs were trained. The vari-

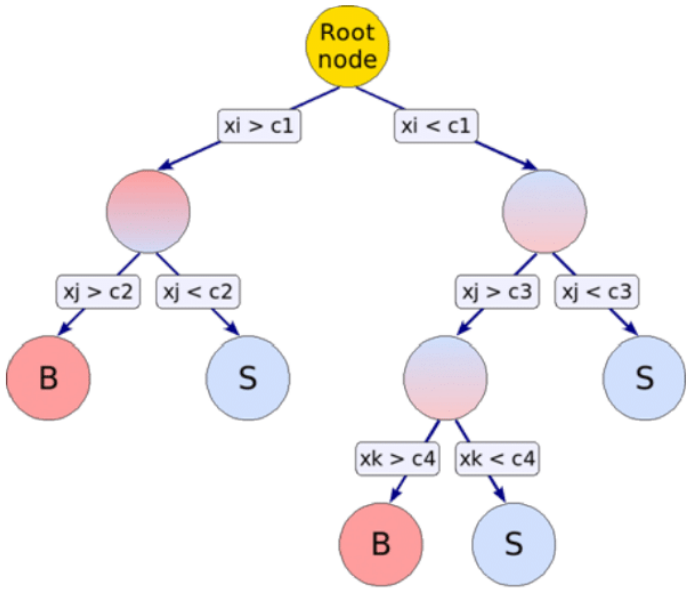


Figure 6.4: Schematic representation of a decision tree.

ables used as inputs can be categorised in variables referring to properties of the Higgs boson, which help to reject background processes without a Higgs boson, variables referring to properties of the di-τ system, which help to discriminate against non-resonant backgrounds and variables referring to the specific properties of the VBF topology. The detailed list of the input variables is the following and they are also summarised in Table 6.1:

- $m_{\tau\tau}^{\text{MMC}}$: the invariant $\tau\tau$ mass reconstructed with the MMC algorithm.
- $m_{\tau\tau}^{\text{vis}}$: the visible mass calculated from the two τ-candidates.
- $p_{\text{T}}^{\tau\tau,E_{\text{T}}^{\text{miss}}}$: the transverse momentum of the Higgs candidate.
- $m_{\text{T}}(\tau_1, E_{\text{T}}^{\text{miss}})$: the transverse mass calculated from the leading τ-candidate and the missing transverse energy.
- $\Delta R_{\tau\tau}$: the angular distance between the two τ-candidates.
- $\Delta\eta_{\tau\tau}$: the pseudorapidity separation between the two τ-candidates.
- $\Delta\phi_{\tau\tau}$: the angular distance in ϕ-direction between the two τ-candidates.
- $E_{\text{T}}^{\text{miss}}$ centrality, $C(\phi^{\text{miss}}/\sqrt{2})$: the transverse plane is transformed such that the direction of the tau decay products are orthogonal, and that the smaller ϕ angle between the tau decay products defines the positive quadrant of the transformed plane. The $E_{\text{T}}^{\text{miss}}$ centrality is then defined as the sum of the x- and y-components of the $E_{\text{T}}^{\text{miss}}$ unit vector in this transformed plane,

$$C(\phi^{\text{miss}}/\sqrt{2}) = \frac{r+s}{\sqrt{r^2+s^2}}$$

where

$$r = \frac{\sin(\phi(E_{\text{T}}^{\text{miss}}) - \phi(\tau_0))}{\sin(\phi(\tau_1) - \phi(\tau_0))} \text{ and } s = \frac{\sin(\phi(\tau_1) - \phi(E_{\text{T}}^{\text{miss}}))}{\sin(\phi(\tau_1) - \phi(\tau_0))}$$

This is used as it helps to distinguish signal events from W decays. Indeed, when the only neutrinos in the event are due to τ decays, their total momentum should be in the first quadrant and centrality is greater than 1. Smaller values are evidence of the visible tau system recoiling against a hard neutrino, like the ones coming from W decays.

- $E_{\text{T}}^{\text{miss}}/p_{\text{T}}(\tau_1)$: the transverse momentum ratio between the $E_{\text{T}}^{\text{miss}}$ and the leading τ candidate.
- $E_{\text{T}}^{\text{miss}}/p_{\text{T}}(\tau_2)$: the transverse momentum ratio between the $E_{\text{T}}^{\text{miss}}$ and the subleading τ candidate.
- m_{jj}: the invariant mass of the two leading jets.
- $p_{\text{T}}^{\text{tot}}$: the transverse momentum of the system of all objects in the final state (the two τ-candidates, the two leading jets and the $E_{\text{T}}^{\text{miss}}$).
- $p_{\text{T}}(j_3)$: the transverse momentum of the third jet.
- $C_{jj}(\tau_1)$: the η centrality of the leading τ candidate with respect to the pseudorapidity of the two leading jets.
- $C_{jj}(\tau_2)$: the η centrality of the subleading τ candidate with respect to the pseudorapidity of the two leading jets.

The latest two variables, $C_{jj}(\tau_1)$ and $C_{jj}(\tau_2)$ are defined as:

$$C_{\eta_1,\eta_2}(\eta) = \exp\left[\frac{-4}{(\eta_1-\eta_2)^2}\left(\eta - \frac{\eta_1+\eta_2}{2}\right)^2\right] \tag{6.19}$$

where η, η_1 and η_2 are the pseudorapidities of the object and the two leading jets, respectively. This variable has a value of unity when the object is halfway in η between the two jets, 1/e when the object is aligned with one of the jets and $< 1/\text{e}$ when the object is not between the jets in η.
For each channel, the variables with the highest discrimination power were chosen among those listed above and it was checked that adding more variables does not have an impact on the sensitivity. As an example, the input features used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are listed in Table 6.2 and their distributions in the VBF region can be seen in Figure 6.5. Finally, Figure 6.6 shows the BDT score output in the same channel. Good agreement is observed between data and Monte Carlo prediction and all the small deviations are

Channel	$\tau_{\text{lep}}\tau_{\text{lep}}$ SF	$\tau_{\text{lep}}\tau_{\text{lep}}$ DF	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
Preselection	Two isolated τ-lepton decay candidates with opposite electric charge			
	$p_{\text{T}}^{\tau_1} > 19^*/15^*$ GeV (μ/e)	$p_{\text{T}}^{e} > 18$ GeV	$p_{\text{T}}^{\tau_{\text{had}}} > 30$ GeV	$p_{\text{T}}^{\tau_1} > 40$ GeV
	$p_{\text{T}}^{\tau_2} > 10/15^*$ GeV (μ/e)	$p_{\text{T}}^{\mu} > 14$ GeV	$p_{\text{T}}^{\tau_{\text{lep}}} > 21$ GeV	$p_{\text{T}}^{\tau_2} > 30$ GeV
	$m_{\tau\tau}^{\text{coll}} > m_Z - 25$ GeV		$m_{\text{T}} < 70$ GeV	$0.8 < \Delta R_{\tau\tau} < 2.5$
	$30 < m_{ll} < 75$ GeV	$30 < m_{ll} < 100$ GeV		$\lvert\Delta\eta_{\tau\tau}\rvert < 1.5$
	$E_{\text{T}}^{\text{miss}} > 55$ GeV	$E_{\text{T}}^{\text{miss}} > 20$ GeV		$E_{\text{T}}^{\text{miss}} > 20$ GeV
	$E_{\text{T}}^{\text{miss, hard}} > 55$ GeV			
	$N_{b\text{-jets}} = 0$			
VBF topology	$N_{\text{jets}} \geq 2$, $p_{\text{T}}^{j_2} > 30$ GeV, $m_{jj} > 300$ GeV, $\lvert\Delta\eta_{jj}\rvert > 3$			
	$p_{\text{T}}^{j_1} > 40$ GeV			$p_{\text{T}}^{j_1} > 70$ GeV, $\lvert\eta_{j_1}\rvert < 3.2$
BDT input variables	$m_{\tau\tau}^{\text{MMC}}$, m_{jj}, $\Delta R_{\tau\tau}$, $C_{jj}(\tau_1)$, $C_{jj}(\tau_2)$, $p_{\text{T}}^{\text{tot}}$			
	$m_{\tau\tau}^{\text{vis}}$, $m_T^{\tau_1,E_{\text{T}}^{\text{miss}}}$, pT^{j_3}		$C(\phi^{\text{miss}})/\sqrt{2}$	
	$\Delta\phi_{\tau\tau}$	$E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{\tau_1}$, $E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{\tau_2}$	$m_{\tau\tau}^{\text{vis}}$, $\lvert\Delta\eta_{\tau\tau}\rvert$	$p_{\text{T}}^{\tau\tau E_{\text{T}}^{\text{miss}}}$, $\lvert\Delta\eta_{\tau\tau}\rvert$
Signal region	$\text{BDT}_{\text{score}} > 0.78$		$\text{BDT}_{\text{score}} > 0.86$	$\text{BDT}_{\text{score}} > 0.87$

Table 6.1: Summary of the event selection for the four analysis channels. The asterisk denotes the lowest p_{T} treshold which may increase due to the trigger used.

covered by the uncertainties. It is visibile that the background is peaked at -1, and the signal is concentrated at output greater than 0.8. Despite the small signal to background ratio of the VBF inclusive region, this last bin achieves an overall purity of $\sim 35\%$ and a significant signal excess over the dominant $Z \to \tau\tau$ and misidentified τ backgrounds.

Input Variable ($\tau_{\text{had}}\tau_{\text{had}}$)
$m_{\tau\tau}^{\text{MMC}}$
$C_{jj}(\tau_1)$
m_{jj}
$p_{\text{T}}^{\text{tot}}$
$C_{jj}(\tau_2)$
$C(\phi^{\text{miss}}/\sqrt{2})$
$p_{\text{T}}^{\tau\tau,E_{\text{T}}^{\text{miss}}}$
$\Delta R_{\tau\tau}$
$\Delta\eta_{\tau\tau}$

Table 6.2: Input BDT variables used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, listed in order of decreasing importance.

6.5.2 Signal and control regions

Signal regions (SRs) are defined by introducing cuts on the BDT score. The exact cuts are the result of a comprimise between sensitivity maximization and background statistics. Indeed, a sufficient background statistics is required in order to have positive entries for each background process in each bin of the SR and to avoid instabilities in the fit. For example, in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the quantity $\frac{s}{\sqrt{s+b+\sigma_b^2}}$ is used as a figure of merit for

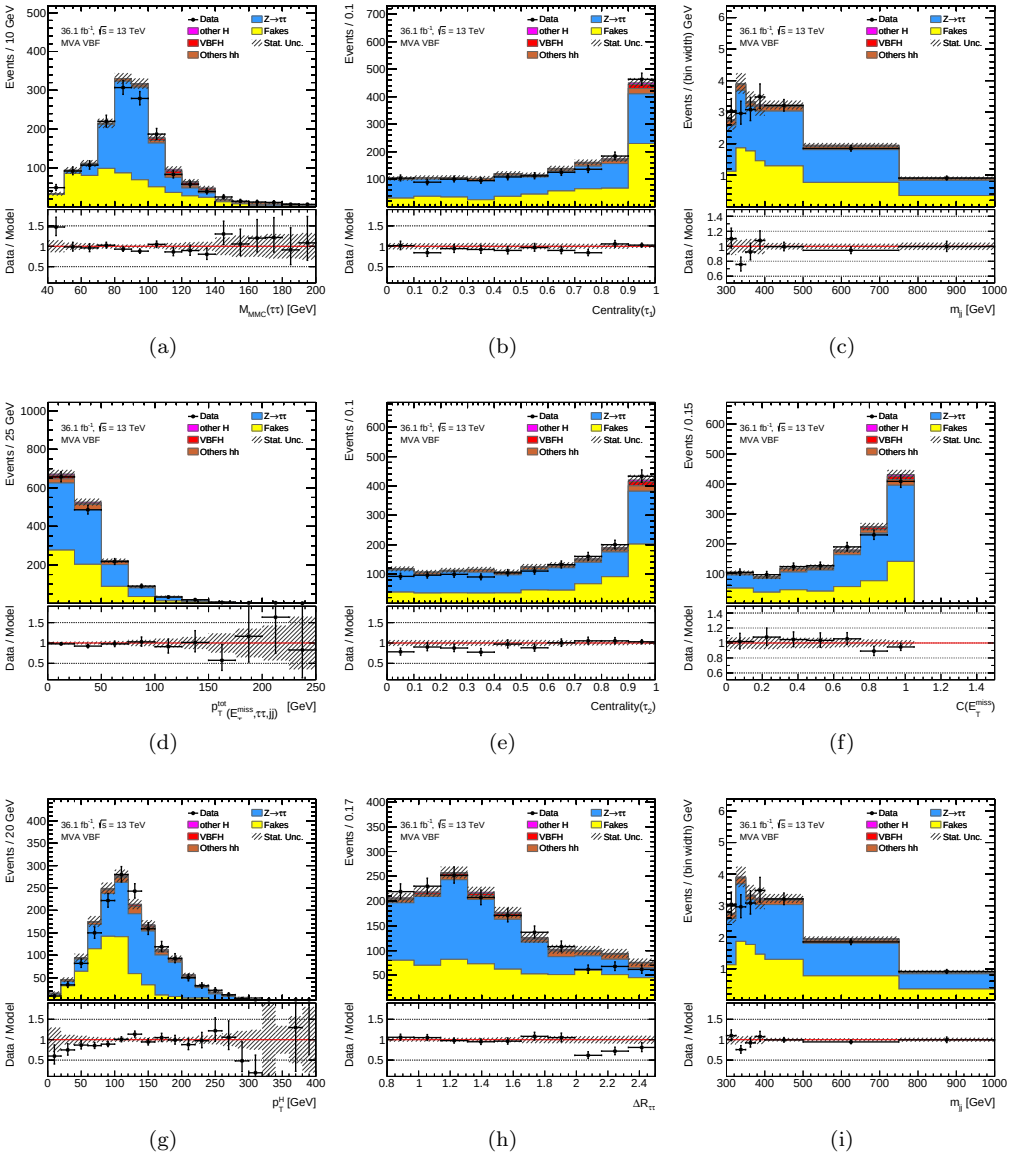
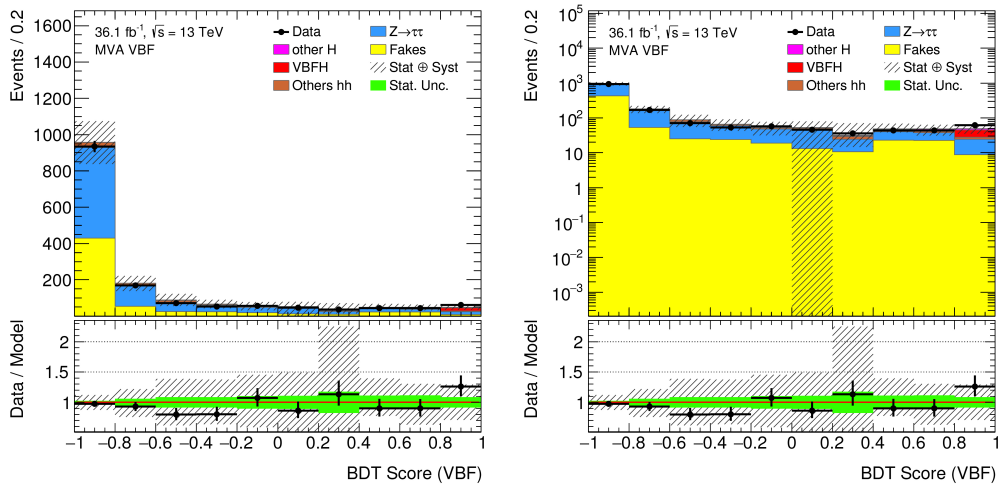


Figure 6.5: BDT input variables distributions in the VBF region in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The uncertainty band includes only statistical uncertainties.



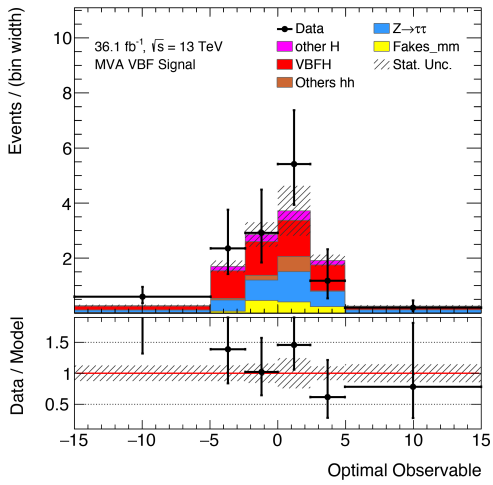
(a) (b)

Figure 6.6: BDT score distributions in the $\tau_{\text{had}}\tau_{\text{had}}$ channel shown in linear (a) and logarithmic (b) scales. The uncertainty band includes both statistical and systematic uncertainties.

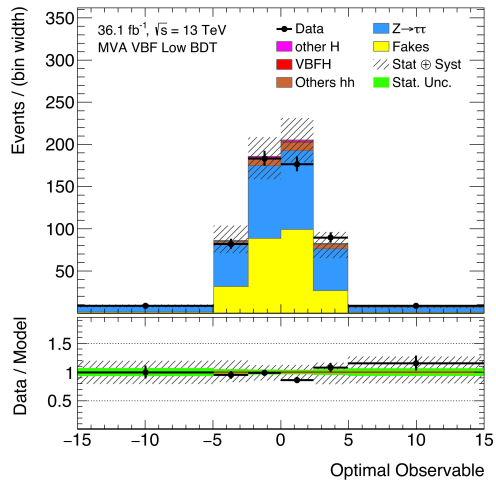
maximizing the sensitivity, where s and b are the signal and background event yields and σ_b is the statistical uncertainty for the backgrounds. This leads to a cut value of 0.90, however due to very small statistics of the misidentified τ background this cut is relaxed to 0.87. The signal region definitions can be found for all the analysis channels in Table 6.1.

A control region (CR) for each channel is defined by inverting the BDT score cut for the signal region. These regions will be referred to as the "low-BDT" CRs and they are fundamental for constraing the $Z \to \tau\tau$ background normalization and in the case of the $\tau_{\text{had}}\tau_{\text{had}}$ channel the misidentified τ background as well (see Section 6.6 for the background estimation). The contribution of $Z \to \tau\tau$ production amounts to 30-54% depending on the channel. In addition, in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel, four more CRs are added: two for the different flavor and two for the same flavor final state channels, for constraining the top-quark and the $Z \to ll$ backgrounds. The top-quark CR is obtained by inverting the veto on b-tagged jets and not applying the BDT score cut. The $Z \to ll$ CR is defined by again not applying the BDT score cut and requiring the dilepton invariant mass to be $80 < m_{ll} < 100$ GeV.

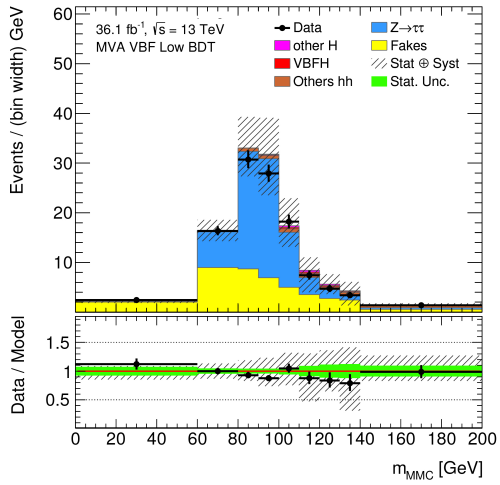
Distributions of the Optimal Observable in the $\tau_{\text{had}}\tau_{\text{had}}$ signal and control regions and of the $m_{\tau\tau}^{\text{MMC}}$ in the $\tau_{\text{had}}\tau_{\text{had}}$ control region are shown in Figure 6.7. The Optimal Observable is then used in the SR to probe for CP invariance. As demonstrated in Figure 6.8 no dependence of the mean value of the Optimal Observable on the BDT score is observed confirming that this cut does not introduce any bias, i.e. any artificial CP asymmetry. In addition, the $m_{\tau\tau}^{\text{MMC}}$ behaviour is shown as this will be the variable used in the CR in the fit. Good modelling within uncertainties is observed for the variables in the CR.



(a)



(b)



(c)

Figure 6.7: Optimal Observable distribution in the $\tau_{\text{had}}\tau_{\text{had}}$ signal region (a) and control region (b) and $m_{\tau\tau}^{\text{MMC}}$ distribution in the $\tau_{\text{had}}\tau_{\text{had}}$ control region (c).

6.6 Background estimation

The same backgrounds described in Section 5.6 for the cross section analysis are considered for this measurement as well and in principle the same estimation strategy is used. A brief summary of the backgrounds involved, highlighting the differences with respect to the cross section analyisis, will be given in this Section.

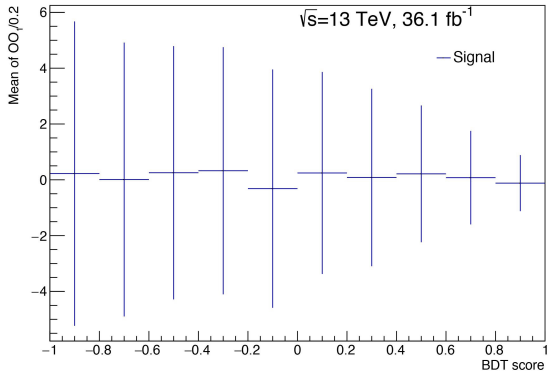


Figure 6.8: Mean values of the Optimal Observable for the signal for different BDT score cuts.

6.6.1 $Z \to \tau\tau$ background

The $Z \to \tau\tau$ Drell-Yan process is the main irreducible background common to all the channels. Its normalization is constrained using the low-BDT CR which provides a good purity for this process as explained in Section 6.5.2. The normalization of $Z \to \tau\tau$ is therefore directly retrieved from the fit to data. To ensure that the normalization is valid in the SR as well, the modelling of this process is validated in a validation region (VR). This VR is built in the same way as for the cross section analysis, therefore it is based on the $\tau_{\text{lep}}\tau_{\text{lep}}$ SF VBF selection and uses $Z \to ll$ events as it is difficult to select a pure $Z \to \tau\tau$ sample. The good modelling of Z boson and jets kinematic properties is then verifed in this region and some example of distributions are shown in Figure 6.9. As for the cross section measurement, a trend is observed in the m_{jj} variable and also in the p_{T} of the third jet. A reweighting based on the m_{jj} behaviour has therefore been applied. Altough the impact on the modelling of the $Z \to \tau\tau$ and on its normalization as well as on the final sensitivity of the analysis is small, it was decided to keep the reweighted $Z \to \tau\tau$ distribution.
The correct modelling of the $Z \to \tau\tau$ has been also investigated in a region more similar to the analysis signal regions, i.e. training new BDTs where the τ-related input variables were replaced by light leptons related variables and a high BDT region was selected by requiring $\text{BDT}_{\text{score}} > 0.6$. Some mismodelling was observed here for the BDT distribution and for some of the jet kinematic variables as shown in Figure 6.10. This mismodelling has been found to be larger where there is more significant contribution of the $Z \to \tau\tau$ electroweak production. This process is simulated in this analysis with Sherpa at LO with two or three partons in the final state. As found in the ATLAS $Z \to \tau\tau$ cross section measurement [159], the Sherpa LO prediction underestimates the cross section, while, for instance, the Powheg NLO prediction is closer to the measured cross section. This is summarised in Table 6.3. It has then been decided to scale the $Z \to \tau\tau$ electroweak production in order to match the ATLAS cross section measurement, since the analysis

VBF regions and the region where the $Z \to \tau\tau$ measurement has been performed are similar. The distributions in Figure 6.10 are shown before and after applying the $Z \to \tau\tau$ electroweak scaling and they exhibit a good improvement of the modelling.

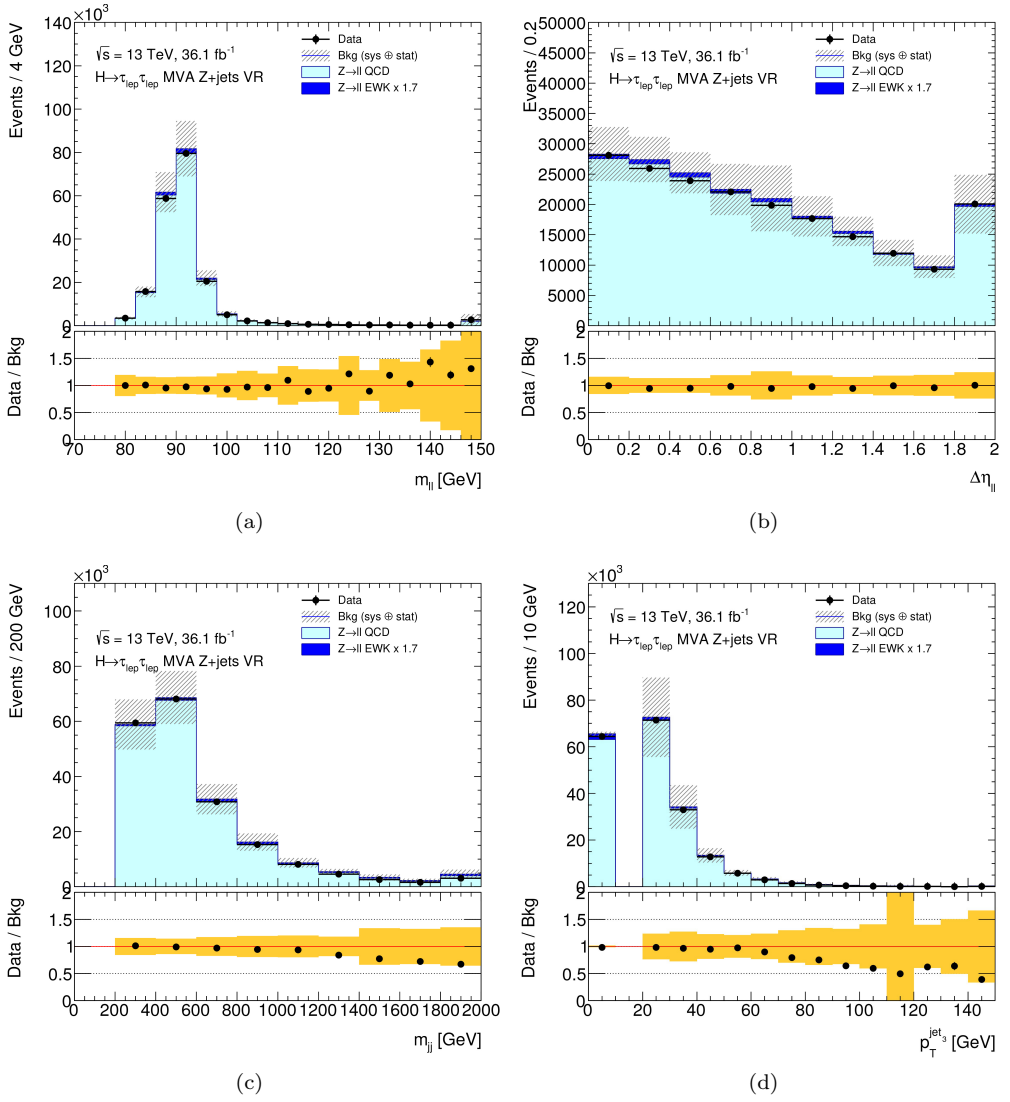
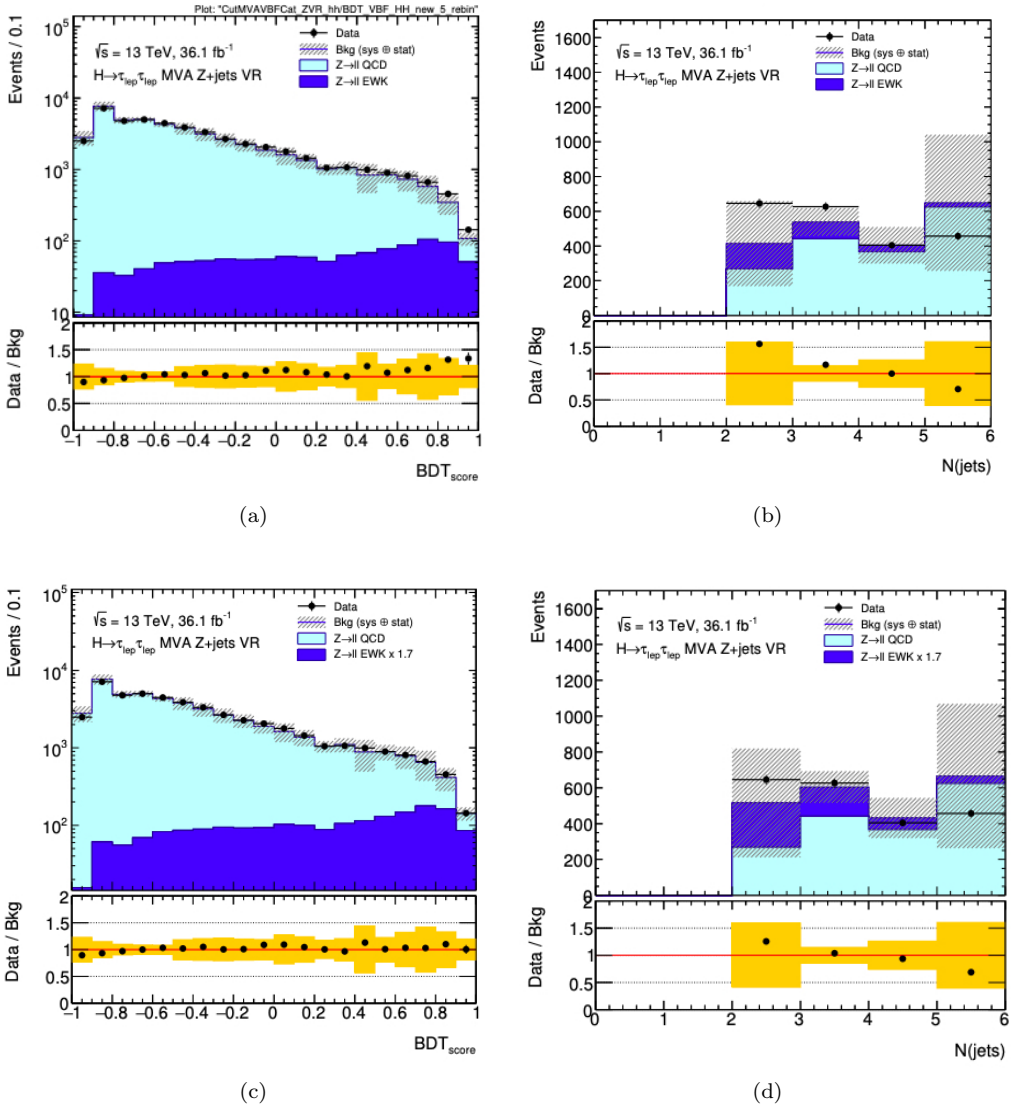


Figure 6.9: Distributions of some selection variables in the Z+jets VR of the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel: tdilepton invariant mass (a), pseudorapidity separation between the two leptons (b), dijet invariant mass (c) and third jet transverse momentum (c). Both statistical and systematic uncertainties are shown in the band in yellow.



(a) (b)

(c) (d)

Figure 6.10: Distribution of BDT score in the $Z \to \tau\tau$ validation region (a) and for the number of jets after selecting an high BDT region (b) ($\text{BDT}_{\text{score}} > 0.6$) in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The same distributions are shown at the bottom ((c) and (d)) after having applied a 1.7 scale factor to the $Z \to \tau\tau$ electroweak process.

6.6.2 Background from misidentified τ

Jets misidentified as $\tau_{\text{had-vis}}$ or as electrons or muons are the second largest contribution to backgrounds, particularly significant for the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channel. They arise mainly from QCD jets, W+jets production and semileptonically decaying $t\bar{t}$ events. The

σ_{EWK} [fb]	$\sigma_{\text{EWK, Sherpa LO}}$ [fb]	$\sigma_{\text{EWK, Powheg NLO}}$ [fb]
118.9	71.2 ± 0.2	$125.2 \pm 0.4(\text{stat.}) \pm 18.8(\text{syst.})$

Table 6.3: $Z \to \tau\tau$ electroweak producion cross section as measured by ATLAS [159] and compared with the SHERPA LO and POWHEG NLO prediction.

same data driven techniques described in Section 5.6.4 are applied for estimating this background. The normalization of the misidentified τ background is derived from CR extrapolation and fixed before performing the fit for the $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels, while in the $\tau_{\text{had}}\tau_{\text{had}}$ channel is directly retrieved from the fit.
Some differences with respect to the cross section analysis affect the $\tau_{\text{had}}\tau_{\text{had}}$ channel. In the cross section analysis the fake template for the $\tau_{\text{had}}\tau_{\text{had}}$ channel was derived in a dedicated control region (nOS) and then reweighted using scale factors dependent on the $\Delta\phi_{\tau\tau}$ between the two τ candidates. These scale factors were derived using a region closer to the signal region (OS anti-ID medium). However the differences between the nOS and the OS anti-ID medium regions found in the cross section analysis and shown in Figure 5.8, are not visible in Figure 6.11 where the $\Delta\phi_{\tau\tau}$ distribution is shown after applying the preselection of this analysis which differs a bit from the preselection of the cross section measurement. Thus, it was decided to not apply this reweighting and consequently the corresponding systematic uncertainty. Another difference concerns the misidentifed τ template used in the SR. Due to the limited statistics, a looser τ ID working point is used for deriving the misidentified τ template in the SR, meaning that a medium requirement for the τ_{had} candidates is used instead of the default tight ID working point. This affects however only the shape since the template derived with the medium ID requirement is scaled to the default template. An associated uncertainty is also applied for taking into account the differences between the two distributions in the SR. More details on this will be given in Section 6.11.

6.6.3 $Z \to ll$ and top-quark backgrounds

Dedicated CRs are used in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel for constraining the normalization of these backgrounds. Differently from the cross section analysis a top-quark CR is not used in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel because the different BDT selection led to a smaller contribution of the top-quark background in this channel. In the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels therefore the normalization is derived from MC simulations and fixed before performing the fit.

6.6.4 Other backgrounds

Minor backgrounds are simulated with Monte Carlo and their normalization is fixed to their theoretical prediction. They include diboson production and electroweak production of W/Z bosons. In addition top-quark production constitutes a minor background in the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels. An important difference to highlight with respect to the cross section analysis is that all the other Higgs production modes are also considered as

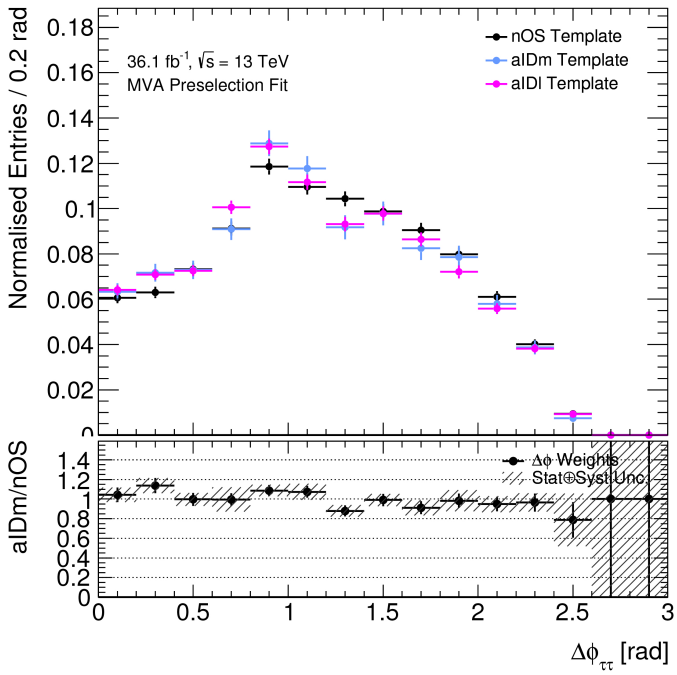


Figure 6.11: $\Delta\phi_{\tau\tau}$ distribution in dedicated control regions for the misidentified τ background estimation after applying the $\tau_{\text{had}}\tau_{\text{had}}$ preselection. The distributions are derived in the nOS region (black), in the OS anti-ID medium region (blue) and in the OS anti-ID loose region (pink). The bottom panel shows the ratio between the distributions in the nOS region and in the the OS anti-ID medium region. The uncertainty band in the ratio is constructed by replacing the medium-not-tight requirement on the subleading τ candidate with the loose-not-tight added in quadrature with the statistical uncertainty on the nominal weight.

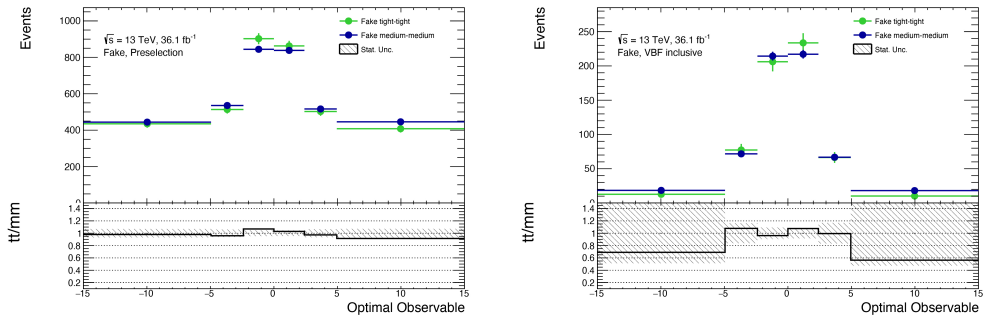
backgrounds. Their contribution is not significant and their normalization is also fixed to their SM prediction.

6.7 Systematic uncertainties

The same sources of systematic uncertainties described in Section 5.7 are considered here as well. They affect both the event yield and the shape of the Optimal Observable in the SRs and of the $m_{\tau\tau}^{\text{MMC}}$ in the CRs.
Both theoretical and experimental systematic uncertainties are taken into account. The theoretical uncertainties are evaluated exactly with the same strategy as for the cross section analysis. Theoretical uncertainties on the total cross section are evaluated for all the Higgs production modes considered in the analysis (VBF, ggF, VH and $t\bar{t}H$) while MC modelling systematics like QCD scale, PDFs and parton shower uncertainties are considered for the VBF signal process. No other theoretical systematics sources are con-

sidered for the Higgs production processes other than VBF, since they are negligible. The only exception is a jet veto uncertainty applied to the ggF process since the BDT cut acts like a third jet veto. This is evaluated using the Stewart-Tackmann method [115].
Theoretical uncertainties are evaluated as well for the $Z \to \tau\tau$ background with the same strategy as in the cross section analysis. These includes scale uncertainties, matching between the matrix element and parton shower, resummation scale uncertainties, PDFs and parton shower uncertainties. As for the cross section measurement, parton shower uncertainties, derived by comparing Sherpa and Madgraph5_aMC@NLO, have limited statistics (in particular for what concerns the Madgraph5_aMC@NLO sample) and therefore they are evaluated in the VBF region and then applied to the actual SRs. Acceptance and shape uncertainties arising from the m_{jj} reweigthing applied for the $Z \to \tau\tau$ sample are evaluated in each region by comparing the reweighted and the non-reweighted sample. Shape uncertainties for both the Optimal Observable and the $m_{\tau\tau}^{\text{MMC}}$ have found to be negligible and therefore only the acceptance uncertainties are considered.
Experimental uncertainties refer to luminosity, pile-up, jet related uncertainties, τ reconstruction, missing transverse momentum reconstruction, electrons and muons related uncertainties and finally uncertainty due to the data-driven estimation of the misidentified τ background. Particularly significant are jet energy scale and resolution, $\tau_{\text{had-vis}}$ energy scale and resolution, as well as $\tau_{\text{had-vis}}$ reconstruction and efficiency. The evaluation strategy is the same as in the cross section measurement except for some differences which will be explained in the following.
The first difference concerns the jet energy resolution uncertainties. The evaluation of these uncertainties considers some parameter variations in both data and MC and then compares data and MC prediction. However, the statistics for data in the SR is small and this led to very large variations causing negative event yields and consequently instabilities in the fit. It was then decided to evaluate these uncertainties keeping only the MC variations.
Another difference relates to the misidentified τ background in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. As already mentioned in Section 6.6.2, the $\Delta\Phi_{\tau\tau}$ reweighting is not applied in this analysis, therefore the uncertainty due to the usage of this method is not taken into account. The other major difference concerns the misidentified τ background template used in the $\tau_{\text{had}}\tau_{\text{had}}$ SR. This is derived by using a medium τ ID working point instead of the default tight ID requirement. Since the template derived using the medium working point is then scaled to the default template, only the shape of the Optimal Observable is affected. A shape uncertainty have been derived in order to take into account differences between the two templates. This uncertainty is estimated by deriving the bin by bin ratio between the two templates in the VBF region and then applying it to the SR. In order to enhance the statistics, a test has also been performed using the preselection region defined by the cuts in Table 6.1 and requiring in addition the subleading jet to have $p_{\text{T}} > 30$ GeV. Figure 6.12 shows the shape difference for the Optimal Observable between the two templates in the preselection and VBF region. The resulting shape uncertainty is small compared to the statistical uncertainty. Furthermore, the fit results did not show any difference between the two regions since this uncertainty is pruned away in both cases.
Figure 6.13 shows some examples of significant uncertainties among the experimental and

theoretical ones in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.



(a) (b)

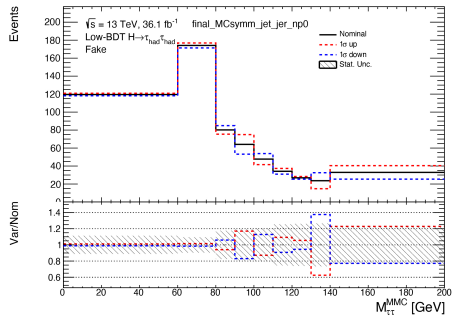
Figure 6.12: Optimal Observable distributions for the misidentifed τ background in the $\tau_{\text{had}}\tau_{\text{had}}$ channel when requiring the two τ candidates to fulfill the tight ID requirement (in green) or when requiring them to fulfill the medium ID requirement (in blue), either in the preselection region (a) and in the VBF region (b).

6.8 Statistical analysis

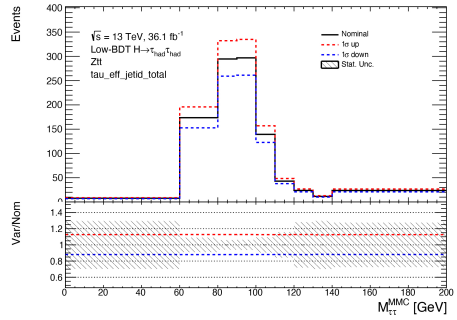
The statistical framework on which the analysis is based is the same described in Section 5.8.1.

The statistical analysis employs a binned likelihood function as well, $\mathcal{L}(x, \epsilon, \theta)$, which is a function of the measured data x, the parameter of interest ϵ and further nuisance parameters θ which takes into account the effect of systematic uncertainties, normalization factors for the backgrounds and bin-by-bin statistical uncertainties. The likelihood is built as the product of the Poisson probability terms for each bin in all the signal and control regions and the terms which accounts for the systematic uncertainties. The Optimal Observable distribution is used as the final discriminant in the SRs while the $m_{\tau\tau}^{\text{MMC}}$ distribution is employed in the CRs for constraining the $Z \to \tau\tau$ background in all the channels and the misidentified τ background in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. In addition, the event yields in the $Z \to ll$ and top CRs are used for constraining the normalization of these backgrounds in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel. A summary of all the regions employed in the fit is given in Table 6.4 with the exact binning used. The binning has been optimized by requiring enough statistics for the signal and backgrounds, in order for the fit to not show instabilities.

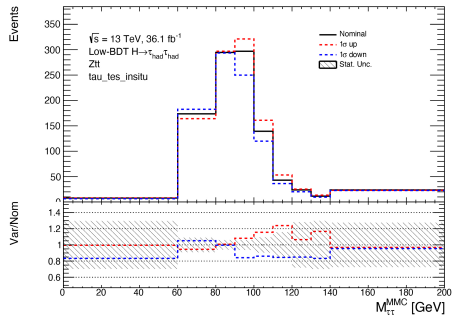
Only the VBF Higgs production mode is considered as signal while the other Higgs production processes (ggF, VH and $t\bar{t}H$) are considered together as backgrounds. The most significant backgrounds, $Z \to \tau\tau$ and the misidentified τ, are also considered separately. In addition, $Z \to ll$ and top-quark backgrounds are considered as additional separate backgrounds in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel. All the other backgrounds with minor contribution are considered together.


√s = 13 TeV, 36.1 fb⁻¹
final_MCsymm_jet_jer_np0
Low-BDT H→τhadτhad
Fake
Nominal
1σ up
1σ down
Stat. Unc.
Events
Var/Nom
$M_{\tau\tau}^{MMC}$ [GeV]

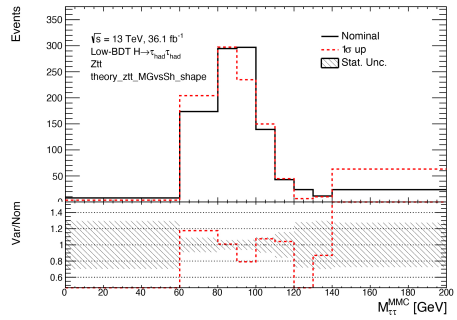
(a)


√s = 13 TeV, 36.1 fb⁻¹
Low-BDT H→τhadτhad
Ztt
tau_eff_jetid_total
Nominal
1σ up
1σ down
Stat. Unc.
Events
Var/Nom
$M_{\tau\tau}^{MMC}$ [GeV]

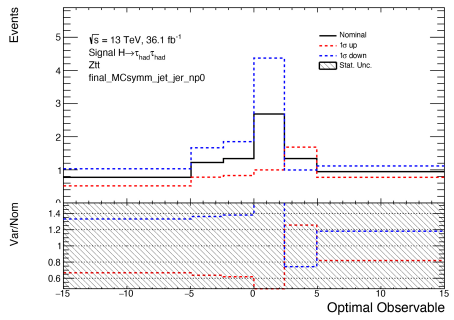
(b)


√s = 13 TeV, 36.1 fb⁻¹
Low-BDT H→τhadτhad
Ztt
tau_tes_insitu
Nominal
1σ up
1σ down
Stat. Unc.
Events
Var/Nom
$M_{\tau\tau}^{MMC}$ [GeV]

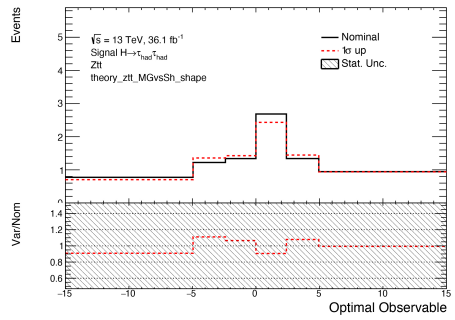
(c)


√s = 13 TeV, 36.1 fb⁻¹
Low-BDT H→τhadτhad
Ztt
theory_ztt_MGvsSh_shape
Nominal
1σ up
Stat. Unc.
Events
Var/Nom
$M_{\tau\tau}^{MMC}$ [GeV]

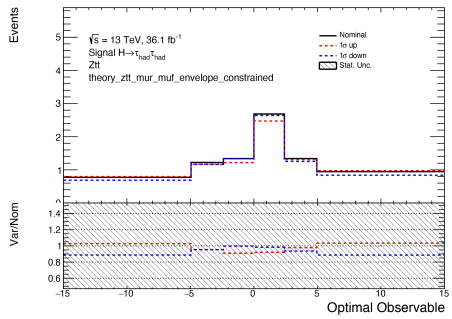
(d)


√s = 13 TeV, 36.1 fb⁻¹
Signal H→τhadτhad
Ztt
final_MCsymm_jet_jer_np0
Nominal
1σ up
1σ down
Stat. Unc.
Events
Var/Nom
Optimal Observable

(e)


√s = 13 TeV, 36.1 fb⁻¹
Signal H→τhadτhad
Ztt
theory_ztt_MGvsSh_shape
Nominal
1σ up
Stat. Unc.
Events
Var/Nom
Optimal Observable

(f)



(g)

Figure 6.13: Example of systematic uncertainties in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, affecting either acceptance and shape of the $m_{\tau\tau}^{\text{MMC}}$ in the low-BDT CR and acceptance and shape of the Optimal Observable in the SR. In particular the following systematics are shown: the leading jet energy resolution uncertainty component for the misidentified τ process in the low-BDT CR (a), one of the jet energy resolution uncertainty components for the $Z \to \tau\tau$ process in the low-BDT CR (b), tau energy scale uncertainty related to the tag and probe measurement for the $Z \to \tau\tau$ process in the low-BDT CR (c), the $Z \to \tau\tau$ parton shower uncertainty in the low-BDT CR (d), one of the jet energy resolution uncertainty components for the $Z \to \tau\tau$ process in the SR (e), the $Z \to \tau\tau$ parton shower uncertainty in the SR (f) and the $Z \to \tau\tau$ scale uncertainties in the SR (g). The dashed band indicates the statistical uncertainty. The distributions referring to the $Z \to \tau\tau$ parton shower uncertainty show only the$+1\sigma$ variation since this is the only one provided and then it is symmetrized before performing the fit. The $Z \to \tau\tau$ scale uncertainty in the SR (g) has been taken into account despite the large statistical uncertainty as the variation was considered a real physical effect and not negligible.

Region		Definition	Distribution, Binning
		$\tau_{\text{lep}}\tau_{\text{lep}}$	
SR	ee$\mu\mu$ eμ	$\text{BDT}_{\text{score}} > 0.78$	Optimal Observable, [-15,-5,0,5,15]
Low-BDT CR	ee$\mu\mu$ eμ	$\text{BDT}_{\text{score}} \leq 0.78$	$m_{\tau\tau}^{\text{MMC}}$, [0,80,100,120,140,200]
Top CR	ee$\mu\mu$ eμ	$N_{\text{b-jets}} \geq 1$	1-bin
$Z \to ll$ CR	ee$\mu\mu$	$80 < m_{ll} < 100\,\text{GeV}$	1-bin
		$\tau_{\text{lep}}\tau_{\text{had}}$	
SR		$\text{BDT}_{\text{score}} > 0.86$	Optimal Observable, [-15,-9,-6.3,-3,-0.9,0,0.9,3,6.3,9,15]
Low-BDT CR		$\text{BDT}_{\text{score}} \leq 0.86$	$m_{\tau\tau}^{\text{MMC}}$, [0,50,70,90,110,130,150,250]
		$\tau_{\text{had}}\tau_{\text{had}}$	
SR		$\text{BDT}_{\text{score}} > 0.87$	Optimal Observable, [-15, -4.95, -2.4, 0, 2.4, 4.95, 15]
Low-BDT CR		$\text{BDT}_{\text{score}} \leq 0.87$	$m_{\tau\tau}^{\text{MMC}}$, [0, 60, 80, 90, 100, 110, 120, 130, 140, 200]

Table 6.4: Summary of signal and control regions for all the channels included in the fit together with the binning used. The regions for the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel are further split into same flavor and different flavor regions with the same distributions and binning.

The nuisance parameters are parametrized using the same pdfs as described in Section 5.8.2. Gaussian distributions are used to parametrize systematic uncertainties while normalization factors are parametrized using log-normal distributions. Finally, the gamma distribution is used for the bin-by-bin statistical uncertainties, which are also encoded in the fit as NPs. Some backgrounds have their normalization free to float in the fit and they have associated normalization factors which the fit is able to determine. Specifically, one NF is associated to the $Z \to \tau\tau$ process correlated across all the channels, one NF is associated to the misidentified τ background correlated across all the $\tau_{\text{had}}\tau_{\text{had}}$ regions and two NFs are used in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel for the $Z \to ll$ and top-quark backgrounds. The fit is then executed in the same way as in the cross section measurement. A maximum likelihood fit is peformed, specifically using the test statistic t_μ given in Equation 5.21, where the tested hypothesis is the signal plus background model and the parameter of interest, extracted from the fit, is the signal strength. However the aim of this measurement is to estimate the parameter $\tilde{d}$. In order to achieve such a measurement, the fit is repeated several times replacing the SM VBF signal with the relevant CP-mixed sample, obtained by using the reweighting tecnique described in Section 6.4, and leaving the background model unchanged. A negative log-likelihood (NLL) curve is then constructed as a function of $\tilde{d}$ and confidence intervals can be determined by using the Neyman construction [164]. Indeed, it can be shown that in the large sample limit the likelihood function has a gaussian behaviour (therefore the logarithm of the likelihood has a parabolic shape) and it is possible to derive the following expression:

$$-\log L(\hat{\tilde{d}} \pm N\sigma) = -\log L_{\text{max}} + \frac{N^2}{2} \tag{6.20}$$

where the negative log-likelihood is chosen for convenience. This means a simple prescription for deriving intervals since changing the estimated parameter by N standard deviations, leads to a a decrease (increase) of $\frac{N^2}{2}$ of the log-likelihood (negative log-likelihood). The difference between the NLL evaluated for each $\tilde{d}$ hypothesis and the minimum found for the NLL (ΔNLL = NLL - NLL_{min}) is then evaluated and the confidence interval for $\tilde{d}$ is determined by reading off the points on the ΔNLL which reach a certain treshold.
The technical framework used is the same employed in the cross section analysis, therefore the same simplifications procedures described in Section 5.8.2 applies also here.

6.9 Results

The results obtained either with the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit and when performing a combined fit with the other channels will be presented in this Section. In order to optimise the analysis sensitivity and to test the analysis strategy, Asimov fits have been performed before looking at the actual data distributions and they are documented in Section 6.9.1 while the observed results are described in Section 6.9.2.

6.9.1 Expected results

Expected results are discussed in this Section and they are obtained using an Asimov dataset where the observed data are replaced by the SM prediction and all the NPs are set to their nominal values.

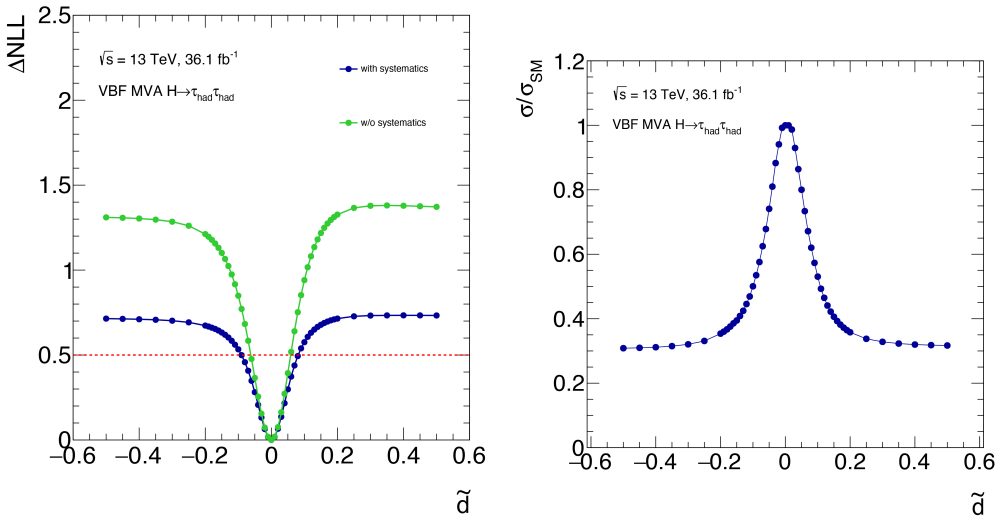
$H \to \tau_{\text{had}}\tau_{\text{had}}$

Expected results obtained with the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit are presented here. As described in Section 6.8, an expected ΔNLL curve is built in order to estimate confidence intervals for $\tilde{d}$. The ΔNLL is shown together with the best-fit signal strength in Figure 6.14 as a function of $\tilde{d}$. Since the fit is performed using an Asimov dataset, the signal strength is constrained very close to 1 which is the assumed value for the signal plus background hypothesis and decreases for less compatible CP-mixed models. The ΔNLL has the minimum value, corresponding to the $\tilde{d}$ parameter that describes data the best, at zero as expected by construction of the Asimov dataset, meaning the the most probable value for $\tilde{d}$ is obtained for the SM hypothesis. The ΔNLL has a parabolic shape in a region close to its minimum value while it reaches a plateau for higher $\tilde{d}$ values. This is a feature of the Optimal Observable of the first order which loses sensitivity when going to high $\tilde{d}$ values.
It is possible to determine a confidence interval for $\tilde{d}$ at 68% CL using only the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The resulted interval is $\tilde{d} \in [-0.090, 0.080]$. In Figure 6.14 it is also shown the behaviour of ΔNLL when no systematics uncertainties are included, meaning that the corresponding NPs are set to constant. The impact of systematic uncertainties is significant, resulting in a large drop of the plateau and in a $\sim 20\%$ decrease in sensitivity looking at the 68% CL interval.
The beahviour of NPs is also monitored by looking at their post-fit values and pulls.

This is done in the fit for $\tilde{d} = 0$, namely in the SM case. Figure 6.15 shows the NPs listed in decreasing order of impact on the parameter of interest, i.e. the signal strength. This result is obtained in the same way as for the cross section measurement by shifting a given NPs to its $\pm 1\sigma$ uncertainty and then repeating the fit for the other NPs. The resulting impact on the signal strength is then shown in the figure as well as the best-fit value for each NP determined by the fit. The highest ranked systematic is one of the jet energy resolution uncertainty components and generally among the ones with the largest impact many uncertainties are related to the jet or missing transverse momentum reconstruction. In addition, some of the tau related uncertainties, particularly the tau energy scale uncertainty on the tag and probe measurement and the $Z \to \tau\tau$ normalization, contribute significantly. As expected, all the parameters are not pulled and have their best-fit values at zero. Figure 6.16 also shows the pulls and constraints for all the NPs involved in the $\tau_{\text{had}}\tau_{\text{had}}$ standalone fit. As in the cross section analysis, constraints are observed for the leading component of jet energy resolution uncertainty ("jet_jer_np_0") and the tau energy scale uncertainty on the tag and probe measurement ("tau_tes_insitu"), and therefore they can be explained by the same reasons discussed in Section 6.9.1. A strong constraint is instead the one exhibit by the $Z \to \tau\tau$ parton shower uncertainty ("Ztt_theory_MGvsSH"). This uncertainty, shown in Figure 6.13 for either the signal and control regions, has a quite large variation. Due to the limited statistics, the shape uncertainty on the Optimal Observable is considered negligible by the fit while the shape uncertainty on the $m_{\tau\tau}^{\text{MMC}}$ is retained. Figure 6.18 shows again the $Z \to \tau\tau$ parton shower uncertainty of Figure 6.13, once with the statistical uncertainty of the default sample (SHERPA) and once with the statistical uncertainty of the MADGRAPH5_aMC@NLO sample. It can be seen how the MADGRAPH5_aMC@NLO sample, used for the evaluation of this systematics, is statistically limited. Furthermore, it can be observed that the uncertainty reaches up to 20% around the Z boson peak. These large variations are the origin of the constraint as data in the low-BDT CR have a high purity in $Z \to \tau\tau$ events. The same effect with a similar behaviour has also been observed in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel. Finally, Figure 6.17 shows the correlations between NPs which are higher than 0.25.
The impact of uncertainties split by categories is given in Table 6.5. The systematic uncertainties are higher than the statistical ones even if the latter are still significant. Among the systematics, jet and missing transverse energy related uncertainties and tau related uncertainties have a considerable contribution.

Combined

The results of the $\tau_{\text{had}}\tau_{\text{had}}$ channel are combined with those of the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ using a combined fit model as described in Section 6.8.
The resulted ΔNLL curve together with the best-fit signal strength are shown in Figure 6.19, where the various channel standalone fit results and the combined results are superimposed in the same plot. All the curves have their minimum at $\tilde{d} = 0$ as expected and the sensitivity gradually decreases when reaching high $\tilde{d}$ values. The combined result allows to determine the confidence interval $\tilde{d} \in [-0.032, 0.031]$ at 68% CL and to reach



(a) (b)

Figure 6.14: Expected ΔNLL curve (a) and best-fit signal-strength (b) as a function of $\tilde{d}$ for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit.

POI μ	Central Value 1
Set of nuisance parameters	Impact on uncertainty
Total	+0.68 / -0.68
Data statistics	+0.42 / -0.38
Total systematic unc.	+0.53 / -0.56
$Z \to \tau\tau$ normalization	+0.18 / -0.04
All normalizations	+0.19 / -0.05
All but normalizations	+0.53 / -0.56
Jets and $E_{\text{T}}^{\text{miss}}$	+0.38 / -0.46
Falvor tagging	+0.00 / -0.00
Electrons and muons	+0.00 / -0.00
Tau	+0.30 / -0.23
Pileup reweighting	+0.05 / -0.13
Misedintified τ	+0.00 / -0.00
Luminosity	+0.04 / -0.01
Signal theory unc.	+0.10 / -0.09
$Z \to \tau\tau$ theory unc.	+0.07 / -0.11
MC statistics	+0.17 / -0.23

Set of nuisance parameters	Fractional impact on uncertainty (square of fraction of total)
Data statistics	+0.39 / -0.31
Total systematic unc.	+0.61 / -0.69
$Z \to \tau\tau$ normalization	+0.07 / -0.00
All normalizations	+0.08 / -0.01
All but normalizations	+0.60 / -0.68
Jets and $E_{\text{T}}^{\text{miss}}$	+0.31 / -0.47
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.00 / -0.00
Tau	+0.19 / -0.11
Pileup reweighting	+0.01 / -0.04
Misedintified τ	+0.00 / -0.00
Luminosity	+0.00 / -0.00
Signal theory unc.	+0.02 / -0.02
$Z \to \tau\tau$ theory unc.	+0.01 / -0.03
MC statistics	+0.06 / -0.11

Table 6.5: Expected impact of uncertainties, grouped in categories, on the signal strength for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit assuming the VBF SM signal. On the left absolute values are shown while on the right the fractional contribution is given.

the 95% CL interval, which is found to be $\tilde{d} \in [-0.12, 0.10]$. The sensitivity is mainly driven by the $\tau_{\text{lep}}\tau_{\text{had}}$ and the $\tau_{\text{had}}\tau_{\text{had}}$ channels.

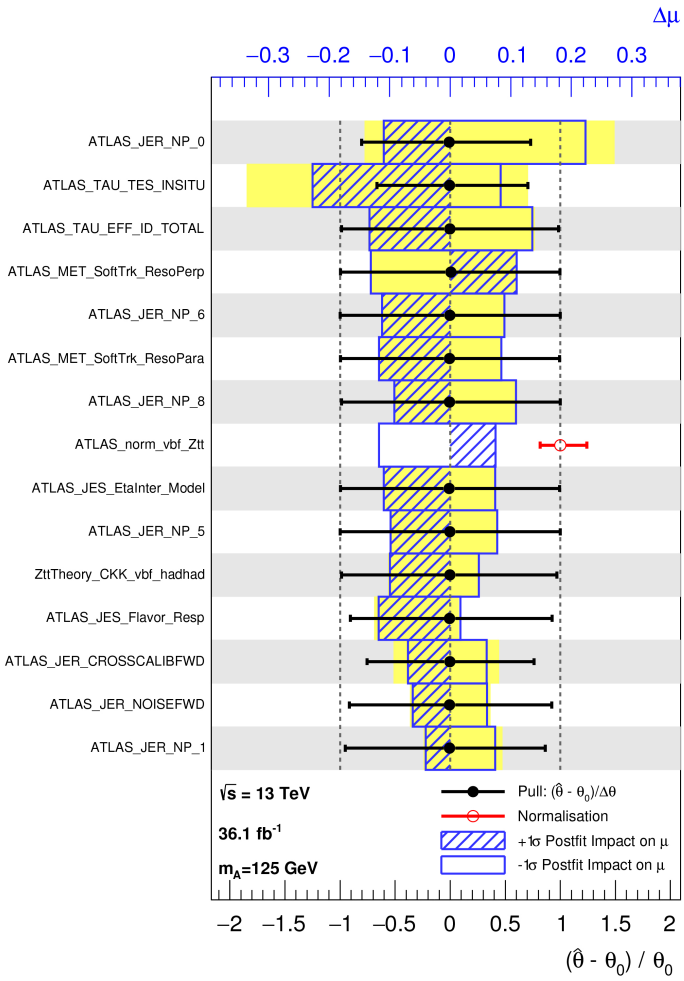


Figure 6.15: Expected impact of nuisance parameters on the signal strength for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit assuming the VBF SM signal. The uncertainties are listed in decreasing order of their impact on μ and only the 15 most highly ranked uncertainties are shown. The hatched blue and the open blue boxes represent the $+1\sigma$ and the -1σ uncertainties respectively while the yellow box represent the pre-fit ones. Their fractional impact is shown referring to the top x-axis. The pulls and constraints of the uncertainties are indicated by the black points and lines, referring to the bottom x-axis. If the uncertainty is not pulled nor constrainted the pull is 0 and the lines lie between $+1$ and -1.

Figure 6.20 shows the NPs with the highest impact on the signal strength. Jet energy resolution uncertainties, as well as VBF parton shower uncertainties have a high impact. The large impact of the latter is driven by the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. In addition, jet energy scale, muon reconstruction and tau reconstruction appear among the uncertainties with

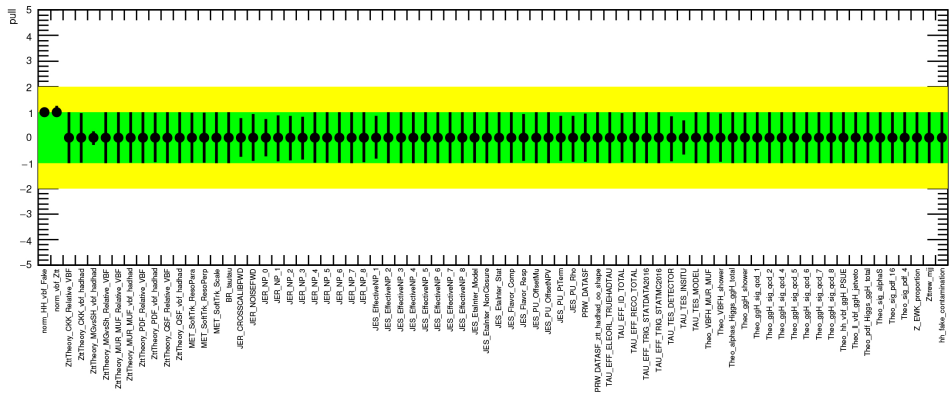


Figure 6.16: Expected nuisance parameters pulls and constraints for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit assuming the VBF SM signal.

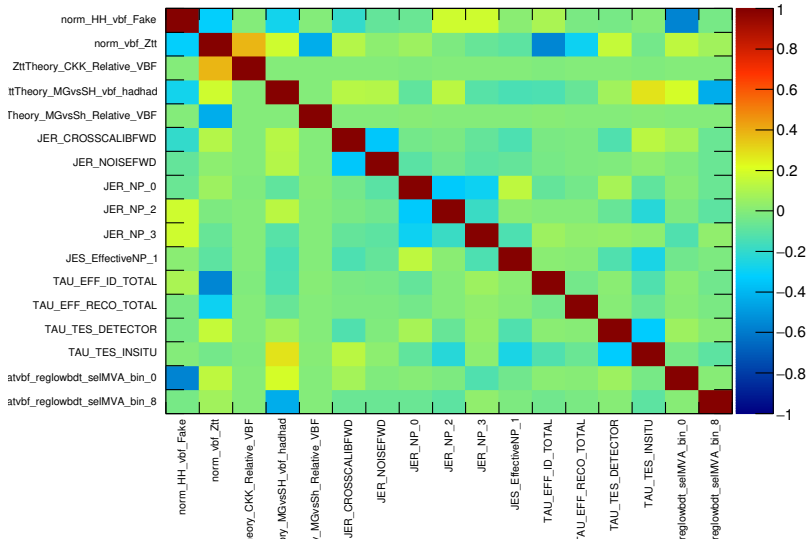
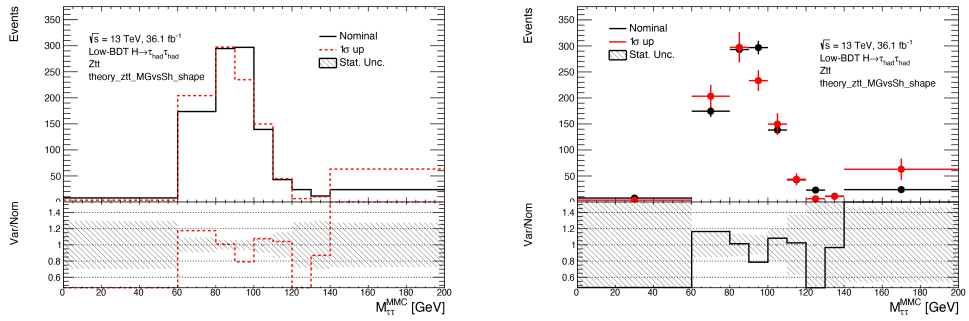


Figure 6.17: Expected nuisance parameters correlation for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit assuming the VBF SM signal. Only the nuisance parameters with absolute correlation higher than 0.25 are shown.

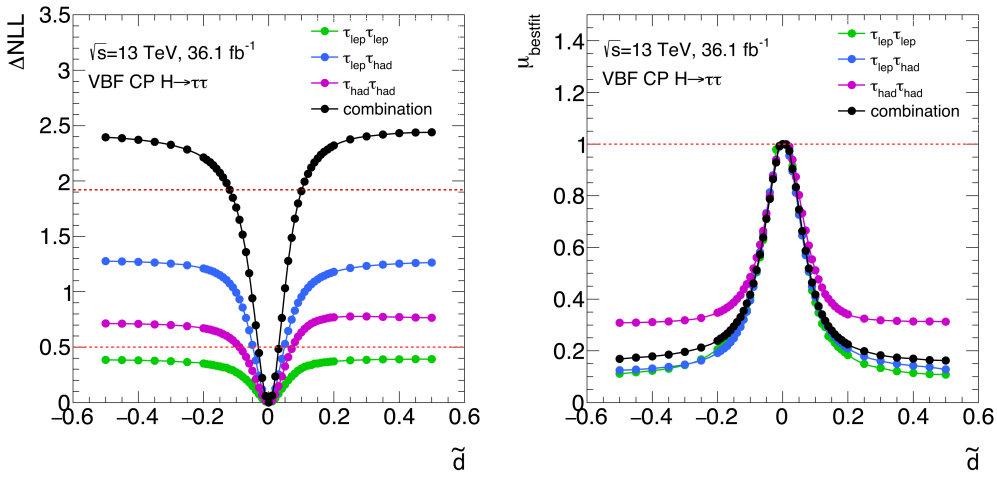
the largest impact as well as the normalization of the $Z \to ll$ background. The NPs pulls and constraints can be seen in Figure 6.21. Here, the largest constraints are obtained for the $Z \to \tau\tau$ parton shower uncertainty, and the reason is explained in the $\tau_{\text{had}}\tau_{\text{had}}$ channel results in this Section, for the jet energy resolution and tau scale uncertainties, and the motivation for these other constraints is the same as for the cross section analysis, explained in Section 6.9.1. Correlations among the NPs can be seen in Figure 6.22. Some large correlations that appear in the Figure are related to the pile-up reweighting uncer-



(a) (b)

Figure 6.18: Behaviour of the $Z \to \tau\tau$ parton shower uncertainty in the low-BDT CR. The same uncertainty is shown once with the statistical uncertainty of SHERPA in the ratio (a) and once with the statistical uncertainty of MADGRAPH5_aMC@NLO in the ratio (b).

tainty in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel and some of the bin-by-bin statistical uncertainties (these in particular arise from some bins in the low-BDT CR of the $\tau_{\text{lep}}\tau_{\text{lep}}$ SF channel).
Finally, the impact on the signal strength of the uncertainties per category is shown in Table 6.6. As already seen for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the highest contribution for the systematic uncertainties arises from jet and missing transverse momentum related uncertainties. In addition, MC statistics and signal theory uncertainties give a large contribution as well as the data statistics. In order to have an estimate of the impact of systematic uncertainties on the extracted confidence interval for $\tilde{d}$, the fit has been repeated several times removing systematic uncertainties corresponding to a certain category. The result is shown in Figure 6.23, where it is possible to notice the large effect of jet related uncertainties on the sensitivity.



(a) (b)

Figure 6.19: Expected ΔNLL curve (a) and best-fit signal-strength (b) as a function of $\tilde{d}$ for the combined $H \to \tau\tau$ fit assuming the VBF SM signal. Both the ΔNLL resulting from the standalone channel fits and the combined one are shown.

POI μ	Central Value 1
Set of nuisance parameters	Impact on uncertainty
Total	+0.42 / -0.40
Data statistics	+0.23 / -0.22
Total systematic unc.	+0.34 / -0.34
$Z \to \tau\tau$ normalization	+0.08 / -0.02
All normalizations	+0.11 / -0.07
All but normalizations	+0.34 / -0.33
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	+0.24 / -0.26
Flavor tagging	+0.02 / -0.02
Electrons and muons	+0.09 / -0.12
Tau	+0.13 / -0.10
Pileup reweighting	+0.05 / -0.04
Misidentified τ	+0.04 / -0.04
Luminosity	+0.03 / -0.02
Signal theory unc.	+0.15 / -0.11
$Z \to \tau\tau$ theory unc.	+0.11 / -0.09
MC statistics	+0.12 / -0.13

Set of nuisance parameters	Fractional impact on uncertainty (square of fraction of total)
Data statistics	+0.32 / -0.30
Total systematic unc.	+0.68 / -0.70
$Z \to \tau\tau$ normalization	+0.04 / -0.00
All normalizations	+0.07 / -0.03
All but normalizations	+0.67 / -0.69
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	+0.33 / -0.41
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.05 / -0.09
Tau	+0.10 / -0.06
Pileup reweighting	+0.01 / -0.01
Misedintifed τ	+0.01 / -0.01
Luminosity	+0.01 / -0.00
Signal theory unc.	+0.14 / -0.07
$Z \to \tau\tau$ theory unc.	+0.07 / -0.05
MC statistics	+0.08 / -0.11

Table 6.6: Expected impact of uncertainties, grouped in categories, on the signal strength for the combined $H \to \tau\tau$ fit assuming the VBF SM signal. On the left absolute values are shown while on the right the fractional contribution is given.

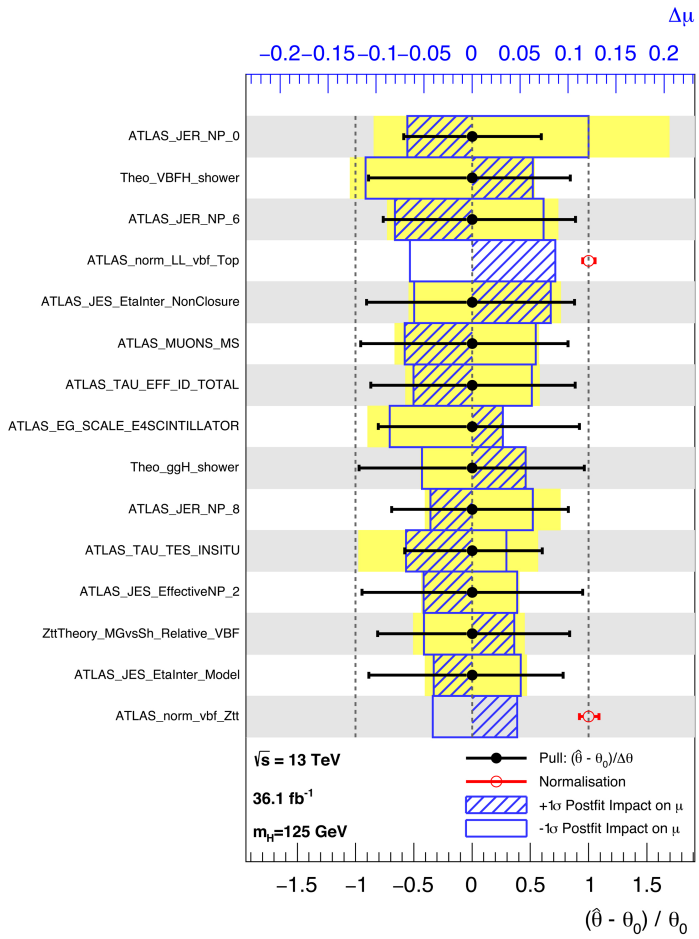


Figure 6.20: Expected impact of nuisance parameters on the signal strength for the combined $H \to \tau\tau$ fit assuming the VBF SM signal. The uncertainties are listed in decreasing order of their impact on μ and only the 15 most highly ranked uncertainties are shown. The hatched blue and the open blue boxes represent the $+1\sigma$ and the -1σ uncertainties respectively while the yellow box represent the pre-fit ones. Their fractional impact is shown referring to the top x-axis. The pulls and constraints of the uncertainties are indicated by the black points and lines, referring to the bottom x-axis. If the uncertainty is not pulled nor constrainted the pull is 0 and the lines lie between $+1$ and -1.

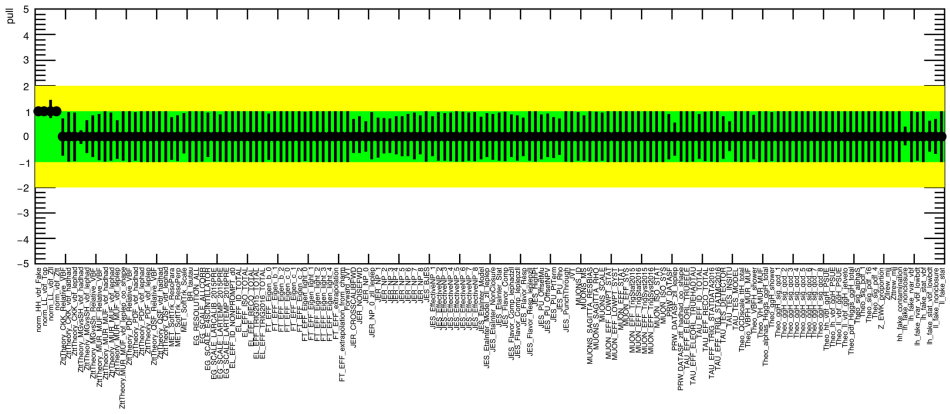


Figure 6.21: Expected nuisance parameters pulls and constraints for the combined $H \to \tau\tau$ fit assuming the VBF SM signal.

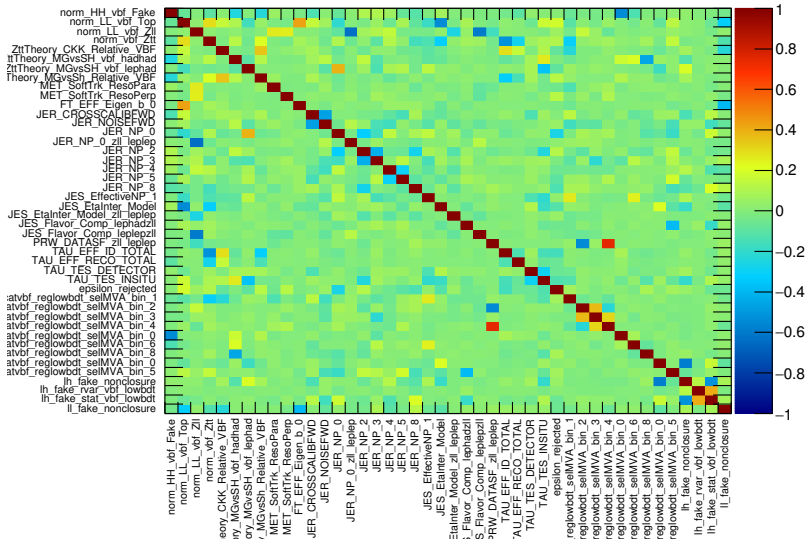


Figure 6.22: Expected nuisance parameters correlation for the combined $H \to \tau\tau$ fit assuming the VBF SM signal. Only the nuisance parameters with absolute correlation higher than 0.25 are shown.

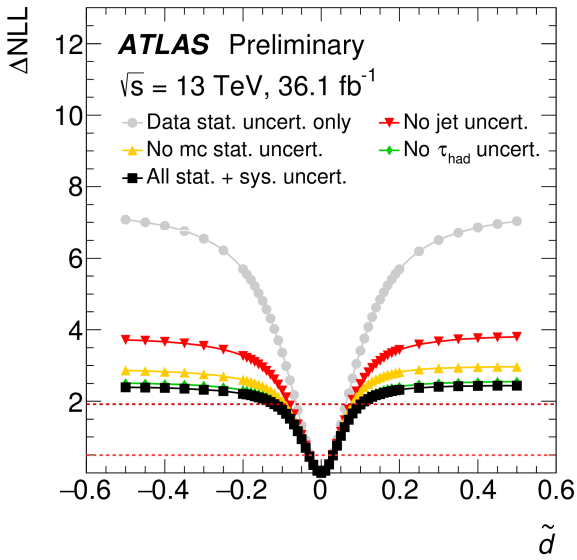


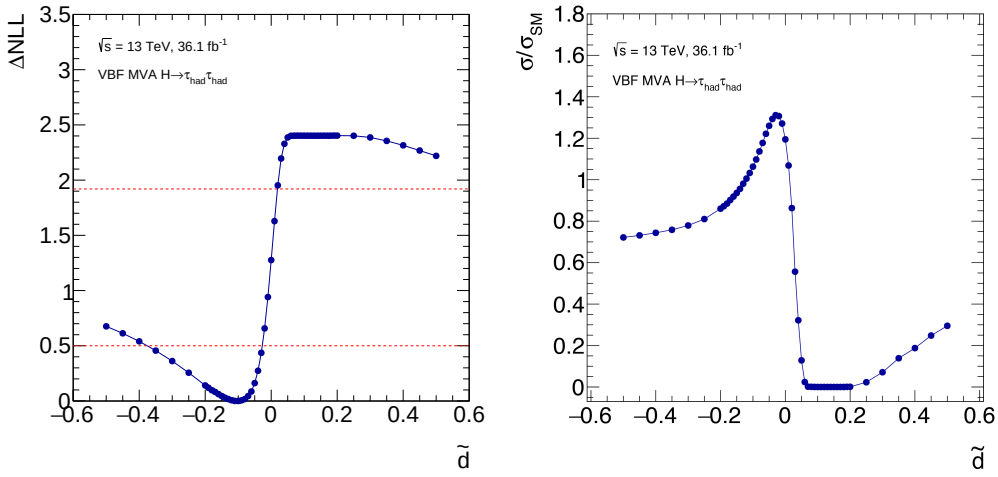
Figure 6.23: Effect of systematic uncertainties on the expected ΔNLL. The grey curve corresponds to the combined expected ΔNLL without systematics, while the other curves are obtained by excluding time by time a certain category of systematics (jet related systematic uncertainties, tau related systematic uncertainties and MC statistics) [145].

6.9.2 Observed results

The final observed results will be presented in this Section. They are obtained by using data in all the signal and control regions.

$H \to \tau_{\mathbf{had}}\tau_{\mathbf{had}}$

The observed ΔNLL and best-fit signal strength curves as a function of $\tilde{d}$ are shown in Figure 6.24. The ΔNLL presents a minimum at $\tilde{d} = -0.1$, corresponding to the signal model that describes data the best. It is also quite asymmetric, resulting in an observed confidence interval larger than the expected. Indeed the confidence interval at 68% CL for $\tilde{d}$ is $[-0.4, -0.04]$. The μ is forced to be positive, hence assuring only physical values. For a better understanding of the observed results the fit performed using the SM signal ($\tilde{d} = 0$) are discussed. Table 6.7 reports the observed best-fit signal strength and the NFs for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit. The resulting μ is $1.20^{+0.74}_{-0.77}$, therefore an excess of data with respect to the signal and background expectations is observed. Furthermore, the $Z \to \tau\tau$ NF is 0.69 which results in a $\sim 30\%$ decrease of the $Z \to \tau\tau$ event yield in the low-BDT CR. These features can be observed in Figure 6.25 as well, where the distributions of the Optimal Observable in the SR and of the $m_{\tau\tau}^{\text{MMC}}$ in the low-BDT CR after performing the fit are shown. The ΔNLL behaviour can be partly explained by looking at the asymmetric Optimal Observable distribution for data in Figure 6.25 and keeping in mind that signals with positive $\tilde{d}$ values tend to have an Optimal Observable distribution shifted to positve values while signals with negative $\tilde{d}$ values have the opposite behaviour. The event yield in the low-BDT CR is instead reduced with respect to the expected and this explains the reduction of the $Z \to \tau\tau$ NF. In Table 6.8 are reported the observed event yields for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit.
The observed impact of NPs on the signal strength is given in Figure 6.26 and shows a similar behaviour as the expected, apart from the tau energy scale uncertainty which is increased, due to adjustments of the fit. The NPs pulls and constraints, shown in Figure 6.27, do not exhibit any unexpected behaviour. The only large pull, not visible in the Figure, refers to the pile up reweighting uncertainty (and it was more pronounced in the combined fit). This pull is due to large shape variations for the $Z \to \tau\tau$ process in the SR and therefore it was decided to decorrelate this specific shape uncertainty in order to not propagate the pull to the other channels in the combined fit. There are indeed two different uncertainties that appear in the plots one called "prw_datasf" and one called "prw_datasf_ztt_hadhad_oo_shape" which refers to the Optimal Observable shape uncertainty for $Z \to \tau\tau$. The correlations among NPs are shown in Figure 6.28.
Finally, the impact of the systematic uncertainties is summarised in Table 6.9. As expected, the major contributions to the uncertainties, beside data statistics, arises from jet and missing transverse momentum related uncertainties, tau related uncertainties and MC statistics.



(a) (b)

Figure 6.24: Observed ΔNLL curve (a) and best-fit signal-strength (b) as a function of $\tilde{d}$ for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit.

Parameter	Value
μ	$1.20^{+0.74}_{-0.77}$
$Z \to \tau\tau$ NF	0.69 ± 0.13
Misedintifed τ NF	1.10 ± 0.13

Table 6.7: Observed signal strength and normalization factors values from the $\tau_{\text{had}}\tau_{\text{had}}$ standalone fit assuming the VBF SM signal.

Combined

The presence of CP violation would cause an asymmetry of the Optimal Observable. Indeed the mean value of the Optimal Observable is expected to be zero under the SM assumption, either for the signal and for the backgrounds. Observing a mean value shifted from zero, would be a sign of the presence of CP violating effects and therefore, directly measuring the Optimal Observable mean value allows to perform an almost model independent test of CP violation. The observed mean values of the Optimal Observable distribution for data are listed in Table 6.10 for each channel and for their combination. All the observed values are consistent with zero and therefore no hint for CP violation is observed.

A maximum likelihood fit is then performed as for the $\tau_{\text{had}}\tau_{\text{had}}$ channel but exploiting the informations from all the signal and control regions of the four analysis channels. The resulting combined ΔNLL and signal strength curves as a function of $\tilde{d}$ are shown in Figure 6.29 along with each channel standalone ΔNLL curves. The minimum value of the

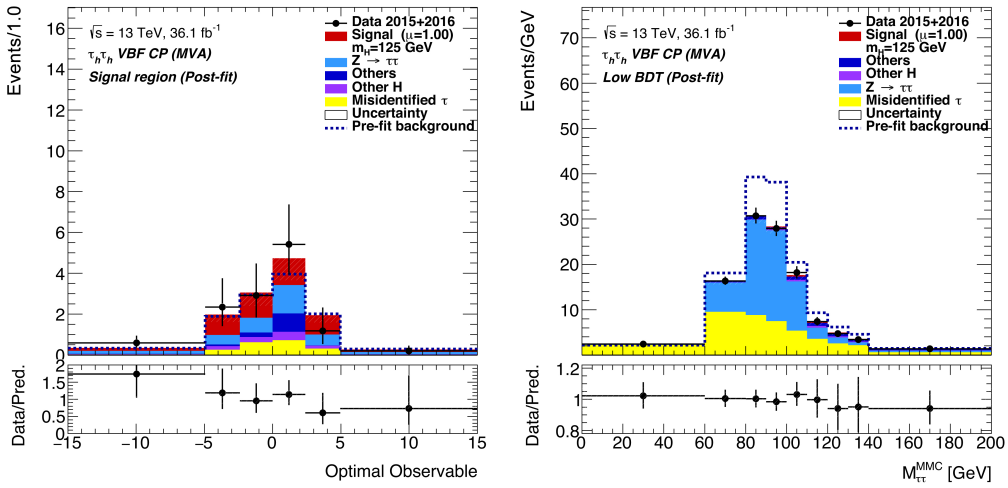
Process	$\tau_{\text{had}}\tau_{\text{had}}$	
Region	Low-BDT CR	SR
Misidentified τ	662.12 ± 62.23	5.04 ± 3.11
Other backgrounds	70.34 ± 9.95	3.13 ± 1.34
$Z \to \tau\tau$	720.51 ± 64.53	10.13 ± 3.65
Other Higgs prod. (ggF, VH, $t\bar{t}H$) $H \to \tau\tau$	15.24 ± 6.72	2.88 ± 1.42
Sum of backgrounds	1468.21 ± 38.73	21.18 ± 4.98
VBF $H \to \tau\tau$ ($\tilde{d} = 0$)	14.07 ± 2.38	13.72 ± 2.08
Data	1479	37

Table 6.8: Event yields for signal and background after performing the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit assuming the VBF SM signal.

POI	Central Value
μ	1
Set of nuisance parameters	Impact on uncertainty
Total	+0.75 / -0.75
Data statistics	+0.44 / -0.40
Total systematic unc.	+0.60 / -0.64
$Z \to \tau\tau$ normalization	+0.22 / -0.06
All normalizations	+0.24 / -0.09
All but normalizations	+0.60 / -0.64
Jets and $E_{\text{T}}^{\text{miss}}$	+0.40 / -0.52
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.00 / -0.00
Tau	+0.38 / -0.36
Pileup reweighting	+0.15 / -0.20
Misidentified τ	+0.003 / -0.01
Luminosity	+0.04 / -0.02
Signal theory u1c.	+0.10 / -0.08
$Z \to \tau\tau$ theory unc.	+0.06 / -0.10
MC statistics	+0.20 / -0.28

Set of nuisance parameters	Fractional impact on uncertainty (square of fraction of total)
Data statistics	+0.34 / -0.28
Total systematic unc.	+0.65 / -0.72
$Z \to \tau\tau$ normalization	+0.08 / -0.005
All normalizations	+0.10 / -0.01
All but normalizations	+0.64 / -0.71
Jets and $E_{\text{T}}^{\text{miss}}$	+0.28 / -0.47
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.00 / -0.00
Tau	+0.25 / -0.23
Pileup reweighting	+0.04 / -0.07
Misedintifed τ	+0.00 / -0.00
Luminosity	+0.00 / -0.00
Signal theory unc.	+0.02 / -0.01
$Z \to \tau\tau$ theory unc.	+0.01 / -0.02
MC statistics	+0.07 / -0.14

Table 6.9: Observed impact of uncertainties, grouped in categories, on the signal strength for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit assuming the VBF SM signal. On the left absolute values are shown while on the right the fractional contribution is given.



(a) (b)

Figure 6.25: Distributions for the Optimal Observable in the SR and the $m_{\tau\tau}^{\mathrm{MMC}}$ in the CR after performing the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel standalone fit.

ΔNLL is observed at $\tilde{d} = -0.01$ and the best-fit signal strength extracted from the fit is $\mu = 0.73 \pm 0.46$. The resulting confidence interval is $\tilde{d} \in [-0.090, 0.035]$ at 68% CL which is larger than the expected one and despite the expectation, no confidence interval at 95% CL can be determined. The observed results can be understood by looking at Figure 6.30 where different ΔNLL curves are shown corresponding to different setups for the fits: beside the combined and expected Asimov fits, other two fits have been performed, once constructing an Asimov dataset with all the NPs and NFs coming from a fit to data using only CRs and with $\mu = 1$ and once constructing the same Asimov dataset but using the best-fit signal strength, $\mu = 0.73$. Therefore, the effect of the NPs and NFs can be seen by looking at the difference between the pre-fit expected curve and the curve obtained for $\tilde{d} = 0$ and $\mu = 1$ when using all the NPs and NFs constrained by a CR-only fit to data. This indicates that the preferred NPs and NFs values, based on the observed data in the CRs, lead to a decrease in sensitivity. A further drop in sensitivity can be noticed by looking at the ΔNLL curve which has the signal strength μ set to the observed value, $\mu = 0.73$, beside all the NPs and NFs set to their best-fit values from the CR-only fit to data. This demonstrates that another main reason for the reduced sensitivity is the observed low signal strength value.
Table 6.11 quotes all the observed best-fit values for the signal strength and for the NFs, both for the combined and the standalone channel fits.
Figure 6.32 and 6.31 show the distributions of the Optimal Observable in the SRs and of the $m_{\tau\tau}^{\mathrm{MMC}}$ in the CRs respectively. These distributions are obtained after performing the fit and using all the parameters as determined by the combined fit. In addition, the signal corresponds to the best-fit $\tilde{d}$ value observed, i.e. $\tilde{d} = -0.01$ and also $\mu = 0.73$. Very good modelling of the $m_{\tau\tau}^{\mathrm{MMC}}$ distribution in the CRs can be observed and this is expected as

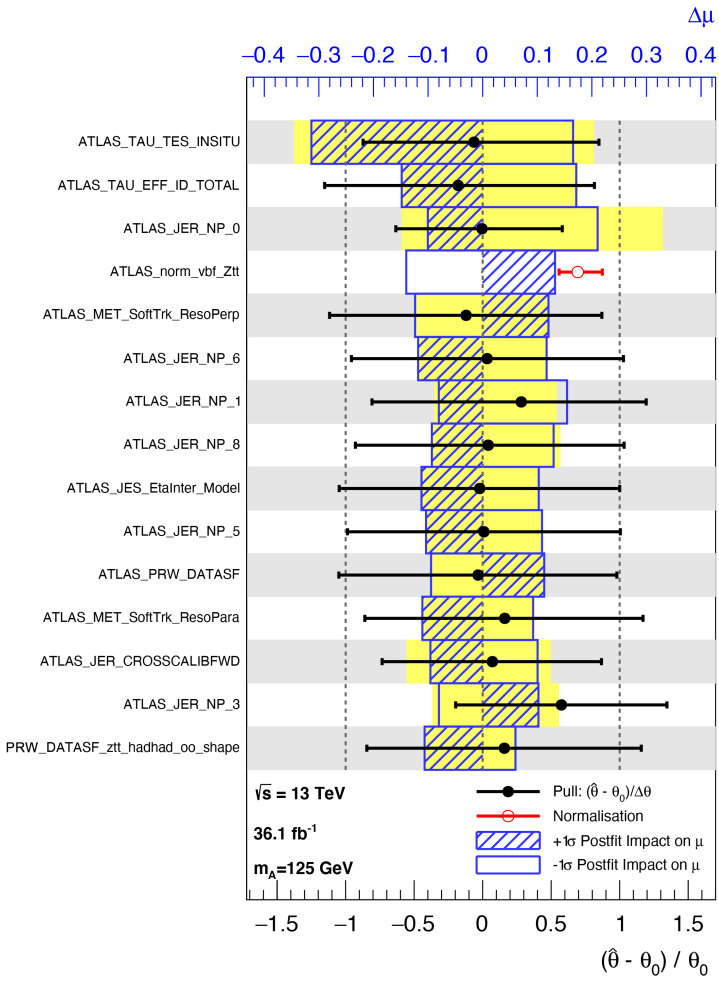


Figure 6.26: Observed impact of nuisance parameters on the signal strength for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit assuming the VBF SM signal. The uncertainties are listed in decreasing order of their impact on μ and only the 15 highly ranked uncertainties are shown. The hatched blue and the open blue boxes represent the $+1\sigma$ and the -1σ uncertainties respectively while the yellow box represent the pre-fit ones. Their fractional impact is shown referring to the top x-axis. The pulls and constraints of the uncertainties are indicated by the black points and lines, referring to the bottom x-axis. If the uncertainty is not pulled nor constrainted the pull is 0 and the lines lie between $+1$ and -1.

this is the variable used in the fit. A good modelling within uncertainties can also be seen for the Optimal Observable distributions in the SRs. The same statement applies to the results reported in Table 6.12 and 6.13, which show the event yields in the SRs for data, for all the background processes and either for the VBF SM signal ($\mu = 1$, $\tilde{d} = 0$) and for

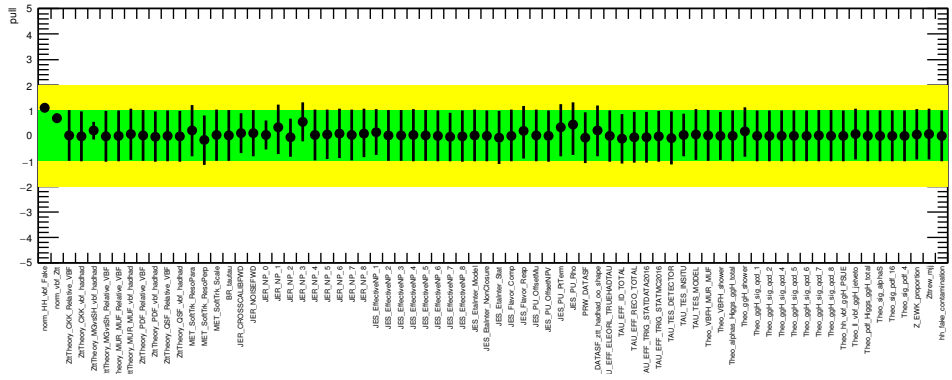


Figure 6.27: Observed nuisance parameters pulls and constraints for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit assuming the VBF SM signal.

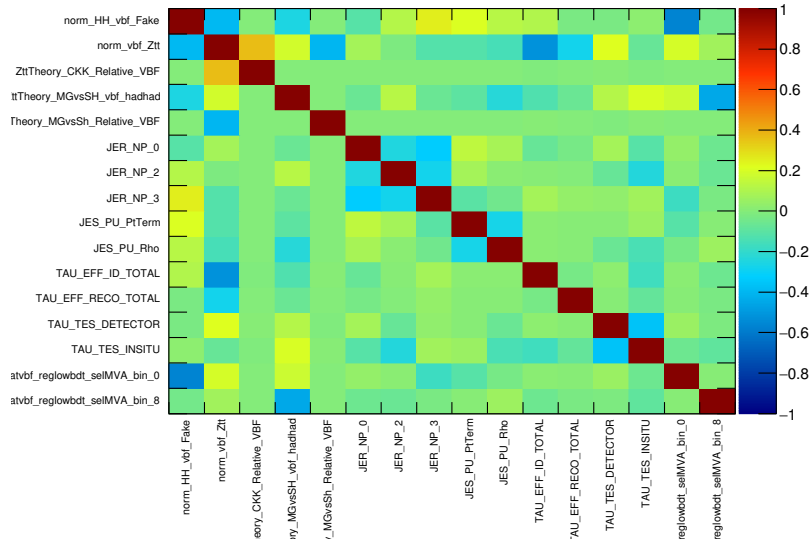


Figure 6.28: Observed nuisance parameters correlation for the $\tau_{\text{had}}\tau_{\text{had}}$ channel standalone fit. Only the nuisance parameters with absolute correlation higher than 0.25 are shown assuming the VBF SM signal.

the VBF signal considering the best-fit $\tilde{d}$ and μ values ($\mu = 0.73$, $\tilde{d} = -0.01$).
Both the combined signal strength and the combined ΔNLL are in some cases lower than the standalone channel curves. This can happen since the signal is correlated across the channels and also are most of the NPs. Thus, the combined ΔNLL has a non-trivial behaviour and it is not simply the sum of the individual ΔNLL. In order to better visualize the contribution of each channel to the combined result, another fit has been performed exploiting the regions of the channel used and only event yield information from the other channels SRs. This ensures that the minimum value for the standalone channel ΔNLL

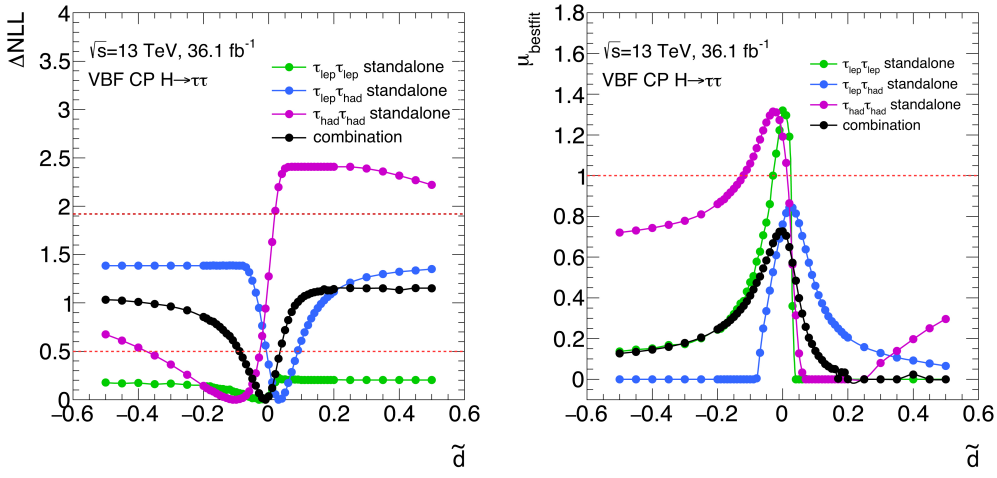
Channel	Optimal Observable
$\tau_{\text{lep}}\tau_{\text{lep}}$ SF	-0.54 ± 0.72
$\tau_{\text{lep}}\tau_{\text{lep}}$ DF	0.71 ± 0.81
$\tau_{\text{lep}}\tau_{\text{had}}$	0.74 ± 0.77
$\tau_{\text{had}}\tau_{\text{had}}$	-1.13 ± 0.64
Combined	-0.19 ± 0.37

Table 6.10: Optimal Observable mean values observed in data for each analysis channels and for their combination. The values are reported along with their statistical uncertainties.

is not biased from the other channels Optimal Observable distributions. This is shown in Figure 6.33. The $\tau_{\text{had}}\tau_{\text{had}}$ sensitivity is therefore reduced while the $\tau_{\text{lep}}\tau_{\text{had}}$ one is enhanced. In addition, the $\tau_{\text{lep}}\tau_{\text{lep}}$ SF and DF contribution are shown separately and it can be noticed the sizeable contribution to the sensitivity given by the $\tau_{\text{lep}}\tau_{\text{lep}}$ SF channel despite the low signal to background ratio of this region (keeping in mind that the analysis aims at the $\tilde{d}$ measurement and not at the cross section or signal strength one).
Finally, the behaviour of the NPs has been analysed for the fit which assumes the SM hypothesis. Figure 6.34 shows the systematic uncertainties impact on the signal strength. The parameters with the highest impact are related to jet energy resolution, tau energy scale uncertainty as well as parton shower uncertainty for the VBF signal. They are in agreement with the expectations apart from the tau energy scale uncertainty on the tag and probe measurement, which has an increased impact driven by the $\tau_{\text{had}}\tau_{\text{had}}$ channel. Figure 6.35 shows the NP pulls and constraints and Figure 6.36 shows the correlation among them. They do not exhibit any unexpected behaviour apart from some pulls in the $\tau_{\text{lep}}\tau_{\text{lep}}$ channel which originate from some specific uncertainty for the $Z \to ll$ process and have been decorrelated as done for the pile-up reweighting uncertainty in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The observed impact on the signal strength of systematic uncertainties is given in Table 6.14. The large impact of systematic uncertainties can be noticed, in particular of the jet and missing transverse momentum related uncertainties, the τ based uncertainties and the MC statistics.

Parameter	Combined	$\tau_{\text{lep}}\tau_{\text{lep}}$	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
μ	0.73 ± 0.47	$1.32^{+1.42}_{-3.07}$	$0.76^{+0.57}_{-0.51}$	$1.20^{+0.74}_{-0.77}$
NF($Z \to \tau\tau$)	0.93 ± 0.08	0.93 ± 0.14	0.90 ± 0.09	0.69 ± 0.13
NF($Z \to ll$)	0.99 ± 0.36	0.99 ± 0.36	$--$	$--$
NF(Top)	1.16 ± 0.06	1.14 ± 0.07	$--$	$--$
NF(Misidentified τ)	0.99 ± 0.09	$--$	$--$	1.10 ± 0.13

Table 6.11: Observed signal stregnth and normalization factors for the backgrounds values exctracted from the combined $H \to \tau\tau$ fit assuming the VBF SM signal. The channel standalone fit values are reported as well.



(a) (b)

Figure 6.29: Observed ΔNLL curve (a) and best-fit signal-strength (b) as a function of $\tilde{d}$ for the combined $H \to \tau\tau$ fit. Both the ΔNLL resulting from the standalone channel fits and the combined one are shown.

Process	$\tau_{\text{lep}}\tau_{\text{lep}}$ SF	$\tau_{\text{lep}}\tau_{\text{lep}}$ DF
Data	26	30
VBF $H \to \tau\tau/WW$ ($\mu = 0.73$, $\tilde{d} = -0.01$)	3.3 ± 2.1	5.1 ± 3.1
VBF $H \to \tau\tau/WW$ ($\mu = 1$, $\tilde{d} = 0$)	4.5 ± 2.9	6.9 ± 4.4
$Z \to \tau\tau$	6.6 ± 3.7	8.2 ± 3.8
Misidentified τ	0.02 ± 0.20	2.3 ± 0.7
Top-quark	3.8 ± 2.3	10.6 ± 5.5
$Z \to ll$	11 ± 18	1.8 ± 1.1
Diboson	0.70 ± 0.59	
$ggF/\ VH/\ t\bar{t}H,\ H \to \tau\tau/WW$	0.49 ± 0.48	0.70 ± 0.30
Sum of backgrounds	23 ± 17	23.6 ± 6.1

Table 6.12: Event yields in the $\tau_{\text{lep}}\tau_{\text{lep}}$ SF and $\tau_{\text{lep}}\tau_{\text{lep}}$ DF SRs, after performing the $H \to \tau\tau$ combined fit. The expected event yields for the VBF signal with the corresponding best-fit signal strength and $\tilde{d}$ values ($\mu = 0.73$, $\tilde{d} = -0.01$) is also shown.

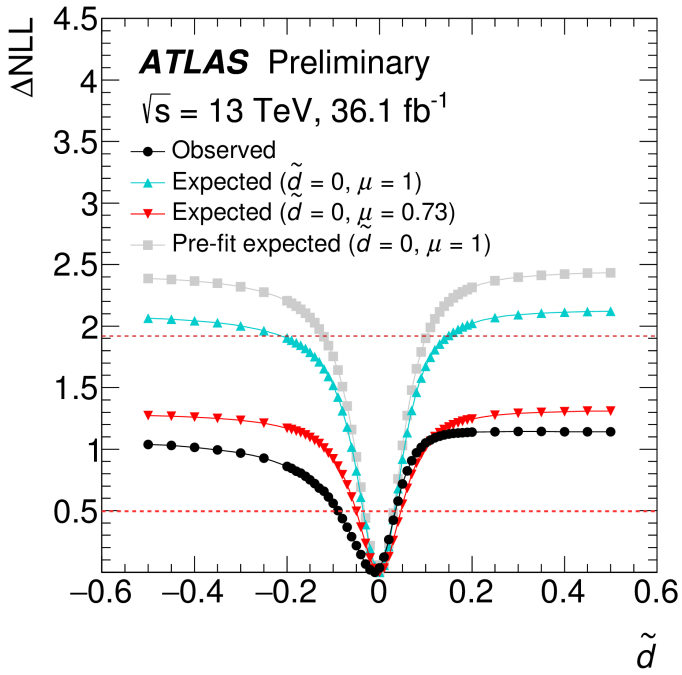
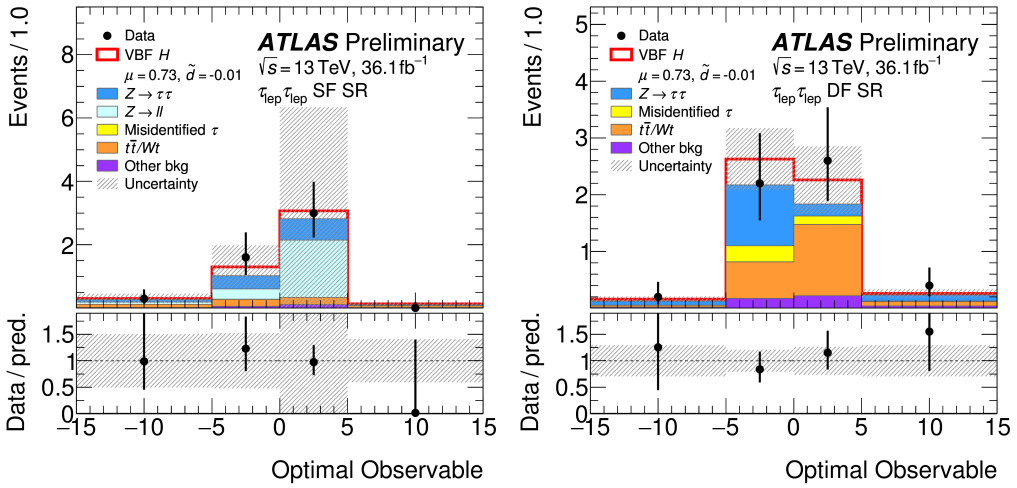


Figure 6.30: ΔNLL curves as a function of $\tilde{d}$ for the combined $H \to \tau\tau$ fit. The combined observed and the expected ΔNLL are shown in the same plot together with other two ΔNLL curves. These are obtained by fitting an Asimov dataset wich is obtained using NPs and NFs obtained in a CR-only fit to data, once using $\mu = 1$ and once using the best-fit signal strength $\mu = 0.73$ [145].

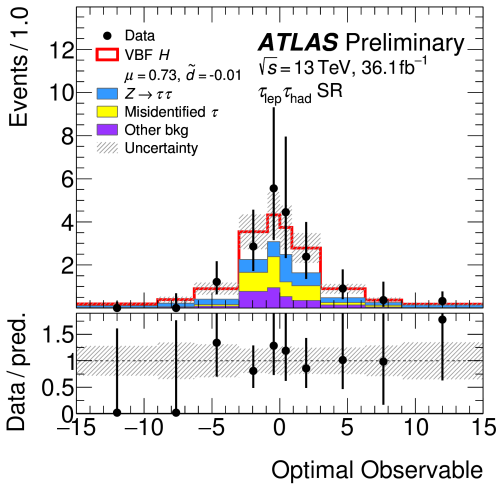
Process	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
Data	30	37
VBF $H \to \tau\tau/WW$ ($\mu = 0.73$, $\tilde{d} = -0.01$)	11.8 ± 7.4	8.9 ± 5.6
VBF $H \to \tau\tau/WW$ ($\mu = 1$, $\tilde{d} = 0$)	16 ± 10	12.3 ± 7.7
$Z \to \tau\tau$	7.8 ± 3.5	15.5 ± 5.2
Misidentified τ	6.2 ± 1.0	5.4 ± 2.7
ggF/ VH/ $t\bar{t}H$, $H \to \tau\tau/WW$	2.1 ± 1.5	2.8 ± 1.4
Other backgrounds	2.8 ± 3.1	2.3 ± 0.8
Sum of backgrounds	19.0 ± 5.5	26.0 ± 6.6

Table 6.13: Event yields in the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ SRs, after performing the $H \to \tau\tau$ combined fit. The expected event yields for the VBF signal with the corresponding best-fit signal strength and $\tilde{d}$ values ($\mu = 0.73$, $\tilde{d} = -0.01$) is also shown.

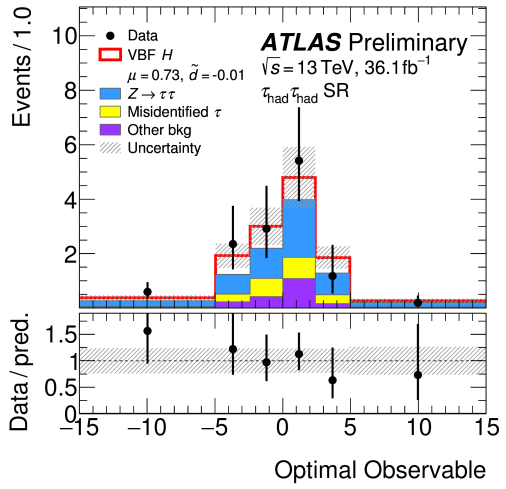


(a)

(b)



(c)



(d)

Figure 6.31: Optimal Observable distributions in the SR, after performing the combined $H \to \tau\tau$ fit for each of the four analysis channel. The values $\tilde{d}$, signal strength and all the nuisance parameters and normalization factors are set to their best-fit values. Both statistical and systematic uncertainties are considered in the hatched band in the ratio.

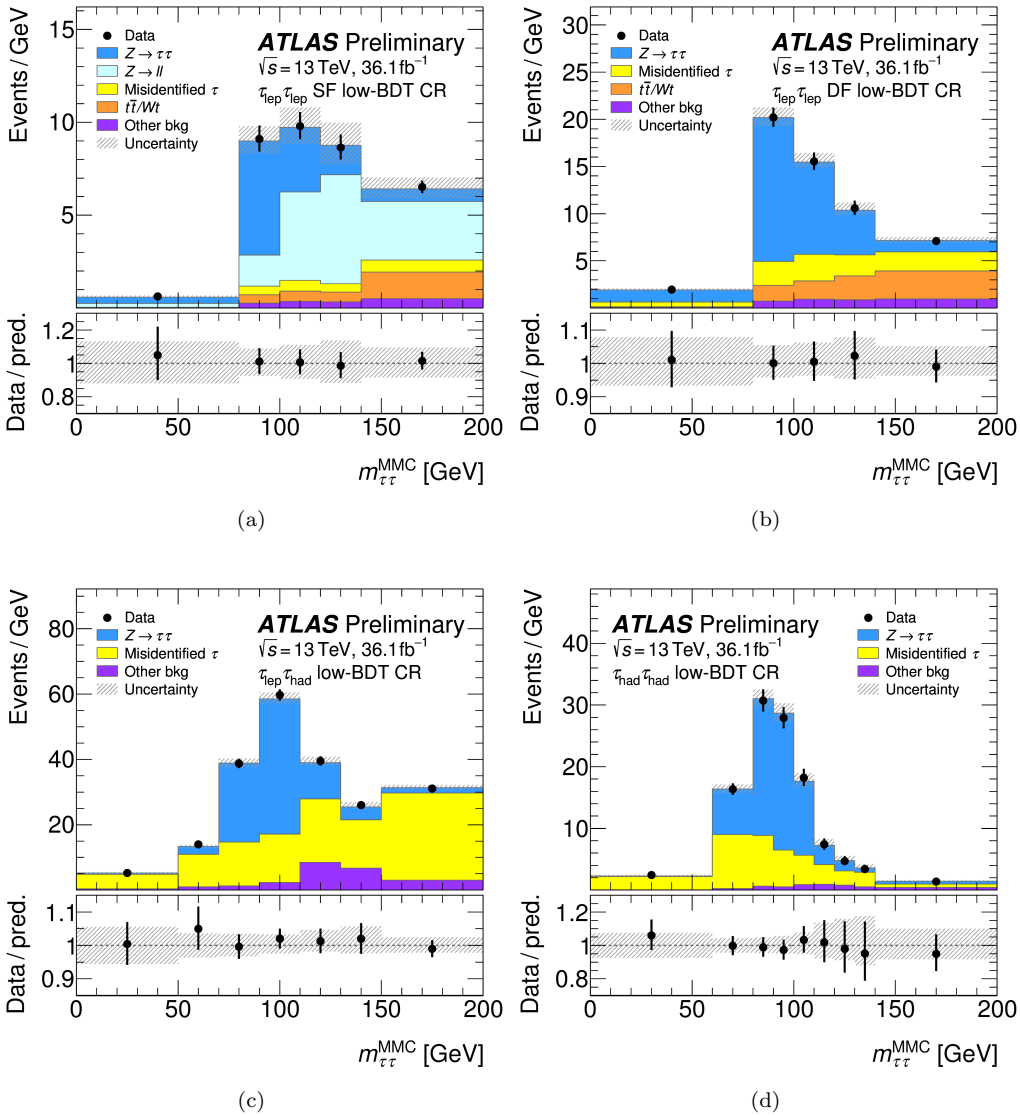


Figure 6.32: $m_{\tau\tau}^{\mathrm{MMC}}$ distributions in the low-BDT CR, after performing the combined $H \to \tau\tau$ fit for each of the four analysis channel. The signal contamination is negligible. Both statistical and systematic uncertainties are considered in the hatched band in the ratio.

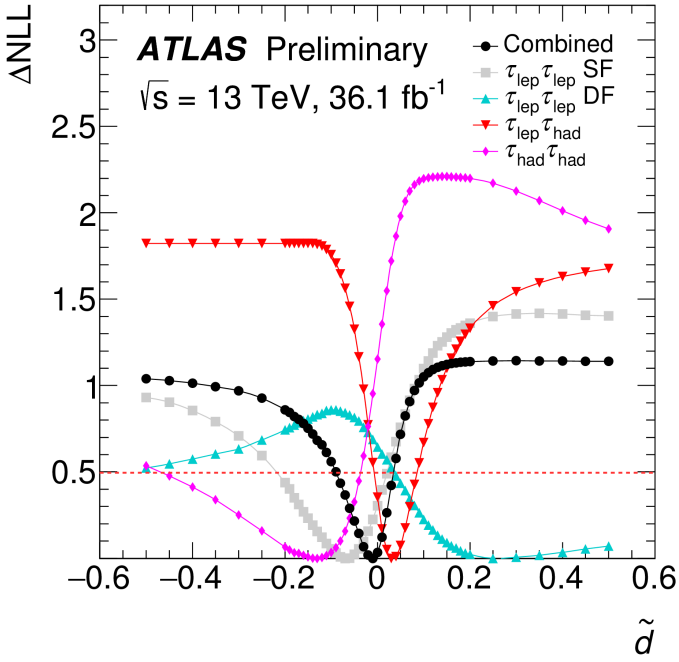


Figure 6.33: Observed ΔNLL curves for each analysis channel and for the combined result as a function of $\tilde{d}$. The individual channel ΔNLL curves are obtained by using all the regions of the featured channel plus the event yields from the SRs of the other channels.

POI μ	Central Value 0.73
Set of nuisance parameters	Impact on uncertainty
Total	+0.49 / -0.48
Data statistics	+0.25 / -0.23
Total systematic unc.	+0.43 / -0.42
$Z \to \tau\tau$ normalization	+0.06 / -0.01
All normalizations	+0.09 / -0.05
All but normalizations	+0.43 / -0.42
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	+0.29 / -0.32
Flavor tagging	+0.03 / -0.03
Electrons and muons	+0.14 / -0.20
Tau	+0.18 / -0.15
Pileup reweighting	+0.06 / -0.08
Misidentified τ	+0.05 / -0.06
Luminosity	+0.03 / -0.02
Signal theory unc.	+0.20 / -0.12
$Z \to \tau\tau$ theory unc.	+0.11 / -0.08
MC statistics	+0.19 / -0.23

Set of nuisance parameters	Fractional impact on uncertainty (square of fraction of total)
Data statistics	+0.25 / -0.24
Total systematic unc.	+0.75 / -0.76
$Z \to \tau\tau$ normalization	+0.02 / -0.00
All normalizations	+0.03 / -0.01
All but normalizations	+0.73 / -0.75
Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	+0.35 / -0.43
Flavor tagging	+0.00 / -0.00
Electrons and muons	+0.08 / -0.17
Tau	+0.13 /-0.10
Pileup reweighting	+0.02 / -0.03
Misedintifed τ	+0.01 / -0.01
Luminosity	+0.00 / -0.00
Signal theory unc.	+0.20 / -0.12
$Z \to \tau\tau$ theory unc.	+0.16 / -0.06
MC statistics	+0.15 / -0.23

Table 6.14: Observed impact of uncertainties, grouped in categories, on the signal strength for the combined $H \to \tau\tau$ fit. On the left absolute values are shown while on the right the fractional contribution is given.

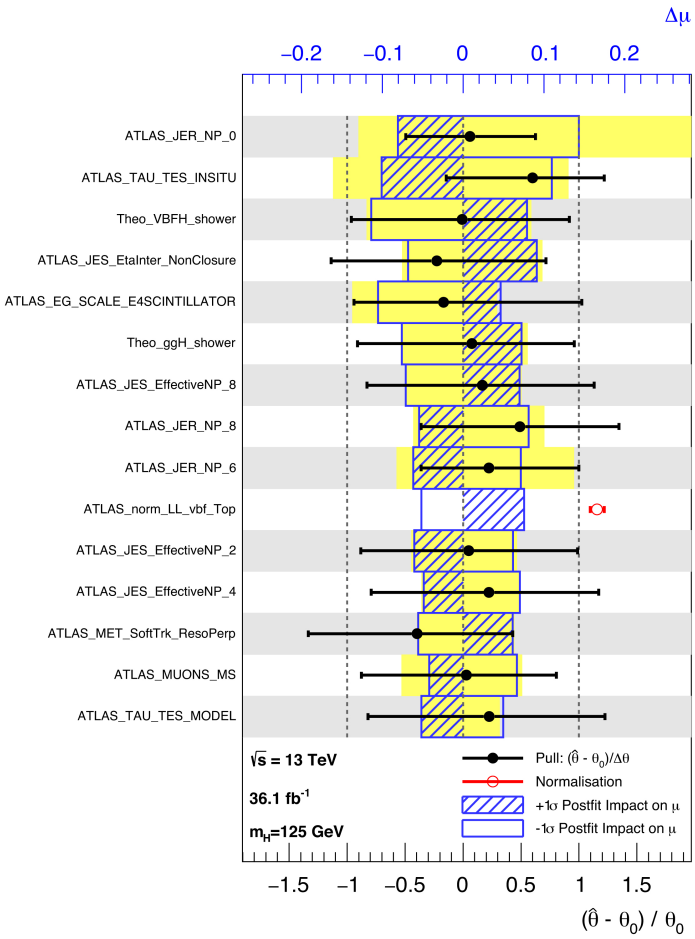


Figure 6.34: Observed impact of nuisance parameters on the signal strength for the combined $H \to \tau\tau$ fit assuming the VBF SM signal. The uncertainties are listed in decreasing order of their impact on μ and only the 15 highly ranked uncertainties are shown. The hatched blue and the open blue boxes represent the $+1\sigma$ and the -1σ uncertainties respectively while the yellow box represent the pre-fit ones. Their fractional impact is shown referring to the top x-axis. The pulls and constraints of the uncertainties are indicated by the black points and lines, referring to the bottom x-axis. If the uncertainty is not pulled nor constrainted the pull is 0 and the lines lie between $+1$ and -1.

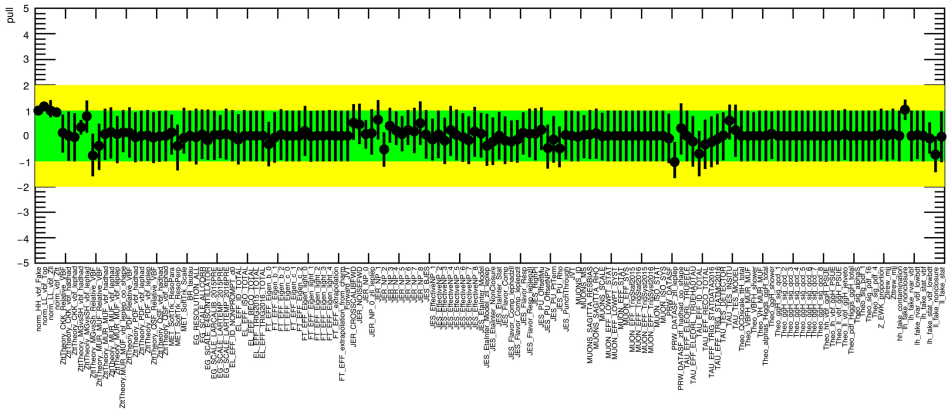


Figure 6.35: Observed nuisance parameters pulls and constraints for the combined $H \to \tau\tau$ fit assuming the VBF SM signal.

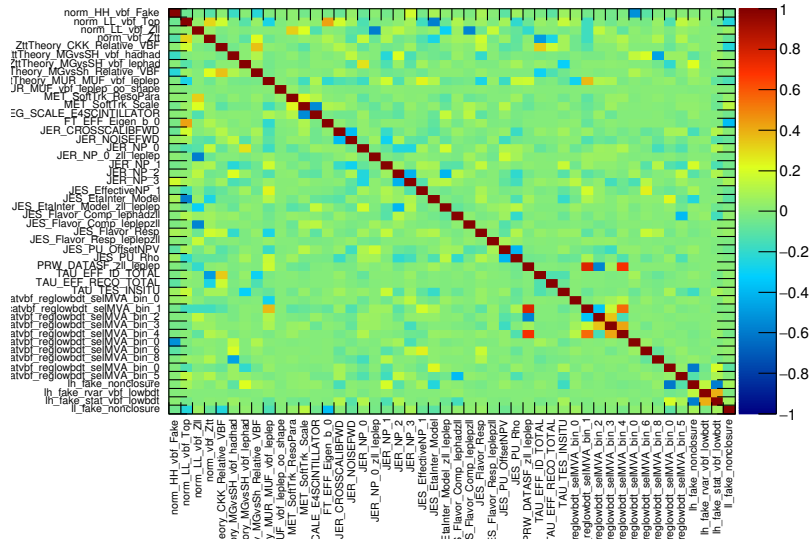


Figure 6.36: Observed nuisance parameters correlation for the combined $H \to \tau\tau$ fit. Only the nuisance parameters with absolute correlation higher than 0.25 are shown.

6.10 Comparison with other analysis

As described in Section 6.1, other analysis in both CMS and ATLAS experiments have studied the structure of the HVV vertex, setting limits on a parameter that quantify CP violation. However they all used different parametrizations.
The ATLAS analysis which study the HVV vertex in the Higgs decay channel ($H \to VV$) use a parametrization which is based on the Higgs characterisation model [165]. The corresponding Lagrangian, considering only the SM and a CP-odd contribution, is the following:

$$\begin{aligned}\mathcal{L}_{eff} = H\Bigg\{ & c_\alpha \kappa_{\rm SM} \left[\frac{1}{2}\frac{2m_V^2}{v} Z_\mu Z^\mu + g_{HWW} W_\mu W^\mu\right] \\ & - \frac{1}{4}\frac{1}{\Lambda} s_\alpha \kappa_{AZZ} Z_{\mu\nu} \tilde{Z}^{\mu\nu} - \frac{1}{2}\frac{1}{\Lambda} s_\alpha \kappa_{AWW} W^+_{\mu\nu} \tilde{W}^{-\mu\nu} \Bigg\}\end{aligned} \tag{6.21}$$

where Λ is the energy scale at which new physics appear, c_α and s_α are the cosine and sine of the α angle respectively, which represent the mixing between the 0^+ and the 0^- states, and v is the vacuum expectation value. The parameter that the $H \to VV$ Run 1 ATLAS analysis uses for the fit is equivalent to $\tilde{d}$ (for the assumption used in the presented analysis, $\tilde{d} = \tilde{d}_B$), i.e.,

$$\tilde{d} = -\tilde{k}_{AVV}/\kappa_{SM} \tan\alpha \tag{6.22}$$

where $\tilde{k}_{AVV}$ has the following relation with the strength of the coupling appearing in the Higgs characterisation lagrangian, k_{AVV},

$$\tilde{k}_{AVV} = \frac{1}{4}\frac{v}{\Lambda} k_{AVV} \tag{6.23}$$

The analysis described in this work and the $H \to VV$ analysis performed in ATLAS during Run 1 data taking can be strictly compared. More recent Run 2 measurements performed in ATLAS, as already mentioned, exploit rate information as well and therefore a comparison can be only carried out with some caveats. In particular, a measurement in the $H \to ZZ$ [139] has been exploited using the same parametrization aforementioned. However, the fitted parameter is $k_{AVV} \tan\alpha$ where α is arbitrarily set to $\pi/4$. In addition, the energy scale Λ is set to 1 TeV and the coupling $k_{\rm SM}$ is set to the SM value of 1. Another fit is also performed with the same setup but with the $k_{\rm SM}$ coupling left free to float.
The CMS collaboration uses the following ansatz for the amplitude $A(H \to VV)$, when only considering a SM like contribution and an anomalous CP-odd contribution:

$$A(H \to VV) = \frac{1}{v} \epsilon_1^{*\mu} \epsilon_2^{*\nu} (g_1 M_V^2 g_{\mu\nu} - 2 g_4 \epsilon_{\mu\nu\alpha\beta} p_1^\alpha p_2^\beta)\,. \tag{6.24}$$

Comparing this with the vertex structure $T^{\mu\nu}(p_1, p_2)$, shown in Section 6.2, we can notice that $\tilde{d}$ is equal to the ratio of $|g_1|$ over $|g_4|$:

$$\tilde{d} = \frac{|g_1|}{|g_4|} \tag{6.25}$$

	Decay	Dataset	exp. CI 68% CL	obs. CI 68% CL	exp. CI 95% CL	obs. CI 95% CL
CMS [140]	$H \to VV$	Run 1	[-1.45, 1.45]	[-0.78, 0.78]	[-2.81, 2.81]	[-1.61, 1.65]
CMS [142]	$H \to VV$	Run 1 + Run 2 (80 fb^{-1})	[-0.11, 0.11]	[-0.10, 0.05]	[-0.76, 0.76]	[-1.12, 0.80]
CMS [143]	$H \to \tau\tau$	Run 2 (80 fb^{-1})	[-0.043, 0.043]	[-0.053, 0.078]	[-0.154, 0.154]	N.A.
CMS [143]		Combination	[-0.039, 0.039]	[-0.042, 0.042]	[-0.086, 0.086]	[-0.81, 0.30]
ATLAS [136]	$H \to VV$	Run 1	N.A.	N.A.	[-2.33, 2.30]	[-2.18, 0.83]
ATLAS [139]	$H \to ZZ$	Run 2 (36.1 fb^{-1}) ($k_{\mathrm{SM}} = 1$)	N.A.	N.A.	[-0.22, 0.22]	[-0.31, 0.31]
ATLAS [139]	$H \to ZZ$	Run 2 (36.1 fb^{-1}) ($\hat{k}_{\mathrm{SM}} = 1.2$)	N.A.	N.A.	[-0.18, 0.18]	[-0.26, 0.26]
ATLAS [144]	$H \to \tau\tau$	Run1	[-0.18, 0.18]	N.A.	N.A.	N.A.
ATLAS [145]	$H \to \tau\tau$	Run2 (36 fb^{-1})	[-0.035, 0.033]	[-0.090, 0.035]	[-0.21, 0.15]	N.A.

Table 6.15: Comparison of observed and expected confidence intervals for $\tilde{d}$ at 68% and 95% confidence level (CL).

Than the CMS analysis use the following cross section ratio:

$$f_3 = f_{g4} = \frac{|g_4|^2\sigma_4}{|g_1|^2\sigma_1 + |g_4|^2\sigma_4} = \frac{r_{41}^2}{1 + r_{41}^2} \tag{6.26}$$

with

$$r_{41} = \frac{|g_4|}{|g_1|}\sqrt{\frac{\sigma_4}{\sigma_1}} \tag{6.27}$$

The σ_1 cross section correspond to the SM cross section obtained for $g_1 = 1$ and $g_4 = 0$ and σ_4 is the cross section for a CP-odd coupling, obtained for $g_1 = 0$ and $g_4 = 1$. The numerical values for the ratio $|\sigma_4|/|\sigma_1|$ are 0.143 (0.153) at a center of mass energy of 8 (13) TeV respectively. The $f_3\cos(\phi_3)$ with $\phi_3 = 0$ and $\phi_3 = \pi$ is then the parameter extracted from the fit.
A summary of the results of all the analysis mentioned is given in Table 6.15 with the caveat that the $H \to ZZ$ Run 2 ATLAS analysis cannot be rigorously compared. The analysis described in this work is highly competitive. Indeed, it provides the best estimation of the 68% CL interval and also the confidence interval determined at 95% CL is larger only than the CMS combined result between the VBF production measurement in $H \to \tau\tau$ and the measurement in the $H \to VV$ decay.

6.11 Conclusion

This Chapter describes a CP violation measurement performed in the $H \to \tau\tau$ decay channel, using data recorded by ATLAS at $\sqrt{s} = 13$ TeV and corresponding to 36.1 fb^{-1}. This measurement has exploited the technique of Optimal Observable which allows to perform a direct test of CP invariance and has shown high potentiality for constraining the $\tilde{d}$ parameter, which quantifies the amount of CP violation.
A 68% CL interval for $\tilde{d}$ is then found to be [-0.090, 0.035] which has to be compared to the expected interval of $\tilde{d} \in [-0.035, 0.033]$. No confidence interval can be determined at 95% CL, despite an expected interval of $\tilde{d} \in [-0.21, 0.15]$. The $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel provide a significant contribution, also being partly responsible of Run 1 measurement [144] improvement, which did not exploit this decay channel. Indeed, the same measurement

performed during Run 1 data found an observed (expected) confidence interval for $\tilde{d}$ of [-0.11, 0.05] ([-0.18, 0.18]) at 68% CL.

Appendices

Appendix A

Pixel detector studies

A.1 Introduction

The work related to the ATLAS Pixel detector [48] [49], mentioned in the introduction, will be summarised in this Appendix.
As described in Chapter 2, the ATLAS pixel detector is the innermost part of the ATLAS detector. It consists of four cylindrical layers in the barrel region and three disks for each end-cap. The first layer, called Insertable B-Layer (IBL), was installed during the first long shutdown, between a new smaller beam pipe and the orignal first layer of the pixel detector.
The pixel detector is the first stage of the tracking system and it has many purposes such as the identification of the primary vertex, the reconstruction of the secondary vertex and the track impact parameter measurement. The pixel detector allows to perform these tasks with high accuracy, thanks to its high granularity and good spatial resolution. The IBL, in particular, has helped to reach better track and vertex reconstruction performance, considering the higher Run 2 luminosities, and to reduce the impact on tracking of the radiation damage on the original innermost layer.
The work done for the pixel detector, has mainly concerned studies on the errors flagged by the front-end chips and the read-out in order to look at possible correlations with occupancy and beam conditions. In addition, I took care of the maintenance and modifications of two related software packages which are able to convert and monitor data and informations coming from the pixel modules. Therefore, both the errors studies and the more technical work on the software, are strictly related to the data quality monitoring, which is essential to produce the final good run lists (GRL) used for physics analysis.

A.2 The ATLAS Pixel detector

The ATLAS pixel detector has been designed in order to achieve the high performances mentioned in Section A.1. These requirements concern good resolution for the transverse impact parameter ($\sim 15\ \mu$m) and in the longitudinal z-coordinate, allowing pimary vertex reconstruction of charged tracks with $\sigma(z) < 1$ mm, the maximum possible solid angle coverage ($|\eta| < 2.5$), very good b-tagging efficiency, minimal amount of material for all the elements in the system in order to reduce multiple scattering and secondary interactions and finally radiation hardness.

Layer	Mean Radius [mm]	Staves	Modules	Pixels ($\times 10^6$)
IBL	25.7	14	224	6.02
B-Layer	50.5	22	286	13.2
Layer 1	88.5	38	494	22.8
Layer 2	122.5	52	676	31.2

Table A.1: Main parameters of the pixel detector in the barrel region.

Layer	Mean Radius [mm]	Sectors	Modules	Pixels ($\times 10^6$)
Disk 0	495	8	48	4.4
Disk 1	580	8	48	4.4
Disk 2	650	8	48	4.4

Table A.2: Main parameters of the pixel detector in the endcap region.

The basic structures of the active part of the pixel detector are the modules which are composed of silicon sensors, front-end chips and control circuits. Modules are mounted in structures which provide mechanical support and cooling and are called staves in the barrel region and sectors for the disk layers in the end-caps. There are 13 modules on a stave and the stave layout is identical for all pixel layers while in the case of IBL there are 16 modules per stave. The most pixels have size 50×400 μm^2, while IBL pixel dimensions are 50×250 μm^2. In Table A.1 and Table A.2 are summarized the pixel detector parameters for the barrel and the endcap regions respectively.

A.2.1 Sensors and front-end electronics

In this Section, sensors and front-end electronics [166] will be briefly summarized because they are preparatory to understand the errors studies. Since both the sensors and the electronics architecture are different among the pixel layers and the IBL, they will be described separately.

Pixel layers

The sensitive part of the pixel detector is composed of solid state ionization chambers made up by a n-type bulk semiconductor with implantations of high positive p^+ and negative n^+ dose regions on the sides of the wafer. The sensor is reverse polarized in order to extend the depletion region. Each sensor tile contains 47232 pixels. Not all the pixels have a separate readout channel, indeed there are a total of 46080 independent channels per sensor tile.
In the case of pixel layers, 16 front-end chips (FEs) are interfaced with one sensor tile, arranged in two columns, each of them composed by 8 FEs. Considering the high amount of channels and the radiation hardness requirements, the FE chips were realized using the so-called FE-I3 architecture, consisting in a 250 nm CMOS technology. Each FE-I3

contains 2880 pixels cells and each of these pixel cells is connected to an analog circuit where the charge signal is amplified before being digitalized. The signal amplitude can be measured by using the Time over Threshold (ToT), namely the width of the preamplifier output. The duration of the ToT is measured by using a 40 MHz clock, which is the frequency of the LHC bunch-crossings. Finally, the signal passes through a discriminator where it is compared to a programmable threshold. The value of the threshold, the threshold current and some other parameters are then handled by a digital circuit. A 8-bit coded time stamp is then assigned to the leading edge and the trailing edge of each hit. When the hit processing is complete, this information is then transferred to the End Of Column (EoC) using a priority mechanism. This buffering is required because all the hits are transferred to the EoC as soon as they are detected. The hit informations waits then for the Level 1 (L1) trigger decision. Only the hits belonging to events accepted by L1 are stored while the other ones are cleared from the EoC buffers. The hits information is then extracted from the EoC buffers, serialized and sent to the Module Controller Chip (MCC).
The MCC [167] is the chip interfaced with the sixteen FE-I3 chips. It has three main tasks: loading the parameter and configuration data in the FEs, distributing timing signals like L1 trigger and bunch-crossings, and FE-chip readout and event building. Figure A.1 shows a block diagram which describes the MCC internal architecture. In particular, it is composed of:

- **Front end port**: this block interfaces the MCC with the FE chips.
- **Module port**: this block interfaces the MCC to the signals going to the Read-Out system.
- **Command decoder**: this block receives and decodes all the commands sent to the MCC like the L1 trigger command.
- **Register bank**: this block contains the set of configuration registers used by the MCC.
- **TTC (Trigger, Timing and Control)**: after a L1 trigger command is received by the MCC, this block generates the L1 trigger to be sent to the FE chips. It also deals with event memory space overflows by stopping the generation of L1 trigger commands if more than 16 events are pending for transmission to the ROD. Indeed, by construction, the MCC is able to receive a maximum of 16 events, one for each FE.
- **Receiver**: this block is repeated 16 times, one for each FE. The hit information received by the FEs is stored in the ReceiverFIFO. This block is also able to detect the End of Event (EoE) word which determines that the hit information is complete.
- **Event Builder**: this block is responsible for the reconstruction of the event. Each time a L1 trigger is decoded by the Command Decoder, the content of the L1 trigger counter and the bunch crossing clock are stored in the PendingLv1FIFO. The EventScoreboard is used to keep track of the events stored in the Receiver.

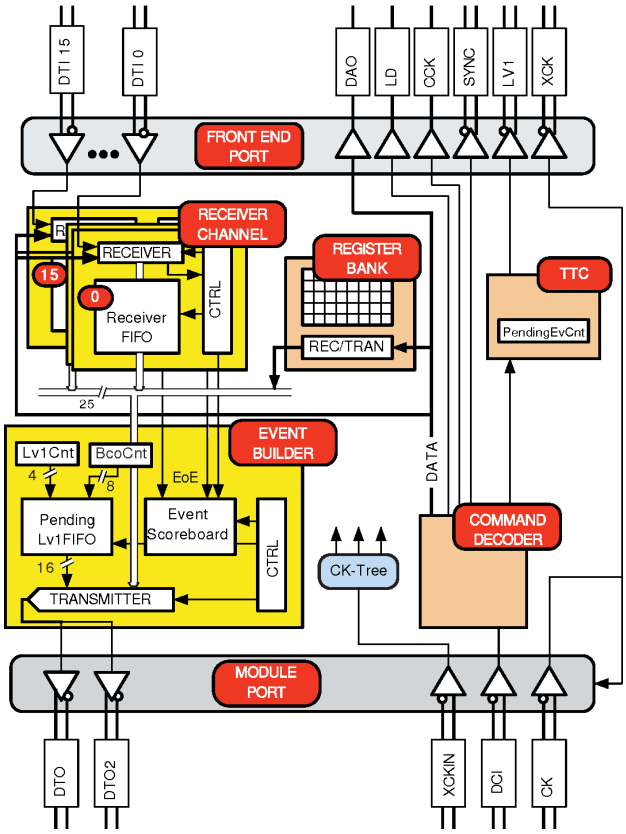


Figure A.1: MCC diagram [167].

When a certain event is fully contained in the MCC, it can be processed by the Event Builder.

IBL

The IBL layer exploits two different technologies for the sensors: both planar and 3D design [168]. Planar sensors are similar to the ones used in the other pixel layers but with stricter requirements. In particular, they have to deal with a fluence of up to $5 \times 10^{15} n_{eq}/\text{cm}^2$. 3D sensors are implemented as well in IBL. They are characterized by a different structure, specifically the electrodes penetrate vertically the bulk instead of being on the surface. 3D sensors need smaller depletion voltage due to smaller distance between the electrodes and have smaller charge-collection distance and consequently shorte pulse rise-time.

The FE electronics interfaced with these sensors is different from that used for the pixel layers. The redesign of the FE-I3 architecture was mainly due to the higher fluence, to wich the IBL is exposed, and the higher hit rates. The main difference is that the FE-I4 is a standalone unit and does not require a MCC. The FE-I4 is composed by an analog and a digital part of the circuit. The analog section is composed by an amplifying architecture

and a discriminator, similarly to the FE-I3 chip, but optimized for having low noise and fast rise-time. In order to deal with the high hit rate, the hits are recorded in local buffers at pixel level. Four pixels share the same digital processing. However, each pixel has its own circuitry where the ToT is measured and the hits are discriminated into small and big hits. Exploiting the fact that the hits are generally clustered, i.e. smaller hits are usually closer to bigger hits, a neighbour logic is used in order to associate small hits to big hits. This results in a reduced time loss. When triggered hits are available, they are sent to the End of Column logic. The peripheral logic, consisting of another block called End of Chip logic, is then responsible for the communication and the operational tunings. These include the 40 MHz clock, trigger commands and global register configuration. When a trigger command is decoded, it is then sent to the End of Column logic and finally to the pixels.

A.2.2 Off-detector readout

Data acquired in the front-end electronics are then moved to the off-detector Read-Out Drivers (RODs). This is relevant for the error studies since errors can also be flagged by the Read-Out system. Each module (FE chip in the case of IBL) is connected to the ROD through optical-fiber links. One link is used to transmit clock, trigger and configuration data, while other two are used for the event readout. The full pixel detector system architecture is summarized in Figure A.2.

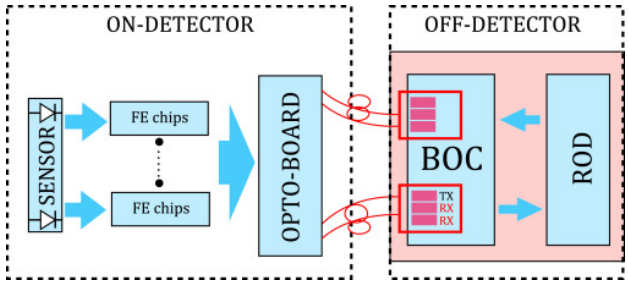


Figure A.2: Pixel detector system architecture split into on-detector electronics and off-detector electronics.

A.3 Errors studies

In this Section, the studies made on the errors flagged by the front-end chips and the read-out system, are described. They are deeply connected to the software maintenance and development work. All the studies have been performed using the ATLAS official software framework called ATHENA [65].

A.3.1 ATHENA framework

ATHENA is the official ATLAS software framework. It is employed for the processing and reconstruction of data and also for simulation and physics analysis. It consists of

many packages written in C++ language, which can be executed using Python scripts, called JobOptions. The latter are very useful for fully configure any Athena component. Two specific packages have been used for these studies: one called "PixelRawDataByteStreamCnv" (Pixel ByteStream Converter) which is able to read and convert data and informations coming from the front-end chips and the read-out and another one called "PixelMonitoring" (Pixel Monitoring) which receives the data decoded by the ByteStream Converter and organizes them in dedicated histograms which are essential for the data quality checks.
The errors studies have been performed running the ATHENA framework over Raw Data, which contains directly the output of the ATLAS detector.

A.3.2 Data format

Pixel

As already mentioned, when a Trigger command is decoded by the MCC data are then trasmitted from the FE chips to the MCC. Each hit information contains a 1-bit header, followed by a 4-bit code for the Level 1 Trigger, an 8-bit row number, a 5-bit column number and finally a 8-bit ToT value. In addition, the FE encodes an End of Event (EoE) word and warning messages. The data format for transmission from the FE to the MCC is shown in Figure A.3.
The data from the 16 FEs are then sent to the MCC. The event data format from the MCC to the ROD, can be seen in Figure A.4. In addition to the hit information aforementioned, 16 bit are dedicated to the MCC- and FE-flag error words. Example one shows an event with two hits belonging to the same FE chip. The second example shows an event with two hits as well but belonging to different FE chips. Finally, the last example shows an event with one hit and one FE-flag error word.

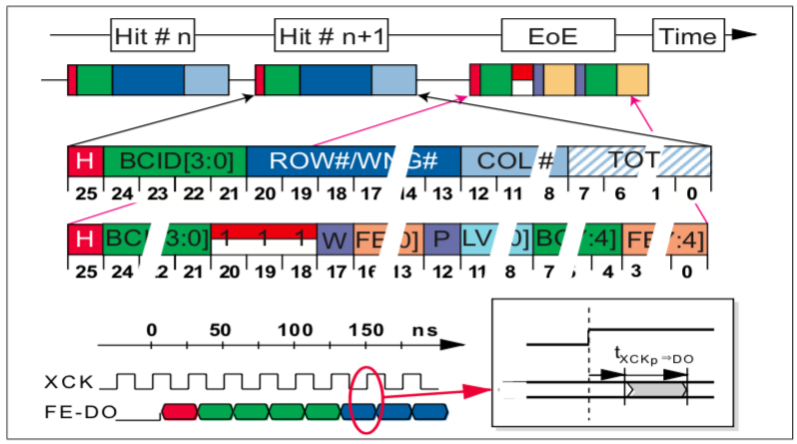


Figure A.3: Data format for event transmission from the FEs to the MCC.

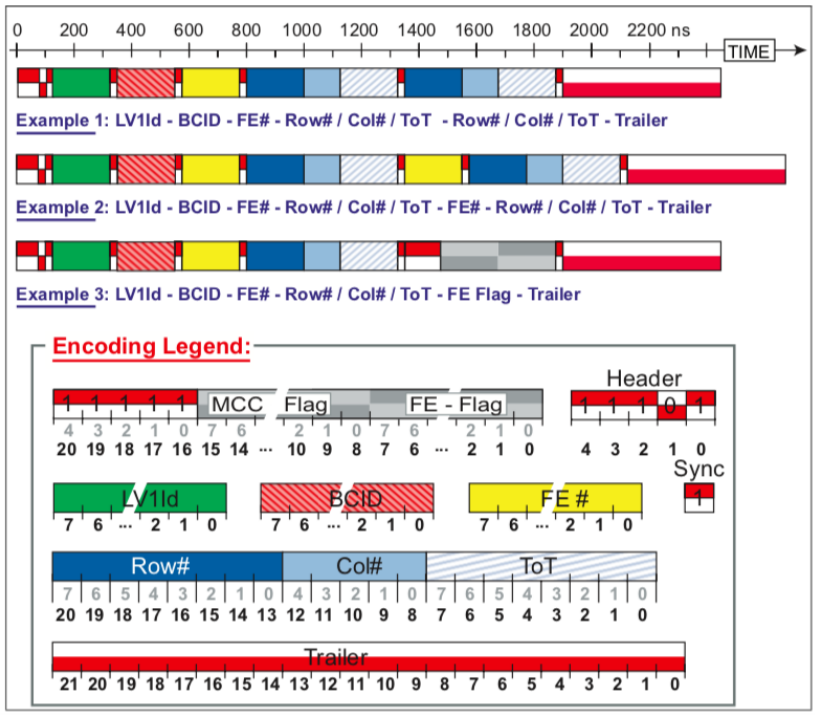


Figure A.4: Event data format at the MCC output.

IBL

In the case of IBL FE architecture, the informations are organized in 24-bit records which can be Data Header, Data Record, Address Record, Value Record or Service Record. The Data Header contains one code for the recognition of the start of transmission (11101), one unique code for data header and a flag which indicates that errors have been logged. In addition, it comprehends the Level 1 trigger counter and the BCID counter. The former has 12 bits, 5 of them are stored in the data header and the other 7 are sent to a special service record when their value changes, while the latter has 13 bits, 10 of them stored in the data header and the remaining 3 sent to a special service record when their value changes. Data Record contains hit informations, namely a 7-bit column number, a 9-bit row number and two 4-bit codes for the ToT where the first one is referred to a given pixel (column, row) and the second one refers to a neighbour pixel in the same column, generally corresponding to row+1. The Address Record gives information on the address of the global register, namely it contains a code for the recognition of the start of transmission (11101), a flag for the Address Record (010), a type information identifying the Global Register or alternatively the Shift Register, and 15-bit code for the address of the Global Register or the Shift Register position. Value Record contains the value of the global register or contained in the shift register, namely a code for the recognition of the start of transmission (11101), a flag for the value record (100) and a 16-bit code for the value. Finally, Service Record is used for services messages like errors

or extra-information. In particular they contain, a code for the recognition of the start of transmission (11101), a flag for the Service Record (111), a 6-bit message code and a 9-bit code for the payload. Service Records are listed in order of priority: at the first available trigger, the highest priority Service Record is written in the bytestream. The payload has the information on how many identical errors occurred since the last publication, therefore a Service Record needs to be weighted by the payload.

A.3.3 Error words

Pixel

The errors studied in this work can be both flagged by the FE or the MCC and by the Read-Out system. In this Section the error words will be listed together with their meaning.
The FE/MCC error words are the following:

- **BCID1**: synchronization error due to inconsistency of the BCID counter inside the receiver. Hits coming from the same FE should have the same BCID number, in the opposite case a BCID1 error word is flagged.
- **BCID2**: synchronization error due to inconsistency of the BCID counter inside the EventBuilder. It checks for consistency between the BCID number of the first FE with data and the current FE.
- **LVL1D**: synchronization error due to inconsistency of the Level 1 trigger counter inside the EventBuilder. It checks for consistency between the LVL1D number of the first FE with data and the current FE.
- **EoC**: truncation error due to full End of Column buffers. It is a warning diectly generated by FEs.
- **Hit Overflow**: truncation error due to rejected hits. This is flagged when the maximum number of hits for a particular receiver is reached and that FE keeps sending data.
- **EoE Overflow**: truncation error due to End of Event words overflow. This happens if, for a particular receiver, there are already 16 End of Event words and another EoE is sent by the FE. In principle, this should never happen since the MCC send a maximum of 16 triggers command to the FE chips by construction. If an EoE overflow error is flagged this is an indication that some data corruption happened in the FE chip, for instance a single event upset, namely a change of state due to a single ionizing particle.
- **SEU Hit Parity**: parity check on the hit words. It is a method which counts the number of one states and produces a single parity bit according to whether an odd or even number of one states was found.
- **SEU Register Parity**: parity check in the command register.

- **SEU Hamming Code**: hit and EoE words are identified with two codes. In the case of a hit word the pattern "0000" is used while for an EoE word the pattern "0111", "1011", "1101", "1110" are used according to the presence of warnings and/or parity errors. In this way an EoE word is checked against pattern with 2 or 4 bit equal to 1 while an hit word is checked against pattern with maximum one bit equal to 1.
- **FE warning**: FE warnings referring to bit-flip.

In addition, the following errors flagged by the ROD can be issued:

- **BCID**: synchronization error due to BCID inconsistency.
- **LVL1D**: synchronization error due to Level 1 trigger inconsistency.
- **H/T Limit error**: truncation error due to the maximum data that can be written in between an header and a trailer word.
- **ROD Overflow**: truncation error due to the maximum allowed amount of data stored.
- **Optical errors**: errors in the optical link.
- **ROD Timeout**: timeout error due to the maximum allowed time for a module to transmit data following a given Level 1 trigger.

IBL

In the case of IBL, the module concept, which is made up by 16 FEs, is not valid anymore, therefore the errors are directly flagged by the single FEs.
The FE error words, also called Service Records, are the following:

- **BCID counter**: synchronization error due to BCID inconsistency in the event.
- **L1 trigger input**: synchronization error referring to the Level 1 trigger input counter written in the End of Double Column Logic (EODCL).
- **L1 trigger request counter**: synchronization error referring to the Level 1 trigger counter read from the EODCL.
- **L1 register**: synchronization error due to full registers. They are used to keep track of the Level 1 triggers sent to the EODCL and pending for readout.
- **L1 trigger ID**: synchronization error due to inconsistency in the L1 trigger ID counter, which is the one stored in the register and sent-out as a part of the Data Header.
- **Skipped trigger**: it indicates that a L1 trigger was skipped due to full registers.
- **Truncated event flag**: it indicates a truncated event.

- **Hamming code**: errors found in hamming codes used for SEU protection.
- **Triple redundant mismatch in CNFGMEM**: SEU detected in the global configuration memory (CNFGMEM). In order to protect against SEU, the logic circuit is triplicated and the output is determined by a majority voting circuit.
- **Bit flip in CMD**: bit-flip detected in the command decoder (CMD) input stream. The CMD is responsible for interpreting the serial commands to control the chip.
- **Triple redundant mismatch in CMD**: SEU detected in the CMD. In order to protect against SEU, the logic circuit is triplicated and the output is determined by a majority voting circuit.
- **Triple redundant mismatch in EFUSE**: SEU detected in the EFUSE, which is another block consisting of triple redundant RAM as CNFGMEM.
- **Readout processor error**: readout error of data in the End of Chip Logic.
- **Write register data error**: SEU detected in the write register storage inside the CMD.
- **Address error**: SEU detected in the address register storage inside the CMD.
- **Other CMD error**: error related to a bit-flip detected in the CMD.
- **Data bus address**: error due to non-existing register address.

The errors flagged by the ROD are instead listed in the folllowing:

- **BCID**: synchronization error due to BCID inconsistency.
- **LVL1D**: synchronization error due to Level 1 trigger inconsistency.
- **Row/Column**: truncation error due to invalid row/column number.
- **H/T Limit error**: truncation error due to the maximum data that can be written in between an header and a trailer word.
- **Optical error**: errors in the optical link.
- **Trailer error**: error flagged when a trailer word is never received after a certain amount of time.
- **Timeout error**: timeout error due to the maximum allowed time for a module to transmit data following a given Level 1 trigger.

A.3.4 Errors

The correlations of errors with occupancy and beam conditions has been investigated in order to estimate their impact on the Run 2 pixel detector operation. This study has also required a work on the Athena software. Indeed, besides small bugs and issues, the software presented two major problems: one concerning the FE flagged errors from pixels which were reported with module granularity and not FE granularity and the other one concerning all the IBL errors which were not implemented. Part of this study was indeed dedicated to help solving these two issues which affected not only the pixel Monitoring package but also the ByteStream Converter.
All the 16 FEs of a module in the pixel layers are able to send error words and by construction there can be a maximum of one error word per FE, therefore a maximum of 16 per module. However, originally, a logic OR was applied to the FE error words, allowing to have a maximum of one error word per module without the possibility to know from which FE the error was flagged. The issue has been fixed allowing to retrieve all the error words with FE granularity.
Some runs during 2016 and 2017 data-taking have been analyzed, preferring those with more extreme conditions like high pile-up and/or high istantaneous luminosity. An example is here documented for a run of 14th October 2016 (run 310574) which was a short run with peak average number of collisions per bunch crossing being, $< \mu >_{\text{peak}} = 50.1$, peak Lumi $= 7.2 \times 10^2 \times 10^{30}$ cm^{-2}s^{-1} and # bunches $= 99$. In the following discussion some results for pixel layers are presented. Layer 2 has however been omitted since there were some disabled modules and the results are consequently unreliable. The output of the Athena monitoring allows to check various histograms which are fundamental for data quality. Errors can be either visualized for individual error words and grouped in categories such as synchronization, truncation, etc.. errors. In addition, different kind of histograms are available: 2D maps which show the number of errors for each module, histograms that show the number of errors as a function of time or more precisely as a function of lumi-blocks, which represent a fraction of time where the instaneous luminosity is approximately constant ($\sim$ one minute) and they are either normalized by the number of events or by the number of events and modules. An example of the Athena monitoring output is displayed in Figure A.5 for the BCID and LVL1D module synchronization errors.

After the update of the Athena framework, it is possible to retrieve all the FE/MCC error words with FE granularity. Figure A.6 shows then the fraction of modules with a given number of FE/MCC error words. The distributions do not follow neither a binomial or a poissonian which implies some kind of correlations among the FEs. Table A.3 summarizes the same information.

Since the purpose of this study was to investigate possible correlations with beam conditions, the number of errors has been checked as a function of the pseudorapidity (therefore integrating over ϕ direction) and as a function of the pile-up. Figure A.7 and Figure A.8 show the results for the BCID and LVL1ID FE/MCC error words respectively. As it can be seen, there is no evidence for a dependence in terms of pseudorapidity or pile-up. The synchronization errors are concentrated on specific modules, with the LVL1ID

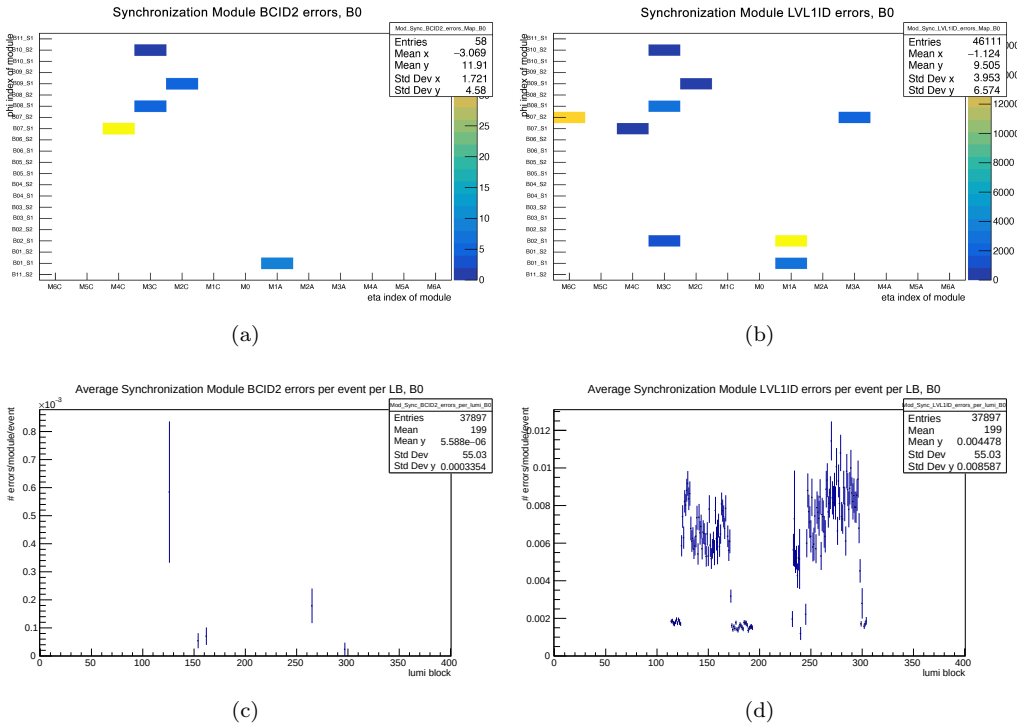


Figure A.5: Map of BCID module synchronization errors (a) and LVL1ID module synchronization errors (b) for Layer 0. In addition, the number of errors per event is shown as a function of lumi-block for the same error words and layer, respectively (c) and (d).

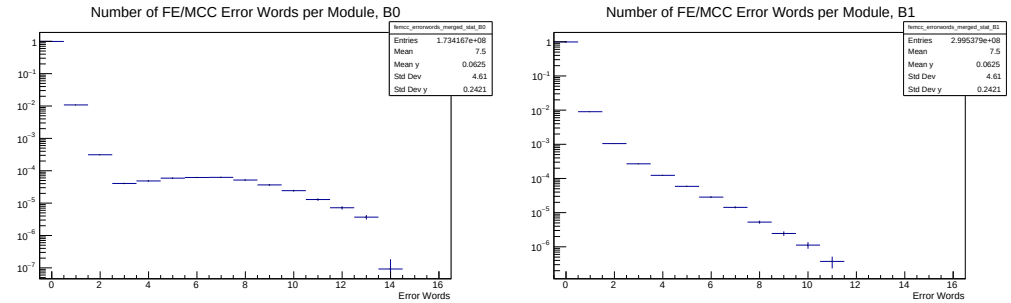


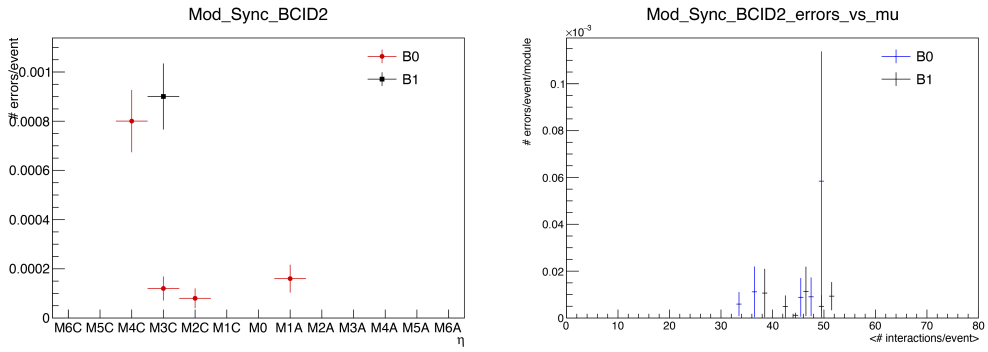
Figure A.6: Fraction of modules with a given number of FE/MCC error words: Layer 0 (a) and Layer 1 (b).

error rate being significantly higher with respect to the BCID.

Figure A.9 and Figure A.10 show again synchronization errors but flagged by the ROD. Very similar behaviour can be noticed between LVL1ID and BCID ROD errors.

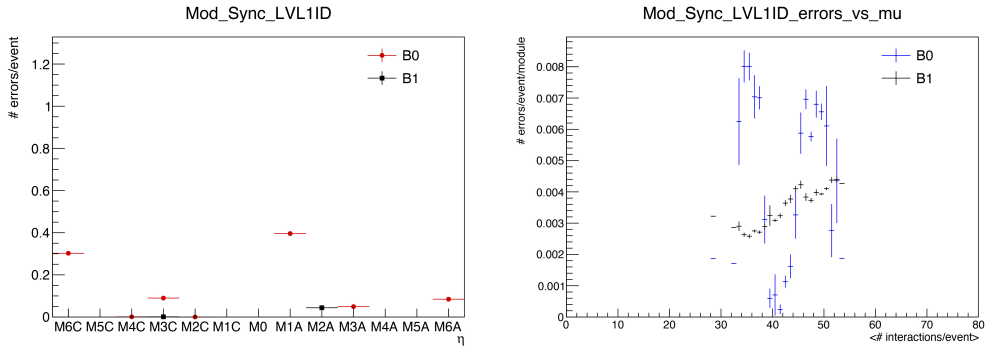
	Layer 0	Layer 1
Frac. of modules with 1 error word	1.1 %	0.9%
Frac. of modules with 2 error words	0.03%	0.1%
Frac.of modules with $\geq$ 2 error words	0.07%	0.2%

Table A.3: Fraction of modules with a given number of FE/MCC error words.



(a) (b)

Figure A.7: Pseudorapidity (a) and pile-up (b) dependence for the BCID FE/MCC synchronization error. Layer 0 and layer 1 behaviour are superimposed in the same plot.

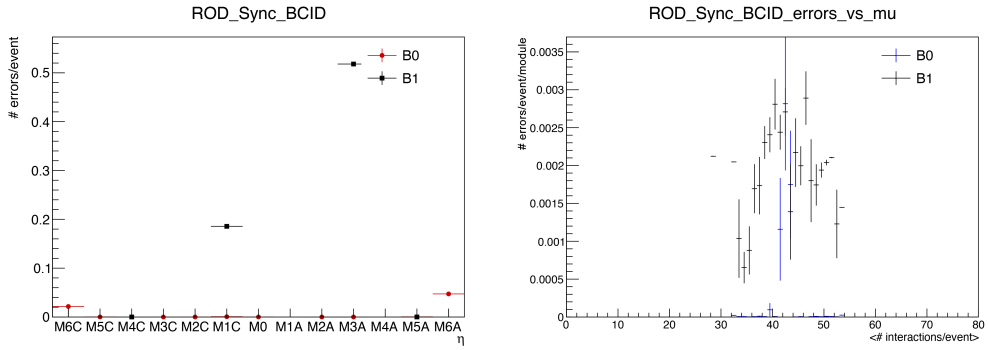


(a) (b)

Figure A.8: Pseudorapidity (a) and pile-up (b) dependence for the LVL1ID FE/MCC synchronization error. Layer 0 and layer 1 behaviour are superimposed in the same plot.

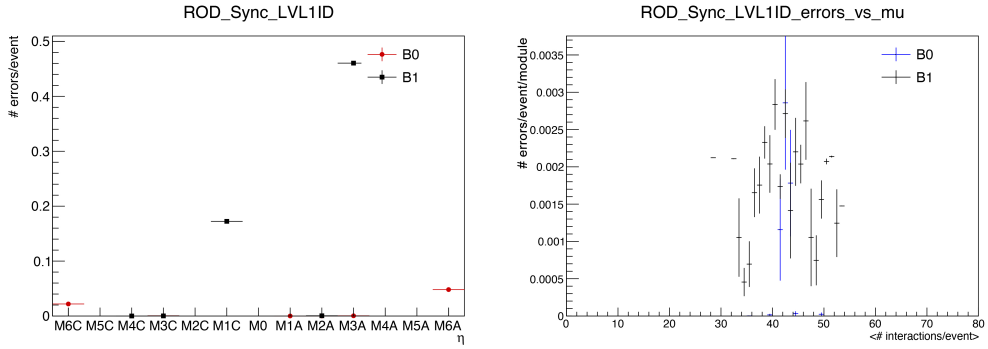
Also for the synchronization errors flagged by the ROD, no specific dependence in terms of pseduorapidity or pile-up can be noticed.

Other frequent errors are the truncation errors flagged by the FEs, namely the EoC which indicates full end of column buffers and the hit overflow which indicates a truncation



(a) (b)

Figure A.9: Pseudorapidity (a) and pile-up (b) dependence for the BCID ROD synchronization error. Layer 0 and layer 1 behaviour are superimposed in the same plot.

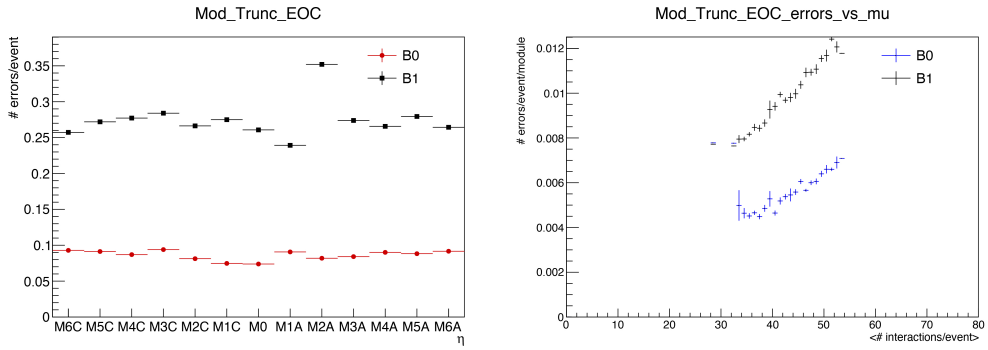


(a) (b)

Figure A.10: Pseudorapidity (a) and pile-up (b) dependence for the LVL1ID FE/MCC synchronization error. Layer 0 and layer 1 behaviour are superimposed in the same plot.

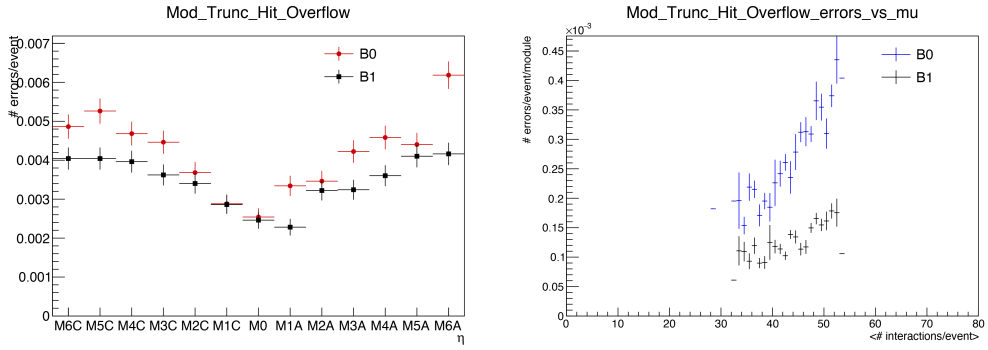
of the event due to the maximum number of hits allowed. Figure A.11 and Figure A.12 show the results for these FE/MCC error words. No significant η dependence can be noticed for the EoC errors, which show a quasi-flat behaviour, while an increase at higher η is shown in the case of Hit Overflow. Instead, as it is reasonably expected, the rate of the errors increase when increasing the average number of interactions per event.

The results obtained for the run 310574 are compared with those of another run with different features. The run used is from 26th October 2016 (run 311481), which has lower average number of interactions per bunch crossing but higher peak luminiosity with respect to the run 310574, namely $< \mu >_{\text{peak}} = 40.1$, peak Lumi $= 1.2 \times 10^4 \times 10^{30}$ cm^{-2}s^{-1} and # bunches $= 2208$.
The FE/MCC synchronization errors are shown in Figure A.13 as a function of the pseudorapidity, comparing the two runs. An higher rate of synchronization errors, either BCID



(a) (b)

Figure A.11: Pseudorapidity (a) and pile-up (b) dependence for the EoC truncation error. Layer 0 and layer 1 behaviour are superimposed in the same plot.



(a) (b)

Figure A.12: Pseudorapidity (a) and pile-up (b) dependence for the Hit Overflow truncation error. Layer 0 and Layer 1 behaviour are superimposed in the same plot.

and LVL1ID, can be seen for the run 311481, as expected as it had an higher trigger rate. The synchronization errors are instead more isolated and concentrated on some specific modules for the run 310574.
Figure A.14 shows the EoC and the hit overflow truncation errors as a function of the pseudorapidity. The EoC, which indicates full end of column buffers, is reported with an higher rate for run 311481 and this is expected because of the higher number of filled bunches. An higher rate of hit overflow errors can be seen for run 311481 as well. The interpretation is more difficult in this specific case as more hits would be expected in the case of the higher pile-up run (310574). The two runs present anyway the same η behaviour.

Other errors have been detected but are not reported here as they are very rare and not statistically significant like single event upset and other ROD errors.

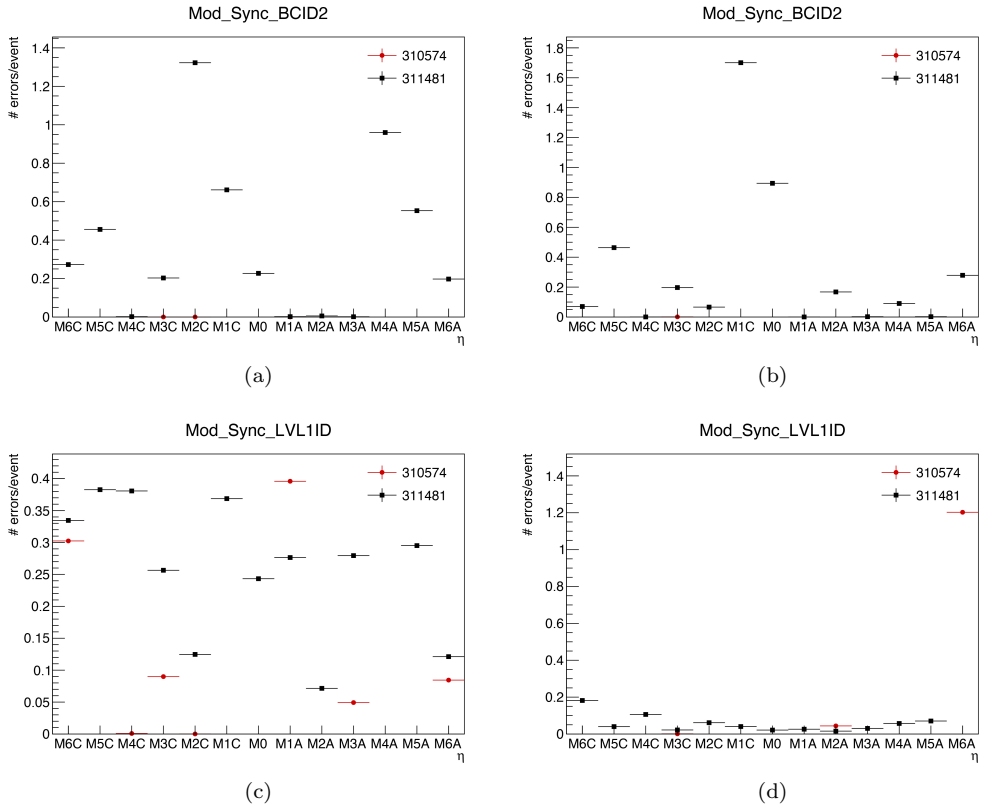


Figure A.13: BCID FE/MCC synchronization errors are shown at the top for Layer 0 (a) and Layer 1 (b) while LVL1ID FE/MCC synchronization errors are shown at the bottom for Layer 0 (c) and Layer 1 (d). The two runs (310574 and 311481) are superimposed in the same plot.

The same analysis can be performed for the IBL FE and ROD errors. However, the IBL errors were not fully implemented in the Athena framework: the monitoring part was not implemented and on the bytestream converter side many modifications have been necessary in order to retrieve errors with FE granularity, which is the only meaningful way to retrieve informations from IBL. Therefore, part of this work required an effort on the software side and has helped to have the final IBL service records and ROD errors available in the monitoring, as demonstrated by Figure A.15.

The studies, presented here, constitute the first approach for investigating the behaviour of FE and ROD flagged errors, and they can be repeated using higher statistics. The statistics is limited by the time-consuming processing of raw data but these studies can be relevant in order to provide inputs for recovery actions during data taking. IBL service records and ROD errors can also be studied more extensively since they are now correctly available in the offline and online monitoring.

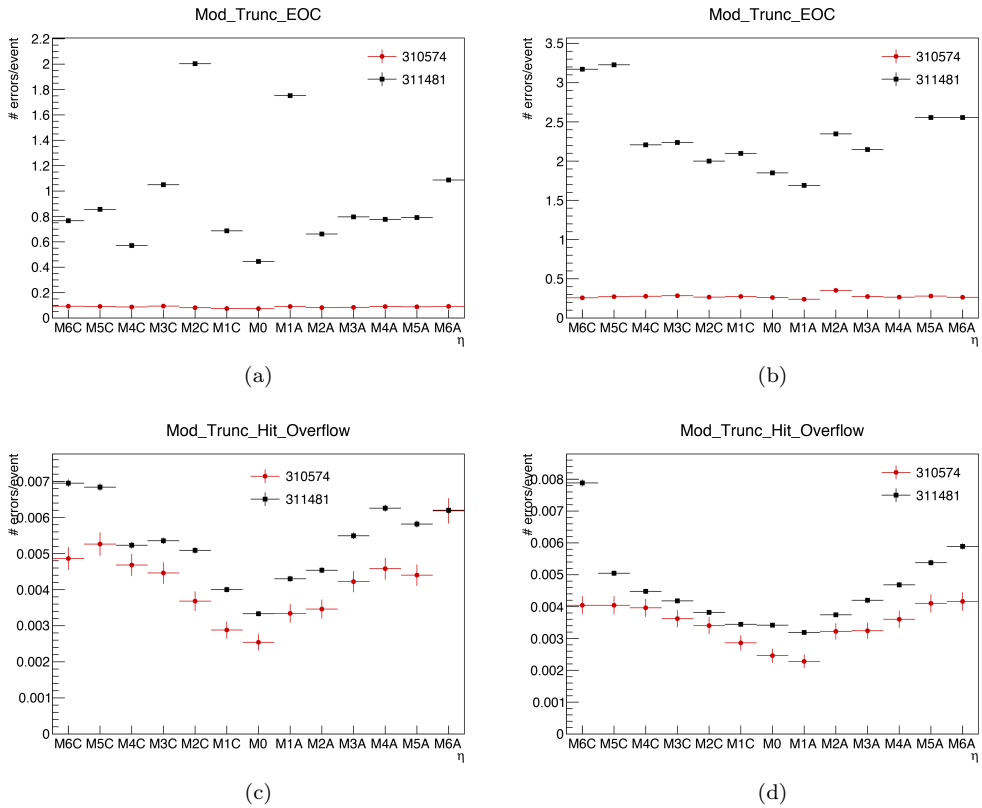
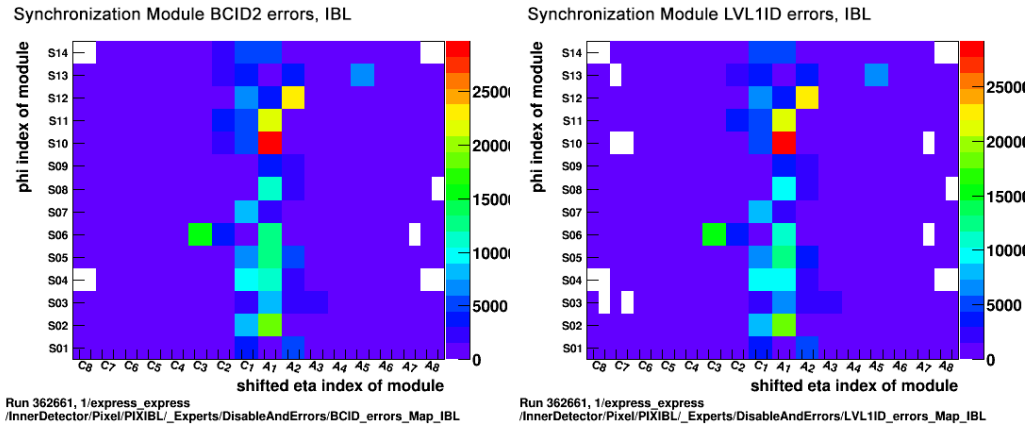


Figure A.14: EoC truncation errors are shown at the top for Layer 0 (a) and Layer 1 (b) while Hit overflow synchronization errors are shown at the bottom for Layer 0 (c) and Layer 1 (d). The two runs (310574 and 311481) are superimposed in the same plot.



(a) (b)

Figure A.15: BCID (a) and LVL1ID (b) service records maps on the online monitoring for a run recorded in 2018 (362661).

Bibliography

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[2] CMS Collaboration, *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] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509, https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[4] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323, https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[5] ATLAS, CMS Collaboration, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv:1503.07589 [hep-ex].

[6] ATLAS, CMS Collaboration, *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].

[7] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32–35, [Usp. Fiz. Nauk161,no.5,61(1991)].

[8] M. Diehl and O. Nachtmann, *Optimal observables for measuring three gauge boson couplings in e+ e- —¿ W+ W-*, pp. , 301–304. 1996. arXiv:hep-ph/9603207 [hep-ph].

[9] M. Schumacher, *Determination of the cp quantum numbers of the Higgs boson and test of cp invariance in the higgs-strahlung process at a future e+ e- linear collider*,.

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

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

[12] A. Salam, *Gauge Unification of Fundamental Forces*, Rev. Mod. Phys. **52** (1980) 525–538, [,306(1980)].

[13] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. http://www.slac.stanford.edu/~mpeskin/QFT.html.

[14] Particle Data Group Collaboration, M. e. a. Tanabashi, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001, https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[15] J. C. Collins, D. E. Soper, and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1–91, arXiv:hep-ph/0409313 [hep-ph].

[16] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298–318.

[17] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641–653, [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[18] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438–450, [Yad. Fiz.15,781(1972)].

[19] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. **C75** no. 5, (2015) 204, arXiv:1412.3989 [hep-ph].

[20] A. Djouadi and R. M. Godbole, *Electroweak symmetry breaking at the LHC*, pp. , 47–74. 2009. arXiv:0901.2030 [hep-ph].

[21] LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3, OPAL, *Search for the standard model Higgs boson at LEP*, Phys. Lett. **B565** (2003) 61–75, arXiv:hep-ex/0306033 [hep-ex].

[22] CDF Collaboration and D0 Collaboration, *Higgs boson studies at the Tevatron*, Phys. Rev. D **88** (2013) 052014, https://link.aps.org/doi/10.1103/PhysRevD.88.052014.

[23] The LEP Electroweak Working Group. http://lepewwg.web.cern.ch/LEPEWWG/.

[24] A Generic Fitter Project for HEP model testing. http://project-gfitter.web.cern.ch/project-gfitter/.

[25] 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], https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG#LHC_Higgs_Cross_Section_Working.

[26] J. Ellis and T. You, *Updated Global Analysis of Higgs Couplings*, JHEP **06** (2013) 103, arXiv:1303.3879 [hep-ph].

[27] S. P. Martin, *A Supersymmetry primer*, arXiv:hep-ph/9709356 [hep-ph], [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[28] V. C. F. W. K. R. M. S. Peterson, C. J.; Rubin, *Extended rotation curves of high-luminosity spiral galaxies. III - The spiral galaxy NGC 7217*, Astrophysical

Journal, Part 1, vol. 226, Dec. 15, 1978.

[29] J. R. Brownstein and J. W. Moffat, *The Bullet Cluster 1E0657-558 evidence shows Modified Gravity in the absence of Dark Matter*, Mon. Not. Roy. Astron. Soc. **382** (2007) 29–47, arXiv:astro-ph/0702146 [astro-ph].

[30] WMAP Collaboration, G. Hinshaw et al., *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, Astrophys. J. Suppl. **208** (2013) 19, arXiv:1212.5226 [astro-ph.CO].

[31] Planck Collaboration, P. A. R. Ade et al., *Planck 2015 results. XIII. Cosmological parameters*, Astron. Astrophys. **594** (2016) A13, arXiv:1502.01589 [astro-ph.CO].

[32] A. Riotto and M. Trodden, *Recent progress in baryogenesis*, Ann. Rev. Nucl. Part. Sci. **49** (1999) 35–75, arXiv:hep-ph/9901362 [hep-ph].

[33] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, JHEP **10** (2010) 085, arXiv:1008.4884 [hep-ph].

[34] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the* 2π *Decay of the* K_2^0 *Meson*, Phys. Rev. Lett. **13** (1964) 138–140, https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[35] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[36] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533, [,648(1963)].

[37] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[38] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*, CERN-2004-003-V1, CERN-2004-003, CERN-2004-003-V-1 (2004).

[39] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[40] LEP Collaboration, *LEP Design Report Vol.1*, CERN-LEP-TH-83-29, CERN-PS-DL-83-31, CERN-SPS-83-26, LAL-RT-83-09 (1983).

[41] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[42] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[43] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[44] ATLAS Collaboration, *ATLAS Luminosity public results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[45] ATLAS Collaboration, *ATLAS magnet system: Technical design report*, CERN-LHCC-97-18 (1997).

[46] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 1*, CERN-LHCC-97-16, ATLAS-TDR-4 (1997).

[47] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 2*, CERN-LHCC-97-17 (1997).

[48] ATLAS Collaboration, *ATLAS pixel detector: Technical design report*, CERN-LHCC-98-13 (1998).

[49] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19 (2010).

[50] ATLAS Collaboration,, *Track Reconstruction Performance of the ATLAS Inner Detector at* $\sqrt{13}$ *TeV*,.

[51] ATLAS Collaboration, *The barrel modules of the ATLAS semiconductor tracker*, Nucl. Instrum. Meth. **A568** (2006) 642–671.

[52] ATLAS Collaboration, *The ATLAS semiconductor tracker end-cap module*, Nucl. Instrum. Meth. **A575** (2007) 353–389.

[53] ATLAS Collaboration, *The ATLAS TRT barrel detector*, JINST **3** (2008) P02014.

[54] ATLAS Collaboration, *The ATLAS TRT end-cap detectors*, JINST **3** (2008) P10003.

[55] ATLAS Collaboration, *ATLAS liquid argon calorimeter: Technical design report*, CERN-LHCC-96-41 (1996).

[56] ATLAS Collaboration, *ATLAS tile calorimeter: Technical design report*, CERN-LHCC-96-42 (1996).

[57] ATLAS Collaboration, *The ATLAS forward calorimeters*, JINST **3** (2008) P02010.

[58] ATLAS Collaboration, *ATLAS muon spectrometer: Technical design report*, CERN-LHCC-97-22, ATLAS-TDR-10 (1997).

[59] ATLAS Collaboration, *ATLAS first level trigger: Technical design report*, CERN-LHCC-98-14, ATLAS-TDR-12 (1998).

[60] ATLAS Collaboration, *ATLAS high-level trigger, data acquisition and controls: Technical design report*, CERN-LHCC-2003-022, ATLAS-TRD-016 (2003).

[61] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, The European Physical Journal C **77** no. 5, (2017) 317, `https://doi.org/10.1140/epjc/s10052-017-4852-3`.

[62] ATLAS TDAQ Collaboration, *The ATLAS Data Acquisition and High Level Trigger system*, Journal of Instrumentation **11** no. 06, (2016) P06008–P06008, `https://doi.org/10.1088%2F1748-0221%2F11%2F06%2Fp06008`.

[63] ATLAS Collaboration, *ATLAS data preparation in run 2*, J. Phys. Conf. Ser. **898** no. 4, (2017) 042050.

[64] ATLAS Collaboration,, *Data Quality public results*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/RunStatsPublicResults2010`.

[65] ATLAS Collaboration, *ATLAS computing: Technical design report*, CERN-LHCC-2005-022, ATLAS-TRD-017 (2005).

[66] S. Höche, *Introduction to parton-shower event generators*, pp. , 235–295. 2015. `arXiv:1411.4085 [hep-ph]`.

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

[68] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. **D93** no. 3, (2016) 033006, arXiv:1506.07443 [hep-ph].

[69] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303, https://geant4.web.cern.ch/.

[70] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, ATLAS-CONF-2012-042 (2012).

[71] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444–450.

[72] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. **C74** no. 10, (2014) 3071, arXiv:1407.5063 [hep-ex].

[73] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2016-015 (2016).

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

[75] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s}$ *=13 TeV*, Eur. Phys. J. **C76** no. 5, (2016) 292, arXiv:1603.05598 [hep-ex].

[76] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at* $\sqrt{s} = 13$ TeV *with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, https://link.aps.org/doi/10.1103/PhysRevD.96.072002.

[77] ATLAS Collaboration, *Properties of jets and inputs to jet reconstruction and calibration with the ATLAS detector using proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-036 (2015).

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

[79] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. **C76** no. 11, (2016) 581, arXiv:1510.03823 [hep-ex].

[80] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012 (2016).

[81] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045 (2015).

[82] ATLAS 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).

[83] ATLAS Collaboration, *Performance of missing transverse momentum*

reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s}$ = 13 TeV, Eur. Phys. J. **C78** no. 11, (2018) 903, arXiv:1802.08168 [hep-ex].

[84] ATLAS Collaboration, *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector*, Phys. Lett. **B786** (2018) 59–86, arXiv:1808.08238 [hep-ex].

[85] CMS Collaboration,, *Observation of Higgs boson decay to bottom quarks*, Phys. Rev. Lett. **121** no. 12, (2018) 121801, arXiv:1808.08242 [hep-ex].

[86] CMS Collaboration, *Observation of $t\bar{t}H$ Production*, Phys. Rev. Lett. **120** (2018) 231801, https://link.aps.org/doi/10.1103/PhysRevLett.120.231801.

[87] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. **B784** (2018) 173–191, arXiv:1806.00425 [hep-ex].

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

[89] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. **D99** (2019) 072001, arXiv:1811.08856 [hep-ex].

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

[91] 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].

[92] 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].

[93] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, *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].

[94] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222, arXiv:1309.0017 [hep-ph].

[95] 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].

[96] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv:1405.0301 [hep-ph].

[97] 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].

[98] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. **C76** no. 4, (2016) 196, arXiv:1512.01178 [hep-ph].

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

[100] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv:1406.3660 [hep-ex].

[101] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv:hep-ph/0201195 [hep-ph].

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

[103] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244–289, arXiv:1207.1303 [hep-ph].

[104] 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].

[105] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv:0811.4622 [hep-ph].

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

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

[108] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv:1207.5030 [hep-ph].

[109] K. Melnikov and F. Petriello, *Electroweak gauge boson production at hadron colliders through O(alpha(s)**2)*, Phys. Rev. **D74** (2006) 114017, arXiv:hep-ph/0609070 [hep-ph].

[110] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, Phys. Rev. **D69** (2004) 094008, arXiv:hep-ph/0312266 [hep-ph].

[111] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv:1112.5675 [hep-ph].

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

[113] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. **A654** (2011) 481–489, arXiv:1012.4686 [hep-ex].

[114] R. K. Ellis, I. Hinchliffe, M. Soldate, and J. J. van der Bij, *Higgs Decay to tau+ tau-: A Possible Signature of Intermediate Mass Higgs Bosons at the SSC*, Nucl. Phys. **B297** (1988) 221–243.

[115] I. W. Stewart and F. J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, Phys. Rev. **D85** (2012) 034011, arXiv:1107.2117 [hep-ph].

[116] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, JHEP **07** (2011) 018, arXiv:1105.0020 [hep-ph].

[117] J. M. Campbell and R. K. Ellis, *An Update on vector boson pair production at hadron colliders*, Phys. Rev. **D60** (1999) 113006, arXiv:hep-ph/9905386 [hep-ph].

[118] J. M. Campbell, R. K. Ellis, and W. T. Giele, *A Multi-Threaded Version of MCFM*, Eur. Phys. J. **C75** no. 6, (2015) 246, arXiv:1503.06182 [physics.comp-ph].

[119] B. R. Webber, *Fragmentation and hadronization*, Int. J. Mod. Phys. **A15S1** (2000) 577–606, arXiv:hep-ph/9912292 [hep-ph], [eConfC990809,577(2000)].

[120] A. Murrone, *Measurements of Higgs boson production cross section and CP violation in the H decay channel with the ATLAS detector*, Nuovo Cim. **C41** no. 3, (2018) 97.

[121] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, https://cds.cern.ch/record/1966419.

[122] M. Guzzi, P. Nadolsky, E. Berger, H.-L. Lai, F. Olness, and C. P. Yuan, *CT10 parton distributions and other developments in the global QCD analysis*, arXiv:1101.0561 [hep-ph].

[123] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[124] C. W. Bauer, S. Fleming, D. Pirjol, and I. W. Stewart, *An Effective field theory for collinear and soft gluons: Heavy to light decays*, Phys. Rev. **D63** (2001) 114020, arXiv:hep-ph/0011336 [hep-ph].

[125] 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].

[126] A. Basalaev and Z. M. and, *The Fast Simulation Chain for ATLAS*, Journal of Physics: Conference Series **898** (2017) 042016, https://doi.org/10.1088%2F1742-6596%2F898%2F4%2F042016.

[127] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** no. 07, (2018) P07017.

[128] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *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)].

[129] P. E. S. Neyman Jerzy and P. Karl, *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character**231** 694–706.

[130] C. C. ATLAS and L. H. C. Group, *Procedure for the LHC Higgs boson search combination in summer 2011*, ATL-PHYS-PUB-2011-011, CMS-NOTE-2011-005 (2011).

[131] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219–228.

[132] F. J., *353QH twice smoothing algorithm*, Proc. CERN School of Computing (1974).

[133] ROOT Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016 (2012).

[134] F. Tackmann, K. Tackmann, K. Tackmann, M. Duehrssen-Debling, and P. Francavilla, *Simplified template cross sections*, `https://cds.cern.ch/record/2138079`.

[135] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. **B726** (2013) 120–144, `arXiv:1307.1432 [hep-ex]`.

[136] ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, Eur. Phys. J. **C75** no. 10, (2015) 476, `arXiv:1506.05669 [hep-ex]`, [Erratum: Eur. Phys. J.C76,no.3,152(2016)].

[137] CMS Collaboration, *Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs*, Phys. Rev. Lett. **110** no. 8, (2013) 081803, `arXiv:1212.6639 [hep-ex]`.

[138] ATLAS Collaboration, *Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. **D98** (2018) 052005, `arXiv:1802.04146 [hep-ex]`.

[139] ATLAS Collaboration, *Measurement of the Higgs boson coupling properties in the $H \to ZZ^* \to 4\ell$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **03** (2018) 095, `arXiv:1712.02304 [hep-ex]`.

[140] CMS Collaboration, *Combined search for anomalous pseudoscalar HVV couplings in VH($H \to b\bar{b}$) production and $H \to VV$ decay*, Phys. Lett. **B759** (2016) 672–696, `arXiv:1602.04305 [hep-ex]`.

[141] CMS Collaboration, *Constraints on anomalous Higgs boson couplings using production and decay information in the four-lepton final state*, Phys. Lett. **B775** (2017) 1–24, `arXiv:1707.00541 [hep-ex]`.

[142] CMS Collaboration, *Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state*, Phys. Rev. **D99** no. 11, (2019) 112003, `arXiv:1901.00174 [hep-ex]`.

[143] CMS Collaboration, *Constraints on anomalous HVV couplings from the production of Higgs bosons decaying to τ lepton pairs*, Submitted to: Phys. Rev. (2019), `arXiv:1903.06973 [hep-ex]`.

[144] ATLAS Collaboration, *Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector*, Eur. Phys. J. **C76** no. 12, (2016) 658, `arXiv:1602.04516 [hep-ex]`.

[145] ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \to \tau\tau$ channel in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2019-050 (2019).

[146] L3 Collaboration, *Search for anomalous couplings in the Higgs sector at LEP*, Phys. Lett. **B589** (2004) 89–102, `arXiv:hep-ex/0403037 [hep-ex]`.

[147] V. Hankele, G. Klamke, D. Zeppenfeld, and T. Figy, *Anomalous Higgs boson couplings in vector boson fusion at the CERN LHC*, Phys. Rev. **D74** (2006)

095001, arXiv:hep-ph/0609075 [hep-ph].

[148] K. Hagiwara, R. Szalapski, and D. Zeppenfeld, *Anomalous Higgs boson production and decay*, Phys. Lett. **B318** (1993) 155–162, arXiv:hep-ph/9308347 [hep-ph].

[149] O. J. P. Eboli, M. C. Gonzalez-Garcia, S. M. . Lietti, and S. F. Novaes, *Probing intermediate mass Higgs interactions at the CERN Large Hadron Collider*, Phys. Lett. **B478** (2000) 199–207, arXiv:hep-ph/0001030 [hep-ph].

[150] A. V. Manohar and M. B. Wise, *Modifications to the properties of the Higgs boson*, Phys. Lett. **B636** (2006) 107–113, arXiv:hep-ph/0601212 [hep-ph].

[151] OPAL Collaboration, *Measurement of W boson polarizations and CP violating triple gauge couplings from W^+W^- production at LEP*, Eur. Phys. J. **C19** (2001) 229–240, arXiv:hep-ex/0009021 [hep-ex].

[152] S. Schael, R. Barate, R. Brunelière, I. Bonis, D. Decamp, C. Goy, S. Jézéquel, J.-P. Lees, F. Martin, E. Merle, M.-N. Minard, B. Pietrzyk, B. Trocmé, S. Bravo, M. Casado, M. Chmeissani, J. Crespo, E. Fernandez, M. Fernandez-Bosman, and I. Videau, *Improved measurement of the triple gauge-boson couplings γWW and ZWW in e^+e_- collisions*, Physics Letters B **614** (2005) 7–26.

[153] R. A. Diaz and R. Martinez, *The Custodial symmetry*, Rev. Mex. Fis. **47** (2001) 489–492, arXiv:hep-ph/0302058 [hep-ph].

[154] OPAL Collaboration, *W^+W^- production and triple gauge boson couplings at LEP energies up to 183-GeV*, Eur. Phys. J. **C8** (1999) 191–215, arXiv:hep-ex/9811028 [hep-ex].

[155] OPAL Collaboration, *Measurement of triple gauge boson couplings from W^+W^- production at LEP energies up to 189-GeV*, Eur. Phys. J. **C19** (2001) 1–14, arXiv:hep-ex/0009022 [hep-ex].

[156] OPAL Collaboration, *Measurement of charged current triple gauge boson couplings using W pairs at LEP*, Eur. Phys. J. **C33** (2004) 463–476, arXiv:hep-ex/0308067 [hep-ex].

[157] OPAL Collaboration, *A Test of CP invariance in Z0 —¿ tau+ tau- using optimal observables*, Z. Phys. **C66** (1995) 31–44.

[158] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, Comput. Phys. Commun. **195** (2015) 161–171, arXiv:1412.5390 [hep-ph].

[159] ATLAS Collaboration, *Measurement of the cross-section for electroweak production of dijets in association with a Z boson in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Lett. **B775** (2017) 206–228, arXiv:1709.10264 [hep-ex].

[160] A. Murrone, *Study of the use of Optimal Observables for the Higgs boson CP determination in the vector boson fusion process*, Master's thesis, Università degli Studi di Milano, 2015. http://www.infn.it/thesis/thesis_dettaglio.php?tid=10011.

[161] http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2015-06/.

[162] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and regression*

trees. Wadsworth, Stamford, 1984.

[163] A. e. a. Hocker, *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide. TMVA - Toolkit for Multivariate Data Analysis*, Tech. Rep. physics/0703039, CERN, Geneva, Mar, 2007. https://cds.cern.ch/record/1019880. TMVA-v4 Users Guide: 135 pages, 19 figures, numerous code examples and references.

[164] J. Neyman, *On the Problem of Confidence Intervals*, Ann. Math. Statist. **6** no. 3, (1935) 111–116, https://doi.org/10.1214/aoms/1177732585.

[165] F. Maltoni, K. Mawatari, and M. Zaro, *Higgs characterisation via vector-boson fusion and associated production: NLO and parton-shower effects*, Eur. Phys. J. **C74** no. 1, (2014) 2710, arXiv:1311.1829 [hep-ph].

[166] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007.

[167] R. Beccherle et al., *MCC: The Module Controller Chip for the ATLAS pixel detector*, Nucl. Instrum. Meth. **A492** (2002) 117–133.

[168] G. F. D. Betta et al., *Recent developments and future perspectives in 3D silicon radiation sensors*, Journal of Instrumentation **7** no. 10, (2012) C10006–C10006, https://doi.org/10.1088%2F1748-0221%2F7%2F10%2Fc10006.

Acknowledgments

At the end of these three years and very intense period of my life, there are many people I would like to thank.
First of all, I would like to thank my supervisor Attilio for his help in developing this work, despite very busy and hard times, for his patience and finally for his extensive knowledge and for giving me precious advices. Then, I would like to extend my thanks to all the Higgs to tau tau team and in particular to the Goettingen and Freiburg guys with which I worked more closely. Thanks to all of you for the hard but worthwhile work of these years. Finally, a special thanks to all the people in the Milano ATLAS team.
Numerous friends accompanied me during this path. I start with those with which I shared many lunches, breaks and chats at university: Betta, Elia, Marta, Benni, Umberto. Thanks to all of you for your friendship, for making my days better even in hard times. There are many other friends who have been important during these years, some of them are a fundamental presence despite the distance, some of them I met again in my path in misterious ways. In particular I would like to thank: Cristina and Samuele with the joyful Giacomo and Chiara, Annalisa and Paolo, Chiara and Andrea and Valeria and Thomas with the little David.
Finally, I would like to thank my family for their continuos support throughout all this path and all the choices I made and for their presence even in distance, in particular thanks to my brothers and sister. I would also like to thank Luca's family for making me feel like at home. A very special thanks to Luca, my new family, for his lovely presence, for his support and for always believing in me. We are looking towards new great adventures during this year.