Unveiling the Higgs sector with tau-leptons: differential cross-section measurements and searches for lepton-flavor-violating decays with the ATLAS detector

Antonio Jesús Gómez Delegido
Departament de Física Atòmica, Molecular i Nuclear
Institut de Física Corpuscular (IFIC, CSIC-UV)

Director
Luca Fiorini



VNIVERSITAT DE VALÈNCIA

Programa de Doctorat en Física – 3126

November 2024

DECLARACIÓ

Dr. **Luca Fiorini**, Profesor Titular de la Universitat de València,

CERTIFIQUE

Que la present memòria, titulada ***Unveiling the Higgs sector with tau-leptons: differential cross-section measurements and searches for lepton-flavor-violating decays with the ATLAS detector***, ha sigut realitzada sota la seua direcció en el Departament de Física Atòmica, Molecular i Nuclear de la Universitat de València per **Antonio Jesús Gómez Delegido**, i constitueix la seua Tesi Doctoral per a optar al títol de Doctor per la Universitat de València una vegada cursats els estudis en el Doctorat en Física.

I, perquè en quede constància en compliment de la legislació vigent, signem el present certificat en Burjassot, a 28 de novembre de 2024.

Signat: **Luca Fiorini**

Contents

Preface **1**

1 Introduction to Higgs boson physics **5**
1.1 A brief summary of the Standard Model 5
1.2 The structure of the SM and the necessity of the Higgs mechanism . 6
1.2.1 Quantum chromodynamics and the strong force 7
1.2.2 The electroweak interaction 8
1.3 The spontaneous symmetry breaking 11
1.4 Production and decay of the 125 GeV Higgs boson: phenomenology for the LHC 15
1.5 Experimental state-of-the-art of Higgs boson physics 20
1.6 Lepton flavor violation in the Higgs sector 28
1.7 The Higgs sector as a portal to new physics beyond the SM 31

2 The LHC and the ATLAS detector **35**
2.1 The Large Hadron Collider 35
2.2 The ATLAS detector 37
2.2.1 ATLAS coordinate system 38
2.2.2 Inner Detector 38
2.2.3 Electromagnetic calorimeter 41
2.2.4 Hadronic calorimeter 41
2.2.4.1 Tile Calorimeter 41
2.2.4.2 End-cap and forward calorimeters 46
2.2.5 Muon spectrometer 47
2.2.6 Magnet system 49
2.2.7 Forward detectors 49
2.2.8 Trigger and Data Acquisition system 50
2.2.9 The LHC computing grid 51

3 The Phase II upgrade of the ATLAS Tile Calorimeter and the simulation of its read-out electronics **53**
3.1 The upgrade of the ATLAS detector towards the HL-LHC 53

3.2 The Tile Calorimeter upgrade . . . 54
3.3 Reconstruction of the Tile Calorimeter signal from Run 1 to Run 3 56
3.4 The Pulse Simulator tool . . . 59
3.5 Updates in the Pulse Simulator for the Phase II upgrade . . . 60
3.5.1 Bit range and gains . . . 61
3.5.2 Number of readout samples . . . 62
3.5.3 Pedestal and noise values . . . 63
3.5.4 Pulse shape . . . 65
3.5.4.1 The TileCal test beam experimental setup . . . 66
3.5.4.2 Pulse shape measurement with test beam data . . . 67
3.5.4.3 The pulse shape undershoot using laboratory data 70
3.5.5 Pile-up simulation . . . 75
3.5.6 Continuous read-out simulation . . . 78

4 Data and simulated samples 83
4.1 Proton-proton interactions at the LHC . . . 83
4.2 Event simulation . . . 84
4.2.1 Matrix element and parton showering . . . 85
4.2.2 Matching . . . 86
4.2.3 Hadronization . . . 87
4.2.4 Underlying event and pile-up . . . 88
4.2.5 Detector simulation . . . 88
4.2.6 Monte Carlo generators . . . 89
4.3 Data and simulated samples . . . 90
4.3.1 Simulation of Higgs boson samples . . . 91
4.3.2 Simulation of background samples . . . 93

5 Object reconstruction and identification 95
5.1 Tracks and vertices . . . 95
5.2 Electrons . . . 97
5.2.1 Reconstruction . . . 97
5.2.2 Identification . . . 98
5.2.3 Isolation . . . 99
5.3 Muons . . . 100
5.3.1 Reconstruction . . . 100
5.3.2 Identification . . . 101
5.3.3 Isolation . . . 103
5.4 Jets . . . 104
5.4.1 b-jet tagging . . . 106
5.5 Hadronic τ-leptons . . . 108
5.6 Missing transverse momentum . . . 110

6 Differential cross-section measurements of the Higgs boson production in the $H \to \tau\tau$ decay channel 113
6.1 Analysis motivation and strategy . . . 113

6.2 Object definition and trigger selection . . . 116
6.2.1 Trigger selection . . . 117
6.2.2 Object definition . . . 118
6.3 Higgs mass and transverse momentum reconstruction . . . 119
6.3.1 Higgs mass reconstruction . . . 119
6.3.2 Transverse momentum reconstruction . . . 122
6.4 Multivariate analysis . . . 124
6.4.1 The VBF tagger . . . 124
6.4.2 The VH tagger . . . 126
6.4.3 The $t\bar{t}H$ multiclassifier . . . 126
6.5 Event selection and categorization . . . 128
6.5.1 Event selection for the STXS measurement . . . 130
6.5.1.1 STXS Stage 1.2 in the $H \to \tau\tau$ analysis . . . 131
6.5.1.2 VBF categorization . . . 133
6.5.1.3 VH categorization . . . 134
6.5.1.4 $t\bar{t}H$ categorization . . . 135
6.5.1.5 Boost categorization . . . 136
6.5.2 Event selection for the fiducial differential cross-section measurement . . . 137
6.5.2.1 Fiducial volume definition . . . 139
6.5.2.2 Event selection at reconstruction level . . . 140
6.5.2.3 Differential variables and binning . . . 141
6.6 Background estimation . . . 142
6.6.1 $Z \to \tau\tau$ background . . . 143
6.6.1.1 Object-level embedding . . . 143
6.6.1.2 Treatment of electroweak $Z \to \tau\tau$ + jets production 145
6.6.2 Misidentified background . . . 146
6.6.2.1 Misidentified background in the $\tau_e\tau_\mu$ channel . . . 147
6.6.2.2 Misidentified background in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel . . . 148
6.6.2.3 Misidentified background in the $\tau_{\text{had}}\tau_{\text{had}}$ channel . . . 150
6.6.2.4 Combination of the misidentified background templates in the VBF categories . . . 151
6.6.3 Other backgrounds . . . 153
6.7 Systematic uncertainties . . . 153
6.7.1 Theoretical uncertainties on the signal prediction . . . 154
6.7.2 Theoretical uncertainties on the background prediction . . . 156
6.7.3 Experimental uncertainties . . . 157
6.8 Statistical analysis . . . 161
6.8.1 Fit model for the STXS measurement . . . 162
6.8.1.1 Simplifying the likelihood model: pruning, symmetrization and smoothing . . . 165
6.8.2 Fit model for the fiducial differential cross-section measurement 167
6.9 Simplified template cross-section measurement results . . . 168
6.9.1 Post-fit distributions . . . 168

6.9.2 Background normalization factors . . . 168
6.9.3 Nuisance parameters . . . 182
6.9.4 Inclusive cross-section measurements . . . 192
6.9.5 STXS measurements . . . 195
6.10 Unfolded fiducial differential cross-section measurement results . . . 203
6.10.1 Interpretation in the Standard Model Effective Field Theory framework . . . 207

7 Searches for lepton-flavor-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ 213
7.1 Analysis motivation and strategy . . . 213
7.2 Object definition and trigger selection . . . 215
7.2.1 Object definition . . . 215
7.2.2 Trigger selection . . . 216
7.3 Event selection and categorization . . . 217
7.3.1 Selection in the MC-based $\ell\tau_{\ell'}$ channel . . . 217
7.3.1.1 Lepton assignment . . . 217
7.3.1.2 Preselection . . . 219
7.3.1.3 Categorization . . . 219
7.3.2 Selection in the $\ell\tau_{\text{had}}$ channel . . . 220
7.4 Background estimation . . . 220
7.4.1 MC-based $\ell\tau_{\ell'}$ channel . . . 223
7.4.1.1 $Z \to \tau\tau$ background . . . 224
7.4.1.2 Top-quark background . . . 225
7.4.1.3 Diboson background . . . 225
7.4.1.4 $Z \to \ell\ell$ background . . . 226
7.4.1.5 Misidentified objects . . . 228
7.4.2 MC-based $\ell\tau_{\text{had}}$ channel . . . 232
7.5 Multivariate analysis . . . 234
7.5.1 Multivariate techniques in the MC-based $\ell\tau_{\ell'}$ channel . . . 234
7.5.1.1 BDTs in the non-VBF category . . . 235
7.5.1.2 BDTs in the VBF category . . . 245
7.5.2 BDTs in the MC-based $\ell\tau_{\text{had}}$ channel . . . 252
7.6 Systematic uncertainties . . . 253
7.6.1 Theoretical uncertainties on the Higgs boson production . . . 254
7.6.2 Theoretical uncertainties on the $Z \to \tau\tau$, $Z \to \ell\ell$ and diboson processes . . . 255
7.6.3 Experimental uncertainties . . . 255
7.6.3.1 Uncertainties on the misidentified background estimation . . . 256
7.7 Statistical analysis . . . 258
7.8 Results . . . 260
7.8.1 Combined measurement of $H \to e\tau$ and $H \to \mu\tau$ . . . 260

7.8.1.1 Normalization factors for the background processes and nuisance parameters . . . 261
7.8.1.2 Best-fit results for the branching ratios and exclusion limits . . . 262
7.8.1.3 Compatibility among categories in the MC-based analysis . . . 266

8 Conclusions and outlook **271**

A Statistical methods: the profile likelihood **279**
A.1 Treatment of Systematic Uncertainties . . . 280
A.2 Statistical uncertainties on the background sample size . . . 282
A.3 Profile-likelihood unfolding . . . 283
A.4 Likelihood minimization . . . 285
A.5 Hypothesis testing . . . 286
A.5.1 Signal Discovery . . . 287
A.5.2 Setting Upper Limits . . . 288

B The VBF-like gluon-gluon fusion contribution in the $H \to \tau\tau$ STXS measurement **289**

C Searches for lepton-flavor-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ final states using the Symmetry method **291**
C.1 Selection in the Symmetry-based $\ell\tau_{\ell'}$ channel . . . 291
C.2 Background estimation in the Symmetry $\ell\tau_{\ell'}$ channel . . . 292
C.3 Multivariate techniques in the Symmetry-based $\ell\tau_{\ell'}$ channel . . . 295
C.4 Results in the Symmetry-based $\ell\tau_{\ell'}$ channel . . . 297
C.4.1 Individual measurements of $H \to e\tau$ and $H \to \mu\tau$ . . . 297
C.4.2 Branching ratio difference in the $\ell\tau_{\ell'}$ channel . . . 300
C.5 Summary . . . 302

Resumen **302**

References **329**

Agradecimientos **357**

Preface

The Standard Model (SM) of particle physics provides the framework for understanding fundamental particles and forces, excluding gravity. This relativistic quantum field theory describes the behavior and interactions of fermions, which constitute matter, and gauge bosons, which mediate the electromagnetic, weak, and strong interactions. The Higgs boson, a scalar particle predicted by the theory, plays a central role in the SM by confirming the existence of the Higgs field. This field is responsible for giving mass to elementary particles through the Higgs mechanism, which involves the spontaneous breaking of electroweak symmetry. The discovery of the Higgs boson in 2012 by the ATLAS and CMS collaborations at CERN's Large Hadron Collider (LHC) completed the SM.

Within the Standard Model, the Higgs boson plays a unique role by coupling to fermions proportionally to their masses and to gauge bosons with a strength proportional to the square of their masses, enabling measurements of its properties to test the model's predictions. The discovery of a Higgs boson with a mass around 125 GeV not only validated the SM but also allowed for precise measurements of its production and decay channels, a window to test theoretical predictions and probe possible deviations. Despite this success, the SM faces limitations. These limitations drive the search for new physics beyond the SM (BSM), with the Higgs boson serving as a fundamental tool, due to its possible connections to BSM phenomena.

Two primary avenues for exploring the Higgs sector are examined in this thesis. The first is the measurement of differential cross-sections of the $H \to \tau^+\tau^-$ decay channel. This decay mode provides direct access to the Higgs boson's Yukawa coupling to leptons, the interaction between the Higgs field and leptons that gives them mass, offering sensitivity to both the SM expectations and possible BSM modifications. Furthermore, it provides a unique sensitivity to the vector boson fusion production of the Higgs boson, a process where two vector bosons (W or Z) emitted by quarks from the colliding protons fuse to produce a Higgs boson.

The analysis of differential cross-sections within this decay channel employs the Simplified Template Cross-Section (STXS) framework, which allows for the measurement of Higgs boson production in specific phase-space regions defined by the kinematics of the final-state particles. The STXS framework, combined with the increased luminosity of LHC's Run 2 dataset at $\sqrt{s} = 13$ TeV, provides enhanced precision and extends the sensitivity of Higgs boson measurements beyond

inclusive approaches. The analysis also performs, for the first time in the ATLAS experiment, a fiducial differential cross-section measurement in the $H \to \tau^+\tau^-$ decay channel. The results of this measurement are further interpreted in the context of an effective field theory.

The other approach explored in this thesis is the search for lepton-flavor-violating (LFV) decays of the Higgs boson into the $e\tau$ and $\mu\tau$ final states. While the SM conserves lepton flavor in the charged lepton sector, observations of neutrino oscillations indicate that lepton flavor is not conserved in nature, raising questions about whether LFV may also manifest in Higgs decays. Detecting such decays would constitute direct evidence of BSM physics and provide insight into potential LFV mechanisms, which could be explained by extensions to the SM. Current experimental limits on LFV decays have placed constraints on their branching ratios, though the small decay width of the Higgs boson leaves room for BSM contributions at levels detectable by the LHC.

This thesis contributes to three primary areas of research within the ATLAS experiment: the development of simulation software for the Tile Calorimeter readout electronics in preparation for the High Luminosity LHC (HL-LHC) upgrades; the measurement of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel, with an emphasis on STXS measurements; and searches for LFV decays into $e\tau$ and $\mu\tau$ final states. These analyses use the complete Run 2 dataset, 140 fb^{-1} of pp collision data collected at a center-of-mass energy of 13 TeV.

The structure of this thesis is as follows: Chapter 1 provides an overview of the Standard Model of particle physics, focusing on the Higgs sector. It covers the theoretical background of the Higgs mechanism, but also the experimental status of the Higgs sector measurements. Chapter 2 introduces the ATLAS detector and the Large Hadron Collider, highlighting the experimental setup and providing an extended description of the ATLAS Tile Calorimeter. Chapter 3 presents the upgrade of the Tile Calorimeter readout electronics for the High Luminosity LHC, focusing on the development of a simulation software that models the output of the electronics with realistic conditions. The simulation of physics processes in pp collisions and the detector response is discussed in Chapter 4. Chapter 5 describes the reconstruction of the physics objects used in the analyses, such as electrons, muons, τ-leptons, jets, and missing transverse momentum. The measurement of the Higgs boson differential cross-sections in the $H \to \tau^+\tau^-$ decay channel is explained in Chapter 6, focusing on the STXS measurement. Chapter 7 presents the search for lepton-flavor-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ final states. These two analyses are based on the same dataset, consisting of 140 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV collected by the ATLAS experiment. Finally, Chapter 8 summarizes the main results and conclusions of the thesis.

Personal contributions

The work presented in this thesis is the result of the author's research activities within the ATLAS collaboration, a large international collaboration with more than 3000 members. The publications and results presented in this thesis are the outcome of a collective effort, and the author's contributions are detailed below, together with the corresponding publications. The level of detail about the different topics covered in the thesis is adapted to reflect the author's contributions to each of them.

- ATLAS Collaboration. *Differential cross-section measurements of Higgs boson production in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.*
  Submitted to JHEP. e-Print: 2407.16320.

  The author contributed to the development and validation of the statistical analysis for the STXS measurement, extracting the measurement results. The author also contributed to the preparation of the internal documentation and the presentation of the results for the publication.

- ATLAS Collaboration. *Searches for lepton-flavour-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector.*
  JHEP 07 (2023), 166. e-Print: 2302.05225. DOI: 10.1007/JHEP07(2023)166.

  The author contributed to the so-called MC-based analysis, specifically to the $\ell\tau_{\ell'}$ channel. The author contributed to the development and validation of the software tools used to develop the analysis strategy, and to the preparation of the inputs for the statistical analysis. The author made a significant contribution to the event selection, the background estimation and the systematic uncertainties evaluation. The development of the multivariate techniques used in the $\ell\tau_{\ell'}$ channel was also a significant contribution of the author. His work included the development and validation of the statistical analysis, and the extraction of the final results, together with the compatibility studies between the different analyses. Furthermore, he contributed to the preparation of the internal documentation and the presentation of the results for the publication.

The author also contributed to the following ATLAS public note:

- ATLAS Collaboration. *Searches of lepton-flavour-violating decays of the Higgs boson with the ATLAS detector at the HL-LHC.* ATL-PHYS-PUB-2022-054.

The author contributed to the preparation of the inputs for the statistical analysis in the original analysis, on which this extrapolation is based. The author is one of the main analyzers of the original analysis.

The author of this thesis has submitted the following conference proceedings accompanying contributions to conferences on behalf of the ATLAS collaboration:

- Antonio Jesús Gómez Delegido. *Higgs boson coupling measurements in ATLAS*, PoS LHCP2024 (2025), 131. DOI: 10.22323/1.478.0131.
- Antonio Jesús Gómez Delegido. *Searches for lepton-flavour-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, JINST 18 (2023) 10, C10001. DOI: 10.1088/1748-0221/18/10/C10001.
- Antonio Jesús Gómez Delegido. *Upgrade of ATLAS Hadronic Tile Calorimeter for the High Luminosity LHC*, Nucl. Instrum. Meth. A 1047 (2023), 167860. DOI: 10.1016/j.nima.2022.167860.
- Antonio Jesús Gómez Delegido. *Performance and calibration of the ATLAS Tile Calorimeter*, PoS EPS-HEP2021 (2022), 748. DOI: 10.22323/1.398.0748.

Beyond the contributions to these publications, the author has also contributed to the operation, calibration and the HL-LHC upgrade of the ATLAS Tile Calorimeter. His contribution to the software development for the simulation of the output of the Tile Calorimeter upgraded electronics is detailed in this thesis. The author has served as Run Coordinator of the Tile Calorimeter during the first months of Run 3, coordinating the daily operations of the detector. He has served as Tile Calorimeter Calibrations Coordinator, supervising the activities of the Tile Calorimeter calibration group. Finally, he has served as Tile Calorimeter Data Quality Leader, monitoring the quality of the data collected by the Tile Calorimeter, and also as expert on-call for the Tile Calorimeter data acquisition system, ensuring the proper functioning of the detector during data taking.

Institutional acknowledgments

The author would like to acknowledge the support received through the FPU 2019 grant, funded by MCIN/AEI/10.13039/501100011033, which made this research possible.

Chapter 1

Introduction to Higgs boson physics

1.1 A brief summary of the Standard Model

In the early 1960s, the fundamental interactions at microscopic scales were described in terms of three forces, each exhibiting distinct strengths and ranges when measured at a typical scale of 1 GeV. The electromagnetic force, governing the most common macroscopic phenomena, had an infinite range and a strength quantified by the fine-structure constant, approximately $\alpha \approx 1/123$. The strong force, responsible for binding neutrons and protons within atomic nuclei, operated at a range of 10^{-15} m and had a strength 100 times greater than the electromagnetic interaction. Lastly, the weak force, governing radioactive decays, had a range of 10^{-17} m and a strength two orders of magnitude weaker than the electromagnetic interaction. The Standard Model (SM) achieved a unified description of the three fundamental forces in the language of a quantum field theory.

The SM is a renormalizable local gauge theory based on the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. The fermionic matter content of the theory is given by the leptons and quarks, which are spin-1/2 point-like particles. Leptons participate into the electromagnetic and weak interactions, while quarks also interact through the strong force. The spin-1 bosons of the theory, mediators of these fundamental interactions, arise from imposing the invariance of the SM under local gauge transformations of the symmetry group. The fermions are organized into three generations that are identical in every attribute except their masses. The origin of the generational structure is not understood. Consequently, there are three charged leptons with an electric charge of -1: the electron (e), the muon (μ), and the tau (τ). Additionally, there are three electrically neutral leptons associated with the charged leptons (ν_e, ν_μ, and ν_τ). Similarly, there are three quarks with an electric charge of $+2/3$: the up (u), charm (c), and top (t), along with three quarks with a charge of $-1/3$: the down (d), strange (s), and bottom (b) quarks.

Quarks exist as triplets within the $SU(3)_C$ group, with an additional quantum property known as color. The strong force is mediated through the exchange of massless gluons among particles possessing color. The eight vector gluons in the

theory carry their own color charges, resulting in self-interaction. The QCD coupling constant, α_s, remains small for high momentum transfers but increases significantly in softer processes. This behavior ultimately leads to the confinement of quarks within color-neutral hadrons. Gluons are neutral under the electromagnetic and weak interactions.

The $SU(2)_L \otimes U(1)_Y$ symmetry group describes the so-called Electroweak Interaction, mediated by three massive vector bosons, the $W^{\pm}$ and the Z. The $W^{\pm}$ couple to particles with weak isospin, while the Z boson couples to particles with weak isospin and hypercharge. The masses of these bosons determine the typical range of the weak force. The massless photon mediated the electromagnetic interaction and couples to particles with electric charge. Since the photon does not have electric charge, it does not couple to itself, and since it is massless, it has an infinite range.

The introduction of mass terms for the vector bosons and the fermions in the SM lagrangian is not allowed by the local gauge invariance of the theory. The $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken through the existence of a scalar field with non-zero expectation value, the Higgs field. This leads to the emergence of massive vector bosons and a spin-0 scalar particle, the Higgs boson. This particle has no color and no electric charge, and couples to all massive particles of the SM, being responsible for their masses.

1.2 The structure of the SM and the necessity of the Higgs mechanism

Gauge invariance does not allow for the introduction of mass terms for the vector bosons or for chiral fermions in the Lagrangian. Introducing mass terms for the vector bosons in the SM lagrangian breaks the invariance of the theory under local gauge transformations. This problem can be avoided in the case of the photon: since the electromagnetic interaction has infinite range, the photon is massless and there is no necessity to introduce a mass term for it. However, the short-ranged weak force indicates the necessity for massive force carriers. Charged fermions masses have been measured, and furthermore they are required in order to generate flavor mixing and keep the theory renormalizable [1]. This evinces the necessity of a mechanism to generate masses for the vector bosons and the fermions in the SM dynamically, preserving the gauge invariance of the theory.

The frequently called Higgs mechanism provides a solution for this problem. It was introduced independently in 1964 by Peter Higgs [2], François Englert and Robert Brout [3], and Gerald Guralnik, Carl Hagen, and Tom Kibble [4].

The idea of this mechanism is to preserve the gauge invariance in which the SM is based, but to break it spontaneously. The lowest energy state, or vacuum, is not symmetric under gauge transformations, and this will induce masses for the particles that propagate through the vacuum. The following paragraphs are

dedicated to introduce the structure of the SM lagrangian, and to show how the Higgs mechanism is implemented in the SM.

1.2.1 Quantum chromodynamics and the strong force

Quantum chromodynamics (QCD) is a non-Abelian gauge theory based on the $SU(3)_C$ symmetry group. The theory arises from the imposition of local gauge invariance under $SU(3)_C$ transformations in the lagrangian of a free quark field. The QCD part of the SM lagrangian can be written as:

$$\mathcal{L}_{SU(3)_C} = -\frac{1}{4}\sum_{i=1}^{8} G^i_{\mu\nu} G^{i\mu\nu} + \sum_{m=1}^{n_f} \bar{q}_m i\gamma^\mu D_\mu q_m, \tag{1.1}$$

where n_f is the number of quark generations. In this expression, the field strength tensors are defined like:

$$G^i_{\mu\nu} = \partial_\mu G^i_\nu - \partial_\nu G^i_\mu - g_s f^{ijk} G^j_\mu G^k_\nu, \tag{1.2}$$

where G^i_μ are the gluon fields, $g_s = \sqrt{4\pi\alpha_s}$ is the strong coupling strength, and f^{ijk} are the structure constants of the $SU(3)$ group, that appear in the commutation relations followed by the group generators, $[T^i, T^j] = if^{ijk}T^k$. Since the generators of the $SU(3)$ group, that are the Gell-Mann matrices, do not commute, the theory is non-Abelian and the gluons are self-interacting. The number of generators determines the number of gauge bosons in the theory. In the case of QCD, eight gluons arise from the eight generators of the $SU(3)_C$ group [5].

The covariant derivative is defined as:

$$D_\mu = \partial_\mu - ig_s \sum_{i}^{8} G^i_\mu T^i, \tag{1.3}$$

and the quark fields are expressed as

$$q_m = \Big(u_m,\ d_m\Big)^{\mathrm{T}}, \tag{1.4}$$

where u_m and d_m are the up-type and down-type quarks, respectively. The index m runs over the three generations of quarks. No mass terms have been included in this lagrangian, even if they would be allowed by a QCD standalone theory. However, as we will see shortly after this lines, the chiral symmetry of the electroweak part of the SM lagrangian will forbid the inclusion of these terms. Ghost and gauge fixing terms have been omitted for simplicity.

An additional term that violates the symmetry of the lagrangian under charge-parity (CP) transformations can be included in the QCD lagrangian without spoiling the gauge invariance:

$$\mathcal{L}_{\mathrm{CP}} = \frac{\theta_{\mathrm{QCD}} g_s^2}{32\pi^2} \epsilon^{\mu\nu\rho\sigma} G_{\mu\nu,a} G^a_{\rho\sigma}, \tag{1.5}$$

where θ_{QCD} is a free parameter of the theory and $\epsilon^{\mu\nu\rho\sigma}$ is a totally antisymmetric tensor. This term would induce a non-zero electric dipole moment for the neutron, but the experimental limits on this quantity are very stringent, constraining θ_{QCD} to be below 10^{-10} [6]. The necessity of fine-tuning this parameter to such a small value is known as the strong CP problem, and has motivated searches for a Goldstone boson associated with the spontaneous breaking of the CP symmetry, the axion [7, 8].

Despite the relatively simple structure of the QCD lagrangian, it induces a very rich phenomenology. It induces not only gluon-quark-antiquark vertices, but also 3-point and 4-point gluon vertices. Two of the most important consequences of the QCD lagrangian are the confinement of quarks and the asymptotic freedom of the strong force. The first one implies that quarks are never observed in isolation, but always in color-neutral hadrons. The interaction between to colored objects increases with the distance. When two colored charges are tried to be separated, the energy stored in the field between them increases, and eventually becomes so large that it is energetically favorable to create a quark-antiquark pair from the vacuum [1]. The second one implies that the strong coupling constant decreases with the energy scale, which allows for the use of perturbation theory at high energies. The QCD interactions become weak at high energies, in hard processes and deep inelastic processes [9].

1.2.2 The electroweak interaction

The weak and the electromagnetic interactions were unified in the SM through the Glashow-Weinberg-Salam model [10–12]. The model required the lagrangian to be invariant under the $SU(2)_L \otimes U(1)_Y$ symmetry group. The $SU(2)_L$ group is a non-Abelian group that contains three generators, σ^i, $i = 1, 2, 3$, that are the Pauli matrices. The $U(1)_Y$ group is an Abelian group that contains a single generator. The lagrangian of the electroweak part of the SM can be written as:

$$\mathcal{L}_{SU(2)_L \otimes U(1)_Y} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{fermions}}, \tag{1.6}$$

where the gauge part of the lagrangian is given by:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F^i_{\mu\nu} F^{i\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \tag{1.7}$$

In this expression $F^i_{\mu\nu}$ and $B_{\mu\nu}$ are the field strength tensors for the $SU(2)_L$ and $U(1)_Y$ groups, respectively. These tensors are defined as:

$$F^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon^{ijk} W^j_\mu W^k_\nu, \tag{1.8}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.9}$$

where W^i_μ and B_μ are the gauge fields for the $SU(2)_L$ and $U(1)_Y$ groups, respectively. The index i runs from 1 to 3, and g is the coupling constant for the $SU(2)_L$ group.

The third field of the $SU(2)_L$ group, W^3_μ, will eventually mix with the B_μ field to give the Z_μ and A_μ fields, which correspond to the Z and photon bosons. The other two fields, W^1_μ and W^2_μ, will give the $W^\pm$ bosons.

As a consequence of Noether's theorem [13], the gauge invariance of the lagrangian implies the existence of conserved currents and charges, as we have seen for the QCD lagrangian. The conserved charges of the $SU(2)_L$ and $U(1)_Y$ groups are the weak isospin $\mathbf{I}$ and the hypercharge Y, respectively. The hypercharge and the third component of the weak isospin are related to the electric charge through the Gell-Mann-Nishijima formula, $Q = T_3 + Y/2$ [14, 15].

Three families of quarks and leptons are involved in the SM, and in each family we can separate a left-handed and a right-handed component. The left-handed components are doublets under the $SU(2)_L$ group, with isospin $I = 1/2$, while the right-handed components are singlets under this group, and therefore have isospin $I = 0$. There is no experimental evidence of right-handed neutrinos, so they are not included in the SM. However, the observation of neutrino masses suggests the need for physics beyond the SM, which could involve the existence of right-handed neutrinos, the possibility of Majorana neutrinos or the generation of neutrino masses through radiative corrections [16]. The doublets and singlets in each family are written as:

$$L - \text{doublets} : q_{mL} = \begin{pmatrix} u_{mL} \\ d_{mL} \end{pmatrix}, \; l_{mL} = \begin{pmatrix} \nu_{mL} \\ e_{mL} \end{pmatrix}, \tag{1.10}$$

$$R - \text{singlets} : u_{mR}, \; d_{mR}, \; e_{mR}. \tag{1.11}$$

The SM is a chiral theory because of this separation: the left-handed and right-handed fields transform differently under the gauge group. This chiral structure has crucial consequences for the mass terms of the fermions: they cannot be introduced in the lagrangian in a gauge-invariant way. A term like $m\bar{q}q$ would break the gauge invariance of the lagrangian. This can be explicitly seen introducing the projection operators $P_L = (1 - \gamma^5)/2$ and $P_R = (1 + \gamma^5)/2$, that project the left- and right-handed components of the fermions (i.e. $q_L = P_L q$ and $q_R = P_R q$, where q is a generic fermion field). A bare mass term can be written as:

$$m\bar{q}q = m\bar{q}(P_L + P_R)q = m\bar{q}_L q_R + m\bar{q}_R q_L. \tag{1.12}$$

Performing a gauge transformation, the right and left-handed components will transform differently, and the mass term will not be symmetric under these transformations. Since no fermion masses are allowed, $\mathcal{L}_{\text{fermions}}$ consists of kinetic terms only:

$$\mathcal{L}_{\text{fermions}} = \sum_{m=1}^{n_f} \Big(\bar{q}_{mL} i\gamma^\mu D_\mu q_{mL} + \bar{l}_{mL} i\gamma^\mu D_\mu l_{mL} + \\ + \bar{u}_{mR} i\gamma^\mu D_\mu u_{mR} + \bar{d}_{mR} i\gamma^\mu D_\mu d_{mR} + \bar{e}_{mR} i\gamma^\mu D_\mu e_{mR} \Big), \quad (1.13)$$

where the covariant derivative is defined as:

$$D_\mu = \partial_\mu - ig\frac{\sigma^i}{2} W_\mu^i - ig'\frac{Y}{2} B_\mu. \quad (1.14)$$

In this expression it is possible to see explicitly that the W_μ fields couple with a strength proportional to g, and the B_μ field has a coupling proportional to g' with particles with hypercharge. The g and g' couplings are free parameters of the theory. The physical neutral fields that couple to the fermions the Z and photon bosons, are linear combinations of the W_μ^3 and B_μ fields:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \quad (1.15)$$

where θ_W is the weak mixing angle, defined as $\cos\theta_W = g/\sqrt{g^2 + g'^2}$. The Z boson couples to particles with weak isospin and hypercharge, while the photon couples to particles with electric charge. The $W^\pm$ bosons can be written as a linear combination of the W_μ^1 and W_μ^2 fields as $W^\pm = (W_\mu^1 \mp iW_\mu^2)/\sqrt{2}$, and they couple to particles with weak isospin.

The masses of the gauge bosons are forbidden to avoid spoiling the gauge invariance of the theory and the mass terms for the fermions are not allowed by the chiral structure of the SM. The SM will need a scalar part to solve these problems. The structure of this part of the lagrangian will be introduced in the following paragraphs, while the spontaneous symmetry breaking of the electroweak symmetry will be discussed in the next section. The scalar part of the lagrangian is:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi), \quad (1.16)$$

where $\Phi = \begin{pmatrix} \phi^+, & \phi^0 \end{pmatrix}^{\text{T}}$ is the Higgs field, ϕ^+ and ϕ^0 are the charged and neutral components of the Higgs field, and $V(\Phi)$ is the so-called Higgs potential. Φ is a complex scalar doublet with hypercharge $Y = 1$ and weak isospin $I = 1/2$. The interaction between the Higgs scalar and the gauge bosons is introduced in this term through the covariant derivative. After the spontaneous symmetry breaking this will generate the masses of the gauge bosons. The potential $V(\Phi)$ is restricted to be of the form:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \quad (1.17)$$

where μ^2 and λ are free parameters of the theory. This term introduces the self-interaction of the Higgs field.

Finally, a term can be added introducing the interaction between the Higgs doublet and the fermion fields [11]. This term is called the Yukawa term, and it is written as:

$$\mathcal{L}_{\text{Yukawa}} = -\sum_{m,n=1}^{n_f} \left(\Gamma^u_{mn} \bar{q}_{mL} \tilde{\Phi} u_{nR} + \Gamma^d_{mn} \bar{q}_{mL} \Phi d_{nR} + \Gamma^e_{mn} \bar{l}_{mL} \Phi e_{nR} + \text{h.c.} \right), \quad (1.18)$$

where $\tilde{\Phi} = i\sigma^2 \Phi^*$ is the conjugate of the Higgs field, and Γ^u_{mn}, Γ^d_{mn}, and Γ^e_{mn} are the Yukawa matrices for the up-type quarks, down-type quarks, and charged leptons, respectively. These matrices are free parameters of the theory, and they are not required to be diagonal by any symmetry of the model.

1.3 The spontaneous symmetry breaking

As we have seen, it is not possible to introduce bare mass terms for the gauge bosons and the fermions in the SM lagrangian. We will see now how these mass terms are generated after the spontaneous symmetry breaking of the electroweak symmetry to electromagnetism. This will generate masses for the fermions and the $W^{\pm}$ and Z bosons.

The Higgs potential has been introduced in equation 1.16. If the potential is required to be bounded from below, the parameter λ must be positive. There are two possible cases as a function of the sign of μ^2. If $\mu^2 > 0$, the potential has a minimum at $\Phi = 0$, but this case is not interesting since the symmetry will remain unbroken. The scalar field would have a mass term, but the gauge bosons and the fermions would remain massless. However, if $\mu^2 < 0$, the potential has an infinite number of minima. The minimum is at a non-zero value of the Higgs field, and without loss of generality (since the potential is invariant under global $SU(2)_L$ transformations) it can be written as:

$$\Phi_0 = \left(0,\ v/\sqrt{2}\right)^{\mathrm{T}}, \quad (1.19)$$

where $v = \sqrt{-\mu^2/\lambda}$ is called the vacuum expectation value of the Higgs field. The lagrangian is still invariant under $SU(2)_L \otimes U(1)_Y$, but the vacuum is not: once a ground state is chosen, the original symmetry is broken to the subgroup that leaves the vacuum invariant, the electromagnetic symmetry, $U(1)_Q$. This will prevent the introduction of mass terms for the photon.

To study the mass terms that are generated after the symmetry breaking, it is convenient to write the scalar doublet in a general form as:

$$\Phi = \exp\left(i\frac{\sigma^j \theta^j}{v}\right) \left(0,\ (v+H)/\sqrt{2}\right)^{\mathrm{T}}, \quad (1.20)$$

where θ^j are the Goldstone bosons, and H is the physical Higgs boson. It is possible to make use of the $SU(2)_L$ symmetry to remove any dependence on these Goldstone

bosons: using the so-called unitary gauge, the degrees of freedom of the Goldstone fields can be absorbed by the longitudinal polarization of the three massive gauge bosons. However, the original Higgs doublet has four degrees of freedom, and one of them remains as a physical scalar particle, the Higgs boson. The scalar part of the lagrangian can be written as:

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2}\partial_\mu H \partial^\mu H + (v+H)^2 \left(g^2 W_\mu^+ W^{-\mu} + \frac{g^2}{8\cos^2\theta_W} Z_\mu Z^\mu \right) - V(H), \quad (1.21)$$

where $V(H)$ is the Higgs potential in terms of the physical Higgs boson. We have rotated the B_μ and W_μ^3 fields to the Z_μ and A_μ fields to explicitly show the mass terms for the gauge bosons and the interaction terms between the Higgs field and the gauge fields. Once the electroweak symmetry has been broken, two quadratic terms have arisen introducing in the lagrangian masses for the $W^\pm$ and Z bosons. As expected due to the invariance of the theory after the symmetry breaking to transformations of the $U(1)_Q$ group, there is no bilinear term between the Higgs field and the photon. The masses of the W and Z bosons are given by:

$$m_W = \frac{1}{2} v g, \qquad m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2} = \frac{m_W}{\cos\theta_W}. \quad (1.22)$$

The lagrangian in equation 1.21 also introduces vertices between one Higgs boson and two gauge bosons and two Higgs bosons and two gauge bosons. Further manipulations in the lagrangian reveal that the strength of the interaction of the Higgs boson with the $W^\pm$ and Z bosons is proportional to their masses squared. These interaction terms are given by:

$$\mathcal{L}_{\text{Higgs-W/Z}} = \frac{2m_W^2}{v} H W_\mu^+ W^{-\mu} + \frac{2m_Z^2}{v} H Z_\mu Z^\mu. \quad (1.23)$$

As introduced in the previous section, in order to generate masses for the fermions, an additional gauge invariant term is required in the lagrangian. This term is the Yukawa term, and has been introduced in equation 1.18. Using the unitary gauge, the Yukawa terms can be rewritten as:

$$\mathcal{L}_{\text{Yukawa}} = -\frac{v}{\sqrt{2}}\left(1+\frac{H}{v}\right) \sum_{m=1}^{n_f} \left(\Gamma^u_{mn} \bar{u}_{mL} u_{nR} + \Gamma^d_{mn} \bar{d}_{mL} d_{nR} + \Gamma^e_{mn} \bar{e}_{mL} e_{nR} + \text{h.c.} \right). \quad (1.24)$$

The notation can be simplified introducing vectors in the flavor space, $\mathbf{u}_L = (u_L, c_L, t_L)$, $\mathbf{d}_L = (d_L, s_L, b_L)$, and $\mathbf{e}_L = (e_L, \mu_L, \tau_L)$, and similarly for the right-handed components. From the proportionality terms in the lagrangian we can define the mass matrices for the fermions as $\mathbf{M}_u = (v/\sqrt{2})\Gamma^u$, $\mathbf{M}_d = (v/\sqrt{2})\Gamma^d$, and $\mathbf{M}_e = (v/\sqrt{2})\Gamma^e$ and also the Yukawa matrices as $\mathbf{Y}_u = \mathbf{M}_u/v$, $\mathbf{Y}_d = \mathbf{M}_d/v$, and $\mathbf{Y}_e = \mathbf{M}_e/v$. With this notation, the Yukawa terms can be written as:

$$\mathcal{L}_{\text{Yukawa}} = -\left(1+\frac{H}{v}\right) \left(\bar{\mathbf{u}}_L \mathbf{M}_u \mathbf{u}_R + \bar{\mathbf{d}}_L \mathbf{M}_d \mathbf{d}_R + \bar{\mathbf{e}}_L \mathbf{M}_e \mathbf{e}_R + \text{h.c.} \right). \quad (1.25)$$

These mass matrices (and therefore the Yukawa matrices) are complex 3×3 matrices that are not necessarily diagonal. However, to identify the physical particle content it is necessary to diagonalize these matrices by unitary transformations that rotate the weak eigenstates. This implies that the mass eigenstates are not the same as the weak eigenstates, that have been employed to define the interaction terms, and a linear combination of weak eigenstates will determine the states with a defined mass. Furthermore, the diagonalization of these matrices will also introduce the flavor mixing of quarks in the SM. These complex matrices can be generally decomposed as the product of a hermitian matrix and a unitary matrix, $\mathbf{M}_f = \mathbf{H}_f \mathbf{U}_f$, where $\mathbf{H}_f$ is defined as $\mathbf{H}_f \equiv \mathbf{M}_f \mathbf{M}_f^\dagger$. Since the matrix $\mathbf{H}_f$ is hermitian, it can be diagonalized by a unitary matrix, $\mathbf{A}_f$, such that $\mathbf{A}_f^\dagger \mathbf{H}_f \mathbf{A}_f = \mathcal{M}_f$, where $\mathcal{M}_f$ is a diagonal matrix containing the masses of the fermions [17]. The rotation represented by the unitary transformation is different for the left- and right-handed components:

$$\begin{aligned} &\mathbf{d}'_L \equiv \mathbf{A}_d \mathbf{d}_L, && \mathbf{u}'_L \equiv \mathbf{A}_u \mathbf{u}_L, && \mathbf{l}'_L \equiv \mathbf{A}_l \mathbf{l}_L, \\ &\mathbf{d}'_R \equiv \mathbf{A}_d \mathbf{U}_d \mathbf{d}_R, && \mathbf{u}'_R \equiv \mathbf{A}_u \mathbf{U}_u \mathbf{u}_R, && \mathbf{l}'_R \equiv \mathbf{A}_l \mathbf{U}_l \mathbf{l}_R. \end{aligned} \tag{1.26}$$

With the rotated fields, the Yukawa terms can be written again, showing explicitly that mass terms for the fermions naturally arise in the lagrangian after the electroweak spontaneous symmetry breaking, but also that the strength of the interaction of the Higgs field with the fermions is proportional to their masses, since the Yukawa matrices are also diagonalized through the aforementioned transformations. The mass terms for the fermions are given by the eigenvalues of the mass matrices:

$$\mathcal{L}_{\text{Yukawa}} = -\left(1 + \frac{h}{v}\right)\left(\bar{\mathbf{u}}'_L \mathcal{M}_u \mathbf{u}'_R + \bar{\mathbf{d}}'_L \mathcal{M}_d \mathbf{d}'_R + \bar{\mathbf{e}}'_L \mathcal{M}_e \mathbf{e}'_R + \text{h.c.}\right). \tag{1.27}$$

From equation 1.27 it is possible to see that the strength of the interaction of the Higgs field with the fermion has the form $y_f = \sqrt{2}\, m_f / v$, where m_f is the mass of the fermion. The Yukawa matrices are diagonal in the SM. Hence, the Higgs boson does not couple to a pair of fermions of different flavor.

The diagonal structure of the Higgs sector has crucial effects in the flavor structure of the SM and the mixing of quarks. Starting from the kinetic terms of the fermions described in equation 1.13, it is possible to derive the interactions of the quarks with the $W^\pm$ bosons, which are not diagonal in the mass eigenstates. This can be seen explicitly by writing the charged current interaction terms in the lagrangian:

$$\mathcal{L}_{\text{fermions}} \supset \mathcal{L}_{\text{CC}} = -\frac{g}{2\sqrt{2}}\left(J_\mu^+ W^{-\mu} + J_\mu^- W^{+\mu}\right), \tag{1.28}$$

where the charged currents J_μ^+ and J_μ^- are defined as a function of the flavor eigenstates:

$$J_\mu^+ = \sum_{m=1}^{n_f} \bar{u}_{mL}\gamma_\mu d_{mL} + \bar{\nu}_{mL}\gamma_\mu e_{mL} \qquad J_\mu^- = \sum_{m=1}^{n_f} \bar{d}_{mL}\gamma_\mu u_{mL} + \bar{e}_{mL}\gamma_\mu \nu_{mL}. \quad (1.29)$$

The unitary transformations that transform up and down quarks from the weak to the mass eigenstates are not generally the same. The charged currents are not diagonal in the mass eigenstates, and the rotation matrices that diagonalize the mass matrices do not leave invariant the charged currents. Once the transformations are introduced in the lagrangian, a unitary matrix of the form $\mathbf{V} = \mathbf{A}_u^\dagger \mathbf{A}_d$ will appear in the charged currents. This matrix is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix, and is a 3×3 unitary matrix that contains the information about the flavor mixing of the quarks. The matrix is conventionally parametrized in terms of three mixing angles and a complex phase that allows for CP violation in the quark sector. However, the amount of CP violation observed in the quark sector is not enough to explain the matter-antimatter asymmetry in the Universe.

In the case of massless neutrinos, where the right-handed sterile neutrinos ν_R are not introduce in the model, it is possible to avoid mixing in the lepton sector by redefining the neutrino fields. The charged current interaction term with the leptons can be written as $\bar{l}_L\gamma_\mu\nu_L W^\mu = \bar{l}_L\gamma_\mu A_l^m \nu_{L,m} W^\mu$, where A_l^m is the unitary matrix that diagonalizes the mass matrix of the charged leptons. The mixing in the lepton sector is then absorbed in a re-definition of the neutrino fields: $\nu'_L = A_l^m \nu_{L,m}$. In the case of massive neutrinos, the right-handed sterile neutrinos ν_R can be introduced in the model, and if a Yukawa term is introduced for the neutrino mass generation, lepton mixing can be introduced in the model allowing also for lepton-flavor violating processes. The evidence for neutrino flavor oscillations indicates the existence of non-zero neutrino masses and a mechanism for neutrino mass generation should be included in the model. However, this mechanism is still unproven.

The neutral current interactions in the lagrangian remain diagonal in the mass eigenstates. The neutral current interaction terms can be written as:

$$\mathcal{L}_{\text{fermions}} \supset \mathcal{L}_{\text{NC}} = -\frac{g}{2\cos\theta_W}\left(J_\mu^3 - \sin^2\theta_W J_\mu^Q\right) Z^\mu, \quad (1.30)$$

where the neutral currents J_μ^3 and J_μ^Q are defined as:

$$J_\mu^3 = \sum_{m=1}^{n_f} \bar{q}_{mL}\gamma_\mu T_3 q_{mL} + \bar{l}_{mL}\gamma_\mu T_3 l_{mL}, \qquad J_\mu^Q = \sum_{m=1}^{n_f} \bar{q}_{mL}\gamma_\mu Q q_{mL} + \bar{l}_{mL}\gamma_\mu Q l_{mL}. \quad (1.31)$$

The neutral currents are diagonal in the mass eigenstates of the fermions. The rotation of the weak eigenstates to the mass eigenstates does not affect the interaction terms between the fermions and the Z boson and prevents the appearance of flavor-changing neutral currents in the SM.

1.4 Production and decay of the 125 GeV Higgs boson: phenomenology for the LHC

After the proposal of the electroweak symmetry breaking by the Higgs mechanism, the LEP collider at CERN and the Tevatron collider at Fermilab performed direct searches for the Higgs boson. The LEP collider excluded the existence of a Higgs boson with a mass below 114.4 GeV at 95% confidence level (CL) [18], while the Tevatron excluded the existence of a Higgs boson with a mass between 149 and 182 GeV at 95% confidence level [19]. Indirect constraints on the Higgs boson mass were also derived from LEP measurements, excluding masses above 185 GeV at 95% confidence level [20].

The Higgs boson was finally discovered at the Large Hadron Collider (LHC) at CERN in 2012 by the ATLAS [21] and CMS experiments CMS [22], marking the beginning of a new era in particle physics. The discovery was announced on July 4th, 2012, and the Higgs boson was observed with a mass of about 125 GeV. After decades of theoretical and experimental efforts and using approximately 10 fb^{-1} of data at a centre-of-mass energy of 7 TeV and 8 TeV, the two experiments announced independently the observation of the boson combining the results of the searches in the main decay modes, where the $H \to \gamma\gamma$, $H \to ZZ^* \to 4l$, and $H \to WW^*$ decay channels had the leading sensitivity. The observed and expected significance for the main decay channels, along with the combined significance at the time of the discovery are shown in figure 1.1. More details on the experimental status of the Higgs boson physics can be found in section 1.5.

The mass of the Higgs boson has been measured by the ATLAS and the CMS experiments to be $m_H = 125.09 \pm 0.24$ GeV [23]. The SM predicts a total decay width for the Higgs boson of $\Gamma_H = 4.07$ MeV [24]. Since the total decay width is much smaller than the mass of the Higgs boson, the scattering cross-section can be expressed as the product of the on-shell production cross-section and the branching ratio for the decay of the Higgs boson (this is the so-called narrow-width approximation).

The Higgs boson is produced at the LHC through the interactions between the gluons and quarks in the proton-proton collisions. Due to the large contribution of the gluon parton distribution function (PDF) at the LHC energy scales, the dominant Higgs boson production mechanism is gluon-gluon fusion (ggF). In this process two gluons from the protons interact through a fermion loop that couples to the Higgs boson, even without a direct coupling between the gluons and the Higgs boson. In the loop process the top quark is the dominant contribution, since it is the heaviest fermion in the SM and therefore has the largest Yukawa coupling. At a center-of-mass energy ($\sqrt{s}$) of 13 TeV the ggF process accounts for approximately 87% of the total Higgs boson production cross-section at the LHC. The process is computed at N^3LO accuracy in QCD [25, 26] and NLO accuracy in electroweak corrections [27]

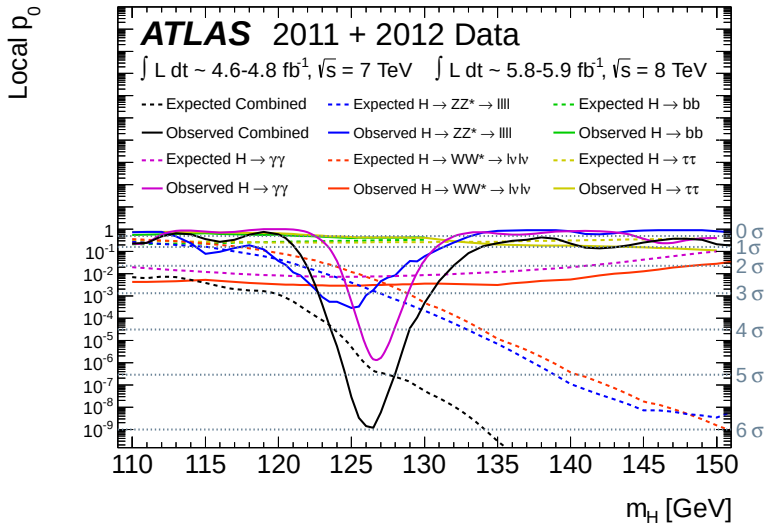


Figure 1.1: Expected and observed local probability p_0 for an experiment where there is no signal to be more signal-like than the observed data. The probability is computed as a function of the Higgs boson mass. Probabilities are shown for the individual channels and their combination. The expected probabilities show in the figure represent the median expected probabilities under the hypothesis of a SM Higgs boson at the given mass. No uncertainties for the expected probabilities are shown [21].

The second most important Higgs boson production mechanism at the LHC is vector boson fusion (VBF). In this process, two quarks from the protons emit a W or Z boson that are fused into a Higgs boson. This process has a characteristic signature of two forward jets in the final state. The VBF process accounts for approximately 7% of the total Higgs boson production cross-section at the LHC with $\sqrt{s} = 13$ TeV and is computed at NNLO accuracy in QCD with NLO accuracy in electroweak corrections[1] [28–30].

[1]These terms refer to the different levels of approximation in perturbative quantum field theory calculations. The "order" corresponds to the power of the coupling constant (e.g., the strong coupling constant, α_s) in the series expansion of a physical quantity, such as a cross-section. Leading Order (LO) is the first and simplest approximation, involving only tree-level Feynman diagrams. Next-to-Leading Order (NLO) includes one-loop corrections, Next-to-Next-to-Leading Order (NNLO) includes two-loop corrections, Next-to-Next-to-Next-to-Leading Order (N^3LO) includes three-loop corrections, and so on. Higher orders yield more accurate predictions by incorporating more complex interactions, but they also require increasingly difficult calculations. Comparing theoretical predictions at different orders with experimental data allows for more precise tests of the underlying theories.

Process	Precision in QCD	Precision in EW	Cross-section ($\sqrt{s} = 13$ TeV) [pb]
ggF	N^{3}LO	NLO	$48.5^{+2.7}_{-3.6}$
VBF	NNLO	NLO	3.78 ± 0.08
qq $\to$ WH	NNLO	NLO	1.37 ± 0.03
qq $\to$ ZH	NNLO	NLO	$0.76^{+0.01}_{-0.02}$
gg $\to$ ZH	NLO	LO	$0.12^{+0.03}_{-0.02}$
ttH	NLO	NLO	$0.51^{+0.03}_{-0.05}$
bbH	NNLO	LO	$0.49^{+0.10}_{-0.12}$
tH	NLO	LO	$0.092^{+0.006}_{-0.011}$

Table 1.1: SM predictions for the Higgs boson production cross-sections at the LHC with a center-of-mass energy of 13 TeV and a Higgs boson mass of 125.09 GeV.

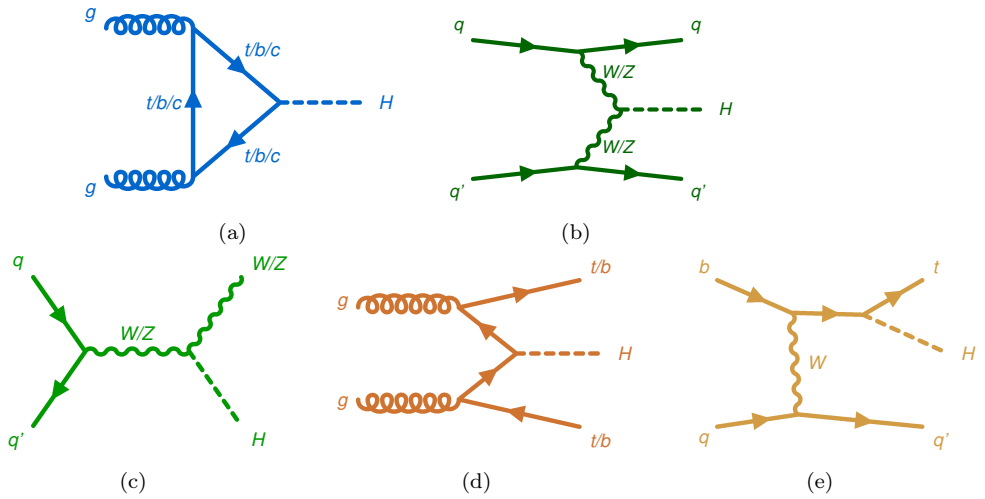


Figure 1.2: Feynman diagrams for the main Higgs boson production mechanisms at the LHC: (a) gluon-gluon fusion, (b) vector boson fusion, (c) associated production with a vector boson, (d) associated production with heavy quark-antiquark pairs, and (e) associated production with a single top quark. Diagrams involving loops are shown in blue, diagrams involving the coupling of the Higgs boson to the vector bosons are shown in green and diagrams where the Higgs boson couples to fermions at tree level are shown in orange [31].

The third most important Higgs boson production mechanism at the LHC is the associated production of a Higgs boson with a vector boson (W or Z), also known as Higgs-strahlung since the vector boson radiates a Higgs boson. The ZH process can be also initiated by the fusion of two gluons. All the associated production processes account for approximately 4% of the total Higgs boson production cross-section at the LHC with $\sqrt{s} = 13$ TeV. The cross-sections are computed a NNLO accuracy in QCD and NLO accuracy in electroweak corrections for the quack initiated processes, and at NLO accuracy in QCD and LO accuracy in electroweak corrections for the gluon initiated processes [32–39].

The next production mechanism at the LHC is the associated production of a Higgs boson with heavy quark-antiquark pairs ($t\bar{t}H$ and $b\bar{b}H$). This occurs through the emission of quark-antiquark pairs from the collision of two gluons. Each of these processes accounts for approximately 1% of the total Higgs boson production cross-section at the LHC at $\sqrt{s} = 13$ TeV. The cross-sections for $t\bar{t}H$ production are computed at NLO accuracy in QCD [40, 41] and electroweak corrections [42, 43], while the cross-sections for $b\bar{b}H$ production are computed at NNLO accuracy in QCD and LO accuracy in electroweak corrections [44–46].

Finally, the Higgs boson can be produced in association with a single top quark. This process accounts for approximately 0.2% of the total Higgs boson production cross-section at the LHC with $\sqrt{s} = 13$ TeV. The cross-sections for these processes are computed at NLO accuracy in QCD and LO accuracy in electroweak corrections [47, 48]. All the cross-sections for the Higgs boson production processes at the LHC are summarized in Table 1.1. These cross-sections are also shown in figure 1.4a as function of the High boson mass for a center-of-mass energy of 13 TeV, which is the center-of-mass energy of the LHC during Run 2. All the Feynman diagrams for the main Higgs boson production mechanisms at the LHC are shown in figure 1.2.

The Higgs boson can decay into both bosons and fermions, due to the coupling to gauge bosons and the Yukawa couplings, respectively. The Feynman diagrams for the main decay modes of the Higgs boson are shown in figure 1.3. The dependence of the branching ratios of the Higgs boson decay modes on the Higgs boson mass is shown in figure 1.4b.

Since the Higgs boson has a mass around 125 GeV, it can not decay to two on-shell vector bosons, at least one has to be off-shell. The branching ratio for the $H \to WW^*$ decay is approximately 22%, while the branching ratio for the $H \to ZZ^*$ decay is roughly 3%. However, the $H \to ZZ^*$ decay has a much cleaner experimental signature than the $H \to WW^*$ decay, since the Z boson can decay leptonically to two electrons or two muons, which can be easily identified and reconstructed in the detector. In the case of the $H \to WW^*$ decay, the W bosons can decay hadronically to two quarks (a signature difficult to distinguish from the large QCD background in a hadron collider like the LHC) or to a lepton and a neutrino. In the later case the neutrino escapes the detector without being detected, imposing challenges in the reconstruction of the final state. The Higgs boson can also decay to two gluons, but this signature is not experimentally accessible at the LHC due to the large QCD background.

The Higgs boson can also decay to two photons through a loop with quarks or W bosons. The branching ratio for the $H \to \gamma\gamma$ decay is approximately 0.2%. The largest background in this case is the photon pair production. The good precision in the energy measurement of the photons make this decay mode a clean experimental signature. The $H \to Z\gamma$ decay is also possible through a loop-induced process, with a small branching ratio of approximately 0.15%.

The Higgs boson can also decay to fermions. The dominant decay mode for the Higgs boson is the $H \to b\bar{b}$ decay, with a branching ratio of approximately 58%. The large backgrounds from multi-jet production make this decay mode difficult to measure at the LHC. However, the large branching ratio motivates the study of this decay, specifically due to the good sensitivity to the associated production of a Higgs boson with vector bosons.

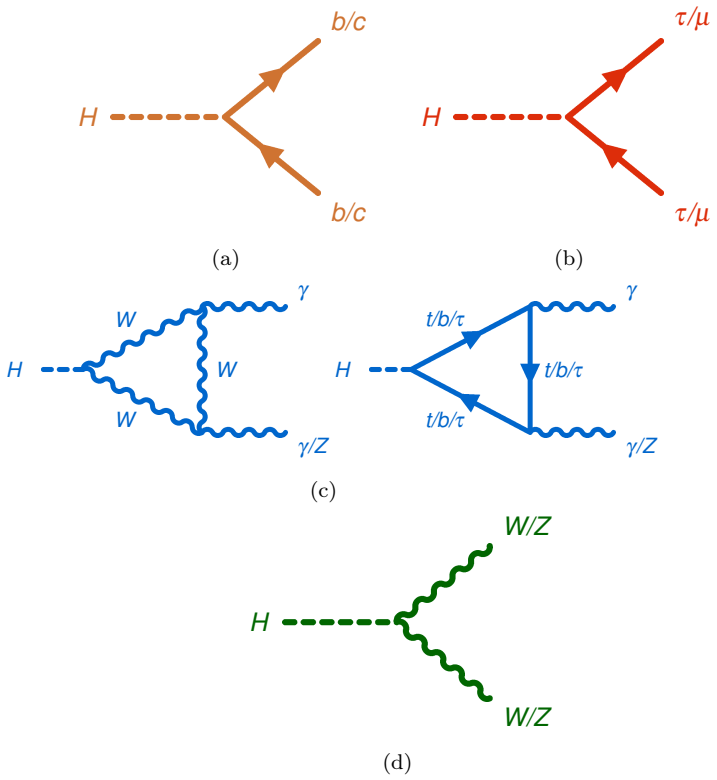
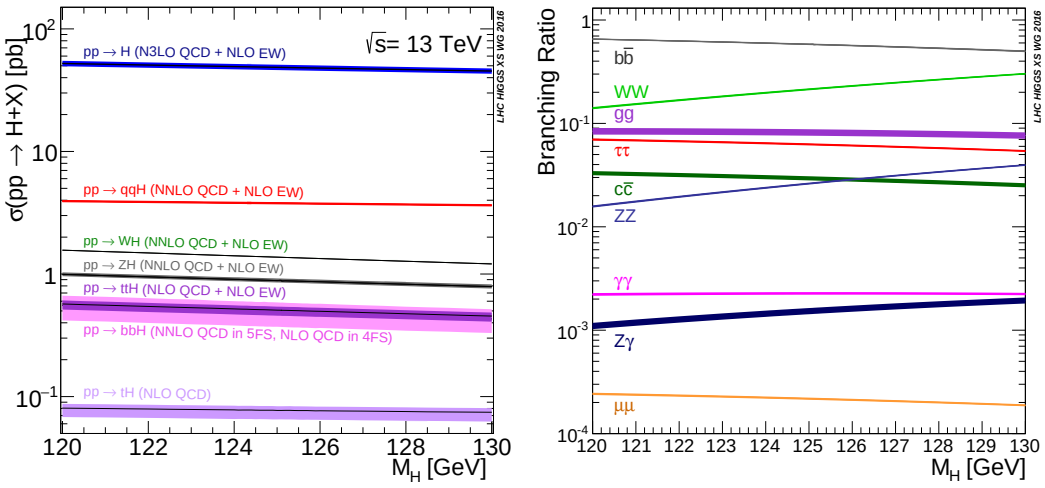


Figure 1.3: Main decay modes of a High boson of 125 GeV: (a) decay to a pair of quarks, (b) decay to a pair of leptons, (c) decay to two photons or to a photon and a Z boson, (d) decay to a pair of vector bosons. The color code is the same as in figure 1.2 [31].

The SM Higgs boson decays to tau leptons with a branching ratio of approximately 6%. This thesis is partially dedicated to study this decay mode. The $H \to \tau\tau$ decay imposes also challenges in the experimental side. Firstly, due to the presence of neutrinos in the tau decays, the reconstruction of the final state is not straightforward. Advanced techniques are required in order to reconstruct the Higgs mass from the di-τ system. Secondly, it suffers from large backgrounds from the $Z \to \tau\tau$ decay, with a cross-section that is orders of magnitude larger than the total Higgs boson production cross-section. The $H \to \tau\tau$ decay offers the possibility to measure the Yukawa coupling of the Higgs boson to the tau lepton and to fermions in general, being the most precise fermionic decay thanks to its large branching ratio. It also allows to study the CP properties of the Higgs boson in its production and decay, and furthermore, it provides a good sensitivity to the vector boson fusion production mode of the Higgs boson.

The Yukawa sector also predicts decays into the first and second generation of fermions. The $H \to \mu\mu$ decay has a branching ratio of approximately 0.02%, and even if it has a clean experimental signature with two muons in the final state, it suffers from a large background from the Drell-Yan process. The $H \to c\bar{c}$ decay has a considerably larger branching ratio of approximately 2.9%, but the large QCD background makes it difficult to measure at the LHC. Furthermore, it requires the identification of charm quarks, a challenging task in a hadron collider. The rest of decays to fermions have extremely small branching ratios, and their direct measurement will not be possible in the near future [24].



(a) (b)

Figure 1.4: (a) Cross-sections for Higgs boson production at 13 TeV as a function of the Higgs boson mass. (b) branching ratios for the Higgs boson decay modes as a function of the Higgs boson mass [24].

1.5 Experimental state-of-the-art of Higgs boson physics

The data recorded by the ATLAS and CMS experiments during Run 1 of the LHC (2010-2012) allowed for the first measurements of the Higgs boson properties. The mass of the Higgs boson was measured to be $m_H = 125.09 \pm 0.24$GeV [49]. The ratio of the measured cross-section to the SM prediction was measured to be $\mu = 1.09 \pm 0.11$ [50]. The ggF production mode was observed by each experiment even using a single decay channel. The combination of the data from the two experiments allowed for the observation of the VBF Higgs boson production modes and the observation of its decay into two photons, two Z bosons, two W bosons, and two tau leptons. [51]. Further measurements with Run 1 data constrained the Higgs width [52, 53] and excluded the possibility of the boson being a spin-2 or a CP-odd spin-0 particle [54, 55] (the possibility of the Higgs boson being a spin-1 particle was already excluded with the observation of the $H \to \gamma\gamma$ decay).

Measurements demonstrated good consistency with the SM predictions within the uncertainties.

The LHC Run 2 (2015-2018) provided the ATLAS and CMS experiments with a larger dataset of 140 fb^{-1} of pp collisions at a $\sqrt{s} = 13$ TeV. This data is crucial to improve the precision of the measurements of the Higgs boson properties, characterize its production and decay modes and search for new physics beyond the SM in the Higgs sector. The following paragraphs synthesize the current status of the Higgs boson physics.

Mass and width measurements

Measuring the Higgs boson mass is fundamental since it is a free parameter in the SM. Furthermore, it is an essential input in order to predict the Higgs boson production cross-sections and branching ratios. Among the different decay modes where the Higgs boson mass can be measured, the $H \to \gamma\gamma$ and $H \to ZZ^*$ are the most precise channels thanks to the good energy resolution and the clean experimental signature, helping to maintain a good control of the backgrounds. The experimental resolution for the Higgs mass measurement in ATLAS is around 1.7 GeV in the $H \to \gamma\gamma$ channel and between 1.5 and 2.1 GeV in the $H \to ZZ^*$ channel (depending on the decay mode of the Z bosons). The combination of both channels provides a mass measurement of $m_H = 125.11 \pm 0.11$GeV [56] using the dataset recorded by the ATLAS experiment.

The Higgs width is determined in the SM once the Higgs boson mass is measured. For a mass of 125 GeV, the SM predicts a total decay width of $\Gamma_H = 4.07$ MeV [24]. The experimental resolution quoted above makes impossible to measure the Higgs width directly. However, the width can be constrained indirectly by measuring the off-shell Higgs boson production [57]. The possibility to constrain this parameter comes from the fact that the Higgs boson on-shell production cross-section is inversely proportional to the Higgs width, while there is negligible dependence of the off-shell production cross-section on the Higgs width. Hence, the ratio of the on-shell to off-shell production cross-sections can be used to indirectly measure the Higgs width. ATLAS reported a best-fit value of $\Gamma_H = 4.5^{+3.3}_{-2.0}$ MeV [58], while CMS measured a best-fit value of $\Gamma_H = 3.2^{+2.4}_{-1.7}$ MeV [59]. Both results are in agreement with the SM prediction.

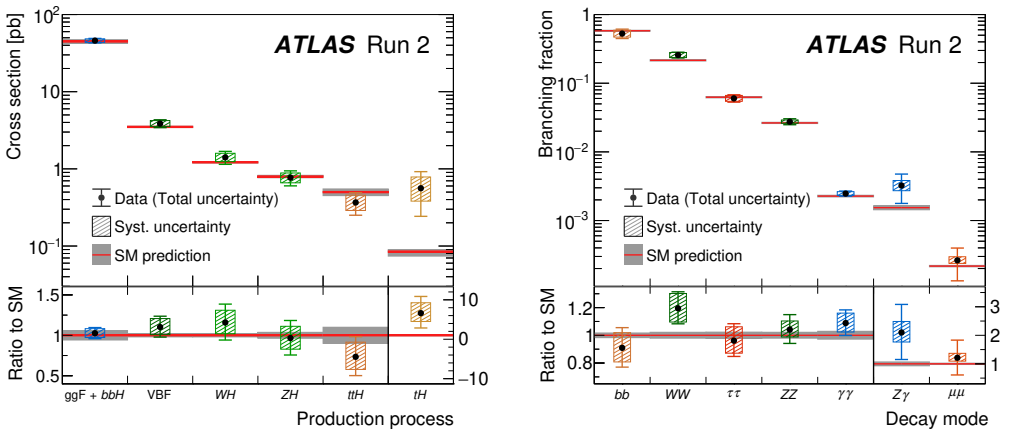
Production cross-section and decay branching ratios

Measurements targeting a Higgs boson signal are usually measuring a signal strength modifier, μ, defined as the ratio of the measured cross-section times branching ratio to the SM prediction (i.e. the total signal yield in the data divided by the total signal yield predicted by the SM). This signal strength can be written as:

$$\mu = \frac{\sigma \times \mathcal{B}}{\sigma_{\text{SM}} \times \mathcal{B}_{\text{SM}}}, \tag{1.32}$$

where $\sigma \times \mathcal{B}$ is the measured Higgs boson production times branching ratio, and the subscripts SM denote the SM prediction. These measurements try to maximize the sensitivity to the Higgs boson signal, but due to its inclusive nature, they are not very sensitive to modifications of the SM predictions phase space areas where few signal events are expected. Furthermore, these inclusive measurements depend on the theory prediction since the uncertainty on the global signal strength will depend on the uncertainty associated to the SM prediction. Additionally, analysis design and selection criteria are established assuming SM kinematics for the expected signal. The latest combination of ATLAS data using different analysis targeting several Higgs boson decay signatures measured the inclusive signal strength to be $\mu = 1.05 \pm 0.06$ [31], well in agreement with the SM prediction. A similar value was reported by the CMS collaboration [60].

If SM values are assumed for the different Higgs boson decay branching ratios, it is possible to perform cross-section measurements combining different analysis by measuring the observed signal yields in the data over the expected signal yields in the SM for each of the Higgs boson production modes. Similarly, it is possible to assume the SM values for the Higgs boson production cross-sections and measure the branching ratios in the same dataset. During the Run 2 of the LHC all the main Higgs boson production and decay modes have been observed. The results of production cross-section and decay branching ratio measurements with the latest combination of different ATLAS analysis can be seen in figure 1.5. All the measurements are well in agreement with the SM predictions. There is evidence for the Higgs boson decay into $Z\gamma$ [61] and $\mu\mu$ [62].



(a) (b)

Figure 1.5: (a) Cross-sections for the different Higgs boson production modes obtained assuming SM values for the Higgs boson branching ratios. (b) Branching ratios for the Higgs boson decay modes assuming SM predictions for production cross-sections. In both figures, the results are obtained combining different ATLAS analysis targeting different Higgs boson decay signatures [31]

To measure the Higgs boson properties with minimal dependency on the theory predictions, it is possible to perform fiducial cross-section measurements.

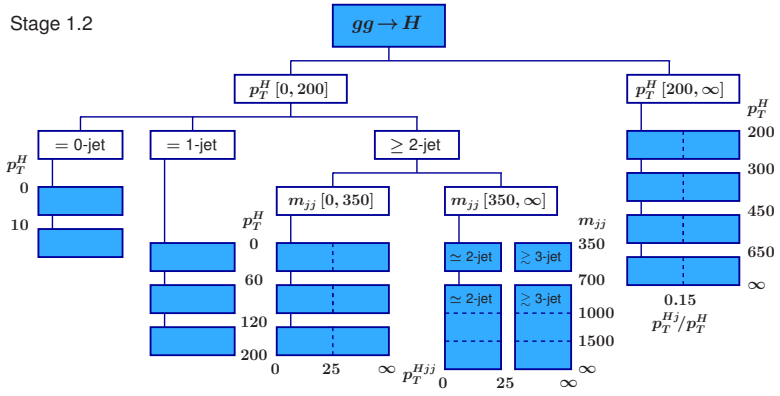
In these measurements a fiducial phase space is defined at particle level[2] keeping it as close as possible to the definition of the reconstructed events to reduce the extrapolation from the measured phase space. The detector effects can be corrected using simulation and the fiducial cross-section can be compared directly to the theory predictions. However, the necessity of keeping similar selections at particle and reconstructed level forces to use simple selections, not necessarily optimal for the signal-to-background ratio. For example, the usage of multivariate techniques to define the selections at reconstructed level is limited in this kind of measurements and the process can be very challenging if the selections imply variables with poor experimental resolution. The technique also allows performing fiducial differential cross-section measurements, where the fiducial cross-section is measured as a function of a kinematic variable.

A compromise between the inclusive and fiducial differential cross-section measurements has been found through the Simplified Template Cross-Section (STXS) framework [63]. This framework divides the phase space of the Higgs boson production into several regions, each of them defined by a set of kinematic variables related to the Higgs boson or to associated objects in the final state. STXS measurements provide differential information, and the definition of the regions is chosen to enhance the sensitivity to Beyond the Standard Model (BSM) effects in the Higgs boson production, while keeping a reasonable independence and control of theory uncertainties. Measurements performed with the STXS framework are not as theory independent as the fiducial cross-section measurements, but they allow for the usage of complex multivariate techniques in the analysis selections. This facilitates the implementation of differential cross-section measurements in decay channels like $H \to \tau\tau$ or $H \to b\bar{b}$, where the reconstruction of the final state is challenging, and the experimental resolution is not as good as in the $H \to \gamma\gamma$ or $H \to ZZ^*$ channels.

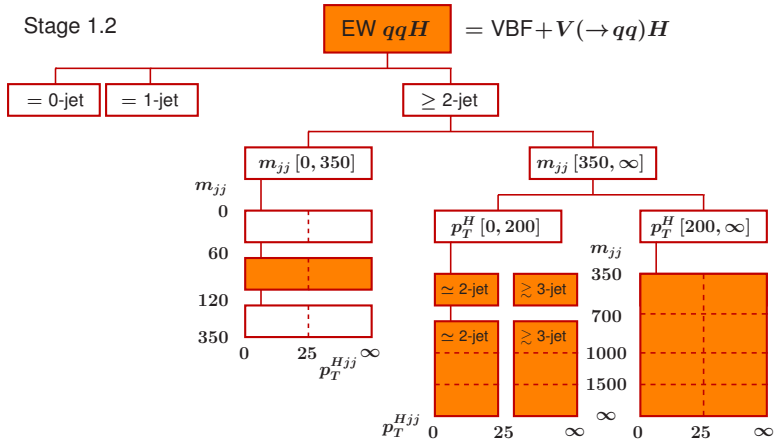
The common definition of the STXS regions allow for the combination of the results from different analyses targeting different Higgs boson decay signatures, which is a fundamental point to maximize the experimental sensitivity.

The current definition of the STXS regions (usually called bins) is shown in figure 1.6 and follows the Stage 1.2, increasing the granularity of the STXS regions with respect to the Stage 1.1 definition [64]. The simplified fiducial volume definition, common for all the analyses that use the STXS framework, requires the Higgs boson to be produced with a rapidity of $|y| < 2.5$ to match the detector acceptance, where the rapidity is defined as $y = 0.5 \ln\left(\frac{E+p_z}{E-p_z}\right)$, with E and p_z being the energy and the longitudinal momentum of the particle, respectively. The details on the ATLAS coordinate system will be presented in section 2.2.1. There are also minimal requirements to define at particle level the associated jets that are used for the definition of the STXS regions [63].

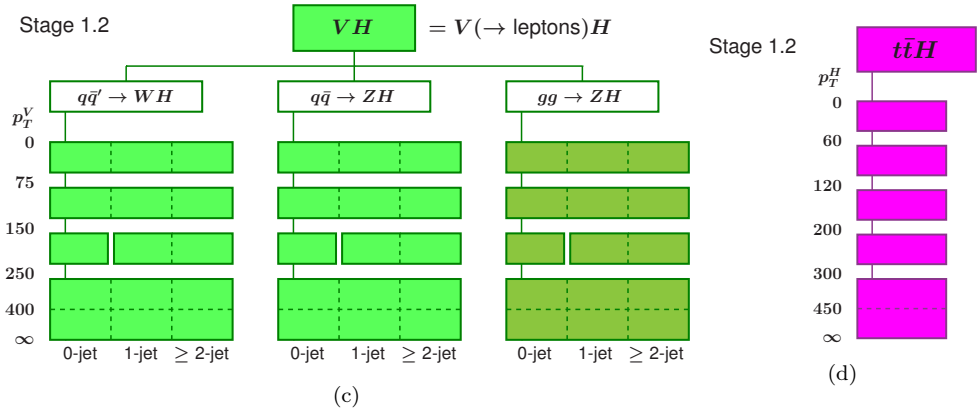
[2]The particle level refers to the level where the objects are defined using stable particles in the final state, without any detector effects.



(a)



(b)



(c) (d)

Figure 1.6: STXS stage 1.2 bin definition for (a) gluon-gluon fusion production, (b) vector boson production and associated production with a hadronically decaying vector boson, (c) associated production with a vector boson with a leptonic decay and (d) associated production with a top quark-antiquark pair [64].

The STXS framework classifies the Higgs boson production in different categories, depending on the production mode and the decay of products of the vector bosons produced in association with the Higgs boson (if any). The first category, called $gg \to H$, includes the Higgs boson production through gluon-gluon fusion and the gluon-induced associated production of a Higgs boson with a Z boson that further decays to hadrons, $gg \to ZH \to q\bar{q}H$. The $b\bar{b}H$ production is also included in this category. The second category, called $qq' \to qq'H$, includes the Higgs boson production through vector boson fusion and the quark-initiated associated production of a Higgs boson with a vector boson where the vector boson decays to hadrons, $qq' \to VH \to qq'H$. The third category, called $VH \to (\ell\ell, \ell\nu)\, H$, includes the vector boson associated production of a Higgs boson where the vector boson decays to leptons. The ttH category includes the production of a Higgs boson with a top quark-antiquark pair. An independent bin is defined for the tH production. Inside each category, the phase space is partitioned into several regions using different variables: the transverse momentum (component of the momentum perpendicular to the beam line) of the Higgs boson (p_{T}^{H}), the transverse momentum of the vector boson (p_{T}^{V}), the number of jets and the invariant mass of the leading and subleading jets in p_{T}.

The binning of the STXS regions is not frozen and can be adapted to the experimental needs. In cases where the experimental sensitivity is not enough to measure the cross-section in a given region, the bin can be merged with a neighbor region. The bins can also be split in the future once the enough data is collected to provide reasonable sensitivity to finer regions.

The last combination of the ATLAS data using the STXS framework [31] provided a measurement of the Higgs boson production cross-section in 36 exclusive kinematic regions. The results of this measurement are shown in figure 1.7. Part of this thesis is devoted to extend the STXS measurements in the $H \to \tau\tau$ decay channel, specifically sensitive to the vector boson fusion production mode. Details on the analysis strategy and the results are presented in Chapter 6. The results quoted in figure 1.7 do not contain the results from the $H \to \tau\tau$ decay channel presented in this thesis.

Couplings of the Higgs boson with other particles

Determining the couplings of the Higgs boson to other particles in crucial as a test of the gauge and Yukawa sectors of the SM. In order to extract from the data the couplings of the Higgs boson to other particles, it is necessary to take into account how modification on the couplings affect at the same time the production cross-section and the decay branching ratios. To parametrize how the cross-section times branching ratio of the Higgs boson in a given production mode and decay channel is modified by a change in the coupling of the Higgs boson to a given particle, coupling strength modifiers are defined in the so-called κ framework [65]. The framework is based on the leading order contributions to the Higgs boson

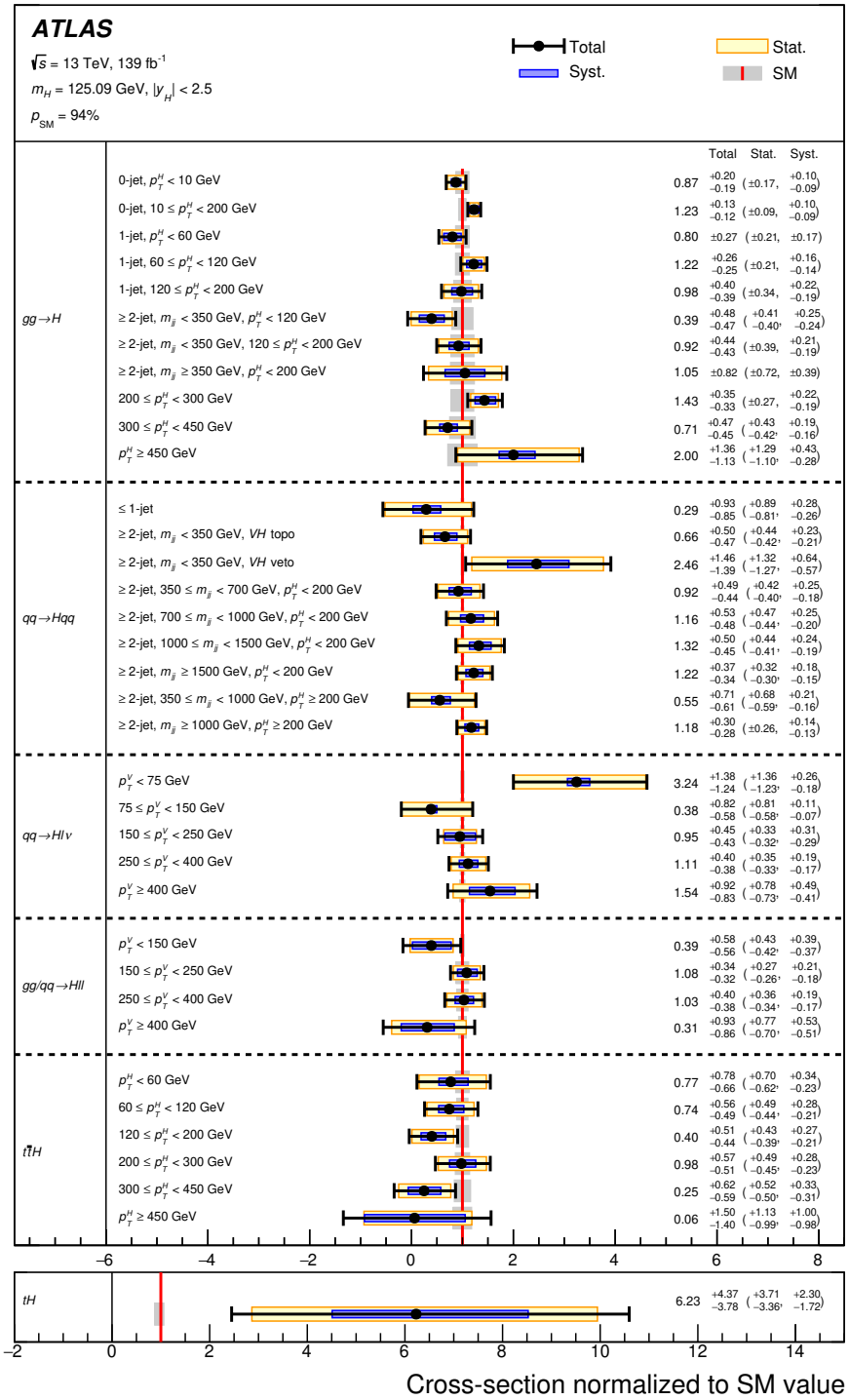


Figure 1.7: Cross-section results normalized to the SM predictions in each of the STXS kinematic regions using the latest combination of data from the ATLAS experiment. All the reported values assume the SM branching ratios for the Higgs boson decays. The error bars represent the total uncertainty in the measurement (in black), the statistical uncertainty (in yellow) and the systematic uncertainty (in blue) [31].

production and decay modes and assumes that modifications in the Higgs boson couplings do not affect the kinematics of the Higgs boson production and decay.

The κ framework defines the coupling strength modifiers as:

$$\kappa_l^2 = \frac{\sigma_l}{\sigma_l^{\mathrm{SM}}} \text{ or } \kappa_l^2 = \frac{\Gamma_l}{\Gamma_l^{\mathrm{SM}}}, \tag{1.33}$$

where σ_l and Γ_l are the production cross-section and the decay width of the Higgs boson in a given production mode and decay channel, l. The corresponding SM values are denoted by σ_l^{SM} and Γ_l^{SM}. With these definitions, it is possible to write the cross-section times branching ratio of the Higgs boson in a given production mode and decay channel as:

$$\sigma_m \times \mathcal{B}_n = \frac{\sigma_m^{\mathrm{SM}}(\boldsymbol{\kappa}) \times \Gamma_n^{\mathrm{SM}}(\boldsymbol{\kappa})}{\Gamma_H(\boldsymbol{\kappa}, \mathcal{B}_{\mathrm{inv}}, \mathcal{B}_{\mathrm{undet}})}, \tag{1.34}$$

where σ_m and $\mathcal{B}_n$ are the production cross-section and the decay branching ratio of the Higgs boson in a given production mode m and decay channel n. The $\boldsymbol{\kappa}$ vector indicates that the several coupling strength modifiers can modify the cross-sections or branching ratios. The total decay width of the Higgs boson is also modified by the coupling strength modifiers, $\Gamma_H(\boldsymbol{\kappa}, \mathcal{B}_{\mathrm{inv}}, \mathcal{B}_{\mathrm{undet}}) = \kappa_H^2(\boldsymbol{\kappa}, \mathcal{B}_{\mathrm{inv}}, \mathcal{B}_{\mathrm{undet}})\Gamma_H^{\mathrm{SM}}$, where κ_H^2 is defined as:

$$\kappa_H^2 = \frac{\sum_p \mathcal{B}_p^{\mathrm{SM}} \kappa_p^2}{1 - \mathcal{B}_{\mathrm{inv}} - \mathcal{B}_{\mathrm{undet}}}. \tag{1.35}$$

In equation 1.35, the sum runs over all the Higgs boson decay channels, except the invisible and undetected decays. The branching ratio to invisible particles is denoted by $\mathcal{B}_{\mathrm{inv}}$ and takes into account decays into BSM particles that do not interact with the detector and escape from it, while the branching ratio to undetected decay modes is denoted by $\mathcal{B}_{\mathrm{undet}}$.

Figure 1.8 shows the reduced coupling strength modifiers for the Higgs boson couplings to fermions and bosons. These reduced modifiers are defined as $\kappa_f m_f/v$ for fermions and $\sqrt{\kappa_V} m_V/v$ for bosons, where m_f and m_V are the masses of the fermions and bosons, respectively. The reduced coupling strength modifiers allow for testing the proportionality of the Higgs boson couplings to the masses of the particles (linear in the case of the fermions and quadratic in the case of the bosons). The reduced coupling strength modifiers shown in figure 1.8 are obtained from the aforementioned combination of the ATLAS data and are well in agreement with the SM predictions. The values are obtained assuming $\mathcal{B}_{\mathrm{inv}} = \mathcal{B}_{\mathrm{undet}} = 0$.

Higgs boson self-coupling

The Higgs boson self-coupling is experimentally accessible through the measurement of the Higgs boson pair production. It also contributes to the single Higgs boson production cross-sections and decay branching ratios. The various possible decay

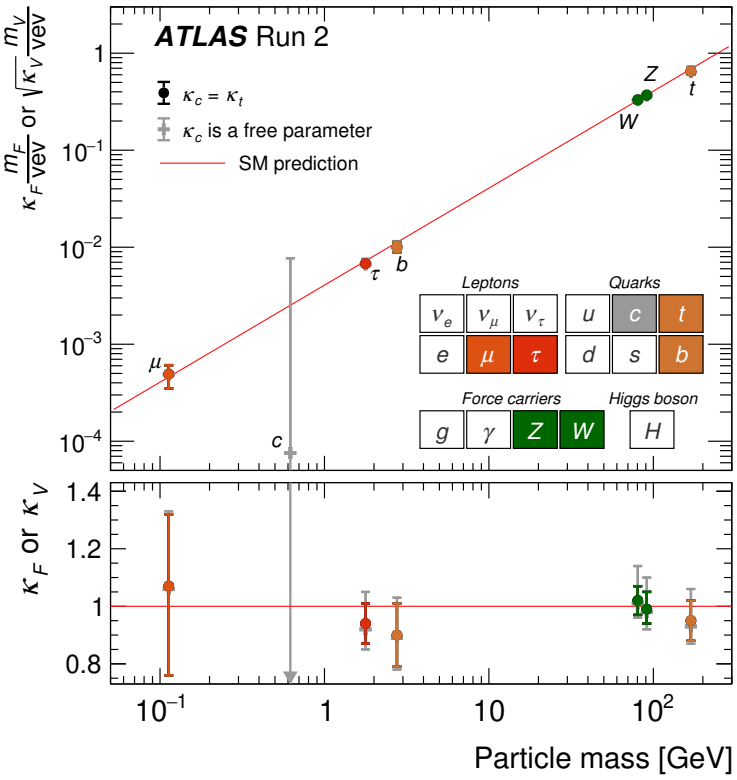


Figure 1.8: Reduced coupling strength modifiers for the Higgs boson couplings to fermions and bosons [31].

modes for the two Higgs bosons allow for the measurement of the Higgs boson self-coupling in different final states. The most sensitive channels are the $HH \to b\bar{b}b\bar{b}$, the $HH \to b\bar{b}\gamma\gamma$ and the $HH \to b\bar{b}\tau\tau$. The ATLAS experiment has set the upper limit on the Higgs boson pair production cross-section at 2.9 times the SM prediction at 95% confidence level [66].

1.6 Lepton flavor violation in the Higgs sector

The conservation of lepton flavor is accidental in the SM, which means that the symmetry is not imposed by the gauge invariance of the theory. The absence of right-handed neutrinos in the SM preserves the lepton flavor conservation in the model, but also implies that neutrinos are massless. However, the observation of neutrino oscillations by the Super-Kamiokande [67] and the Sudbury Neutrino Observatory [68, 69] experiments demonstrated that neutrinos are massive particles.

As shown in Section 1.3, not only the fermion masses arise from the Yukawa terms in the SM Lagrangian, but also the quark flavor mixing. The Yukawa matrices that give rise to the fermion masses (both quarks and leptons) are diagonalized by unitary rotations of the fermion fields in the flavor space. The unitarity of the rotations guarantees that the neutral currents in the SM are flavor diagonal

and no flavor-changing neutral currents are present at tree level. Similarly, it has been shown that the terms introducing vertices between the Higgs boson and the fermions are diagonalized by the same rotations that diagonalize the fermion mass matrices. This implies that the Higgs boson couplings to the fermions are flavor diagonal in the SM and do not introduce charged lepton flavor violation.

On the other hand, diagonalizing at the same time the Yukawa matrices that correspond to the up-type and down-type quarks required $\mathbf{A}_u \neq \mathbf{A}_d$. This misalignment introduces the quark flavor mixing in the SM, controlled by the CKM matrix. The absence of right-handed neutrinos allows the SM to preserve the lepton flavor conservation. The terms in equation 1.18 that involve the lepton doublets are invariant under three unitary rotations of the lepton fields in the flavor space (one for each generation). The invariance of the lagrangian under these rotations implies the conservation of three charges: the lepton family numbers L_e, L_μ and L_τ.

If the SM is extended in order to accommodate massive neutrinos, lepton flavor violation naturally arise. As in the case of the quark sector, the matrices needed to diagonalize the charged leptons and the neutrinos are not the same, and an equivalent of the CKM matrix for the leptons is introduced: the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. Any minimal extension of the SM with massive neutrinos introduces lepton flavor violation in the charged lepton sector through loop diagrams. Some of these processes are the $\mu \to e\gamma$ decay, the $\tau \to \mu\gamma$ decay and the $\tau \to e\gamma$ decay, that even if allowed due to the neutrino oscillations, are extremely suppressed.

Many extensions of the SM introduce further sources of lepton flavor violation in the leptonic sector, including lepton flavor violating decays of the Higgs boson. Some examples are Minimal Supersymmetric SM extensions [70, 71], models with more than one Higgs doublet [72], composite Higgs scenarios [73], Ranadall-Sundrum models [74] or models with wrapped extra dimensions [75]. However, searching for lepton-flavor violation in the Higgs boson decays does not necessarily require to assume a specific model. Searches presented in this thesis are model-independent and just assume that the branching ratio of the lepton flavor violating decay of the Higgs boson is small enough to not modify significantly the Higgs boson width, and thus the couplings of the SM with other SM particles. An extensive presentation of these searches can be found in Chapter 7.

With independence of the model that introduces lepton flavor violation in the Higgs sector, an independent parametrization of an effective lagrangian that includes lepton flavor violation in the Higgs sector can be written as:

$$\mathcal{L}_{\text{LFV}} \supset -\sum_{i,j=1}^{n_f} \left(m_{e_i} \bar{e}_{iL} e_{jR} \delta_{ij} + Y^e_{ij} \bar{e}_{iL} e_{jR} h \right) = -\sum_{i,j=1}^{n_f} \left(m_{e_i} \delta_{ij} + Y^e_{ij} h \right) \bar{e}_{iL} e_{jR}, \tag{1.36}$$

where the lepton couplings do not necessarily need to be diagonal anymore. These terms can be introduced just by higher order operators of the type $\bar{l}_L e_R \Phi(\Phi^\dagger \Phi)$

that can arise after integrating out heavy particles in a theory beyond the SM. The existence of non-diagonal couplings would allow for lepton-flavor violating decays of the Higgs boson of the type $H \to \ell_i \ell_j$, but the same couplings that allow that decay would also contribute at loop level to the $\ell_i \to \ell_j \gamma$ decays. This has implications for the orders of magnitude of the branching ratios of the lepton-flavor violating decays of the Higgs boson. The low energy experimental searches for flavor-violating lepton decays impose constraints on the possible values of the off-diagonal Yukawa matrix elements.

It is possible to employ the current bounds on the branching ratios of $\mu \to e\gamma$, $\tau \to \mu\gamma$ and $\tau \to e\gamma$ to set bounds on the quadratic sum of the off-diagonal Yukawa matrix elements, $\sqrt{\left|Y_{ij}\right|^2 + \left|Y_{ji}\right|^2}$, where i and j run over the three lepton generations. Then, once the bounds on the couplings are obtained, it is possible to translate them on bounds on the branching ratios of the lepton-flavor violating decays of the Higgs boson, $H \to e\mu$, $H \to e\tau$ and $H \to \mu\tau$. This can be done through the following relation:

$$\left|Y_{ij}\right|^2 + \left|Y_{ji}\right|^2 = \frac{8\pi}{m_H} \frac{\mathcal{B}\left(H \to \ell_i \ell_j\right)}{1 - \mathcal{B}\left(H \to \ell_i \ell_j\right)} \Gamma_H^{\mathrm{SM}}, \tag{1.37}$$

where $\Gamma_H^{\mathrm{SM}} = 4.07$ MeV is the total decay width of the Higgs boson in the SM for a mass of 125 GeV. The derivation of equation 1.37 can be found in Ref. [76]. The bounds on the quadratic sum of the off-diagonal Yukawa matrix elements from the low energy experiments, and the corresponding bounds on the branching ratios of the lepton-flavor violating decays of the Higgs boson are shown in Table 1.2.

Process	$\mathcal{B}\left(\ell_i \to \ell_j \gamma\right)$	H LFV Decay	$\mathcal{B}\left(H \to \ell_i \ell_j\right)$
$\mu \to e\gamma$	$< 2.4 \times 10^{-12}$ [77]	$H \to e\mu$	$< 1.6 \times 10^{-8}$ [76]
$\tau \to e\gamma$	$< 3.3 \times 10^{-8}$ [78]	$H \to e\tau$	< 0.19 [76]
$\tau \to \mu\gamma$	$< 4.4 \times 10^{-8}$ [78]	$H \to \mu\tau$	< 0.24 [76]

Table 1.2: Bounds on the branching ratios of the lepton-flavor violating decays of the Higgs boson from the low energy experiments. The bounds on the branching ratios of the lepton-flavor violating decays of the Higgs boson are obtained assuming that the off-diagonal Yukawa couplings are the only source of flavor violation. The bound on the $\mu \to e\gamma$ decay that is reported in [77] is not the latest and most stringent one, but it is used in this table since it was used in [76] to set the bounds on the $H \to e\mu$ decay.

The bound on the $H \to e\mu$ decay is the most stringent one, dominated by the strong constraints induced by the $\mu \to e\gamma$ limit. The process is too suppressed to expect any possible sensitivity to deviations from the SM predictions in the LHC. However, indirect bounds are also model dependent, and direct searches for the $H \to e\mu$ decay are performed in the LHC. The constraints induced by the $\tau \to e\gamma$ and $\tau \to \mu\gamma$ decays are less stringent, leaving the possibility to observe branching ratios of even 20% for the $H \to e\tau$ and $H \to \mu\tau$ decays. The sensitivity of the

LHC to these branching fractions is already below these indirect bounds. This fact reinforces the importance of the direct searches for these processes at the LHC.

Indirect measurements impose also constrains in the product of $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ due to the fact that the same couplings that induce the $H \to e\tau$ and $H \to \mu\tau$ decays at tree level also contribute to the $\mu \to e\gamma$ and $\mu \to e$ conversion in nuclei at the 1-loop level. These branching ratios are related by the following relation:

$$\mathcal{B}(H \to e\tau) \times \mathcal{B}(H \to \mu\tau) = \\ = 7.95 \times 10^{-10} \left[\frac{\mathcal{B}(\mu \to e\gamma)}{10^{-13}}\right] + 3.15 \times 10^{-4} \left[\frac{\mathcal{B}(\mu \to e)_{\text{Au}}}{10^{-12}}\right], \quad (1.38)$$

which assumes that the Higgs boson width is the one predicted by the SM [79]. Using the current best experimental limit on the $\mathcal{B}(\mu \to e)_{\text{Au}} < 7 \times 10^{-13}$ [80] and the $\mathcal{B}(\mu \to e\gamma) < 5.7 \times 10^{-13}$ [81], it is possible to set a limit on the product of the branching ratios of the $H \to e\tau$ and $H \to \mu\tau$ decays of the order of 10^{-4}.

The ATLAS and CMS experiments have performed direct searches for lepton-flavor violating decays of the Higgs boson to $e\mu$, $e\tau$ and $\mu\tau$ using the Run 1 and Run 2 datasets, since it is the best way to probe the off-diagonal Yukawa couplings. No significant excess of events has been observed in any of the searches, and the results have been used to set upper limits on the branching ratios of the lepton-flavor violating processes. The ATLAS experiment results with the full Run 2 dataset will be presented in chapter 7. The CMS results and the ATLAS result that are not included in this thesis are summarized in Table 1.3.

Process	Limit on $\mathcal{B}$	Luminosity	Limit on $\mathcal{B}$	Luminosity
	ATLAS		CMS	
$H \to e\mu$	6.1×10^{-5} [82]	139 fb^{-1}	4.4×10^{-5}	138 fb^{-1} [83]
$H \to e\tau$	4.7×10^{-3} [84]	36.1 fb^{-1}	2.2×10^{-3}	137 fb^{-1} [85]
$H \to \mu\tau$	2.8×10^{-3} [84]	36.1 fb^{-1}	1.5×10^{-3}	137 fb^{-1} [85]

Table 1.3: Limits on the branching ratios of the lepton-flavor violating decays of the Higgs boson from the ATLAS and CMS experiments.

1.7 The Higgs sector as a portal to new physics beyond the SM

Even if more than 10 years have passed since the discovery of the Higgs boson, the exploration of the Higgs sector is still far from being completed. The couplings of the Higgs boson to the gauge bosons have been measured with a precision better than 10%, and the couplings to the third generation fermions have been measured with precision raging from 7% to 20%. Achieving higher precision in the

measurements of the Higgs boson couplings is a fundamental goal in order to test possible extensions of the SM affecting the Higgs boson production cross-sections and decay branching ratios produced by new physics scenarios involving particles beyond the energy scale of the LHC, but that could manifest themselves affecting the Higgs boson properties accessible at the LHC. The High Luminosity LHC (HL-LHC) will provide the necessary data to achieve precision of the order of 1% in the Higgs boson couplings to the gauge bosons and the third generation fermions [86].

However, the goal of the next years is not only to achieve higher precision in the measurements of the Higgs boson properties, but also to measure new interactions. As mentioned before, evidence of the coupling of the Higgs boson to muons is arising, but the sensitivity is still limited. The observation and confirmation of the compatibility of the Higgs boson couplings with muons is still a pending task for the field.

The confirmation of the Higgs boson couplings to charm quarks is a more challenging task. The large multi-jet background from QCD interactions and the poor resolution to charm jet identification impose difficulties for the observation of the $H \to cc$ decay, even at the HL-LHC [87–89]. Measuring the coupling of the Higgs boson to the up, down and strange quarks is beyond the reach of the HL-LHC, and the experimental path to achieve this goal is still unclear.

Finally, to fully understand the mechanism of electroweak symmetry breaking, measuring the Higgs boson self-coupling is required. The Higgs coupling trilinear coupling is accessible through the measurement of the Higgs boson pair production. The observation of this process is of great interest since it will probe experimentally the shape of the Higgs potential, and is within the reach of the High Luminosity LHC.

Beyond the genuine interest in measuring the Higgs boson properties and testing with precision the current SM predictions, the Higgs sector is also a portal to new physics beyond the SM and could be connected to the solution of some of the most important open questions in particle physics.

Modifications in the Higgs potential could have important cosmological implications. The existence of a first order phase transition in the early universe is a necessary condition for the generation of the baryon asymmetry of the universe through the so-called electroweak baryogenesis mechanism. The SM expects a smooth crossover, but modifications in the Higgs potential could induce a first order phase transition [90, 91]. The Higgs field could also have played a role in the inflationary period of the universe [92].

The mass of the Higgs boson is also connected to the stability of the vacuum. The Higgs potential has a non-zero minimum at the electroweak scale, and the shape of this potential is strongly affected by the mass of the Higgs boson, but also by the top quark mass, that affects the Higgs boson self-coupling through loop corrections. According to present measurements, at high energies a second

minimum of the Higgs potential appears, and the possibility of the universe to tunnel to this minimum could lead to an unstable vacuum [93].

The amount of CP violation allowed in the SM through the complex phases of the CKM and PMNS matrices is not enough to explain the matter-antimatter asymmetry in the universe. CP violation in the Yukawa sector and the Higgs production, even if not predicted by the SM, could be a source of new CP violation sources [94].

As the Yukawa sector is responsible for the generation of the mass of the quarks and the charged leptons, the Higgs sector could be connected to an explanation of why the masses of the three generations of fermions are so different [95, 96], or whether the Higgs sector is related to the generation of the neutrino masses [97].

About 85% of the matter in the universe is dark matter, but the nature of this matter is still unknown. Only gravitational interactions have been observed. The SM does not provide a candidate for dark matter, but the Higgs sector could be a portal to a new sector of particles constituting the dark matter content of the universe [98]. Furthermore, the Higgs boson itself could be responsible for the generation of masses of the dark matter particles [99].

In conclusion, the discovery of the Higgs boson at the LHC opened a new era in particle physics. The Higgs sector is a fundamental part of the SM, and the exploration of the Higgs boson properties is a fundamental test of the SM predictions. The experimental results have progressed significantly in the last years and have gone beyond the original expectations, but the exploration of the Higgs sector is still far from being completed. The precision measurements of the Higgs boson properties and the search for yet unobserved interactions are the main goals for the next years. Deviations with respect to the SM predictions would be a clear sign of new physics beyond the SM connected to some of the most fundamental open questions in particle physics.

Chapter 2

The LHC and the ATLAS detector

The proton-proton collisions data used to perform the analyses presented in this thesis were collected by the ATLAS detector at the LHC at CERN. The LHC is the world's largest and most powerful particle accelerator, designed to collide beams of protons and heavy ions up to energies of 14 TeV and 5.6 TeV per nucleon pair respectively. The ATLAS detector is one of the four main experiments at the LHC, and is a general-purpose detector where different subsystems are designed to measure the properties of particles produced in the collisions. This chapter introduces general aspects of the LHC and a brief description of the ATLAS detector.

2.1 The Large Hadron Collider

The LHC [100] is an approximately circular particle accelerator located at CERN. It is located in a tunnel of 26.7 km in circumference, at a depth between 45 and 175 meters below the surface. It is designed to accelerate protons and collide them at a maximum center-of-mass energy of 14 TeV. The protons are accelerated in two opposite beams, that are kept in a circular path by a magnetic field generated by superconducting magnetic dipoles arranged along the tunnel.

The LHC tunnel is divided into eight straight sectors and eight arcs. Particles are bent in the horizontal plane of the LHC ring thanks to the 1200 superconducting dipole magnets installed in the arcs, that are cooled down to a temperature of 1.9 K using superfluid helium to produce vertical magnetic fields of up to 8.3 T.

The straight sections are used to house the four interaction points where the beams collide and the main LHC experiments are located, and to host the radiofrequency cavities that are used to accelerate the beams. In total, the LHC has 16 radiofrequency cavities, 8 per beam, that provide accelerating fields of up to 5 MV/m. The frequency of the fields in the cavities is 400 MHz. The oscillating fields ensures that particles that have already reached the desired energy keep their energy while particles that have not yet reached the desired energy are accelerated

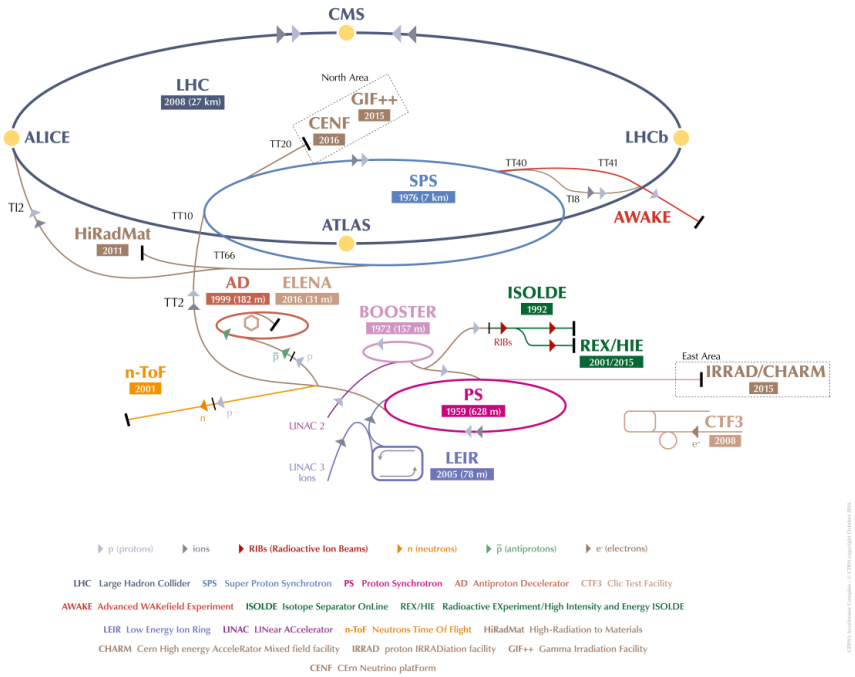


Figure 2.1: Illustration of the CERN accelerator complex [101].

(or decelerated if they are moving too fast). This ensures that particles are kept in concentrated bunches that are later made to collide in the four interaction points.

The LHC has also 854 quadrupole magnets that are used to focus the beams in the vertical and in the horizontal plane. There are also higher order multipole magnets that are used to correct higher order chromatic aberrations.

Protons are injected into the LHC from the SPS (Super Proton Synchrotron) at an energy of 450 GeV. The LHC has a complex injection chain that starts with protons being extracted from a hydrogen bottle and accelerated in the LINAC2 linear accelerator, a linear accelerator that accelerates protons to an energy of 50 MeV. The protons are then injected into the Proton Synchrotron Booster (PSB), where they are accelerated to an energy of 1.4 GeV. The next step is the Proton Synchrotron (PS), where protons are accelerated to an energy of 26 GeV. Finally, the protons are injected into the SPS, a synchrotron with a circumference of 7 km, where they are accelerated to an energy of 450 GeV. The SPS is the last accelerator before the LHC, and it is used to inject protons into the LHC. The accelerator complex is shown in Figure 2.1.

The LHC has four main experiments: ATLAS [102], CMS [103], LHCb [104] and ALICE [105]. ATLAS and CMS are general-purpose detectors, designed to measure a wide range of particles and interactions. LHCb is a forward spectrometer designed to study the properties of B mesons and other particles containing b-quarks. ALICE is a detector designed to study heavy ion collisions producing states with very high energy densities, like the quark-gluon plasma.

2.2 The ATLAS detector

ATLAS (A Toroidal LHC ApparatuS) is a general-purpose detector located at one of the four interaction points of the LHC. The detector consists of a cylindrical structure with a length of 44 meters and a diameter of 25 meters. Several subsystems are arranged and placed in layers around the interaction point to measure the properties of particles produced in the collisions. The cylindrical structure is closed by two end-caps, providing almost 4π coverage of the solid angle. An illustration of the ATLAS detector is shown in Figure 2.2.

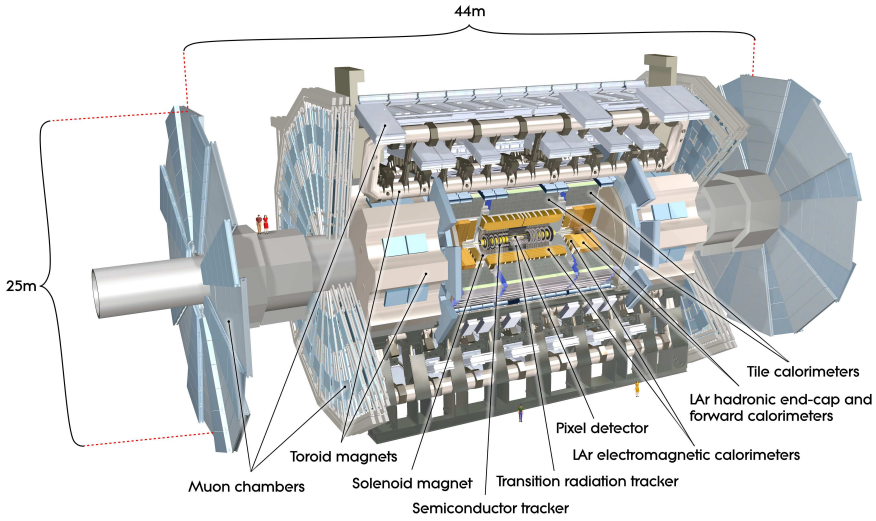


Figure 2.2: Illustration of a cut-away view of the ATLAS detector [102].

The program of the ATLAS experiment covers a wide range of physics topics. The SM predictions can be tested at the LHC with unprecedented precision and sensitivity. The discovery of the Higgs boson in 2012 was a major milestone for the experiment, and the study of its properties is one of its main scientific goals. Furthermore, the LHC is a top quark factory, and the study of top quark properties is also a major goal of the experiment since it provides a unique opportunity to test the SM at the highest energies, but also to search for new physics beyond the SM in the top quark sector. The program searches for new exotic particles and interactions, requiring the ATLAS detector to measure a wide variety of signatures. Finally, the ATLAS experiment also contributes to flavor physics measurements related to the study of B mesons and other particles containing b-quarks and other heavy flavor states and to the study of heavy ion collisions.

In order to achieve these goals, the ATLAS detector is divided into several subsystems, each designed to measure the properties of different types of particles. In the closest region to the interaction point, the inner detector is designed to measure the properties of charged particles produced in the collisions that are bent by a 2 T magnetic fields produced by a superconducting solenoid. The inner detector is surrounded by a calorimeter system, composed by a liquid argon electromagnetic calorimeter and a hadronic calorimeter based on a scintillating

barrel and liquid argon end-caps. The calorimeter system is used to measure the energy of the particles that produce electromagnetic and hadronic showers in the detectors. Finally, the outermost region of the detector is the muon spectrometer, designed to measure the properties of muons produced in the collisions and that are bent by a 4 T magnetic field produced by a toroidal magnet system. All these subsystems are described in the following sections, together with the forward detectors and the trigger and data acquisition system.

2.2.1 ATLAS coordinate system

ATLAS has a right-handed coordinate system with its origin at the nominal interaction point in the center of the detector. The z-axis is along the beam pipe, the x-axis points from the interaction point to the center of the LHC ring, and the y-axis points upwards. The azimuthal angle ϕ is measured around the beam pipe starting from the x-axis, and the polar angle θ is measured from the z-axis. The positive and negative z-axis are defined as the A-side and C-side respectively. The polar angle is usually replaced by the pseudorapidity η, defined as $\eta = -\ln(\tan(\theta/2))$. This quantity is equal to the rapidity, defined as $y = 0.5\ln((E+p_z)/(E-p_z))$, in the limit of massless particles, and is particularly useful to describe the angular distribution of particles produced in the collisions since differences in rapidity are invariant under Lorentz boosts along the beam axis. The angular separation between two particles is usually measured in terms of the distance in the $\eta - \phi$ plane, defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The ATLAS coordinate system is illustrated in Figure 2.3.

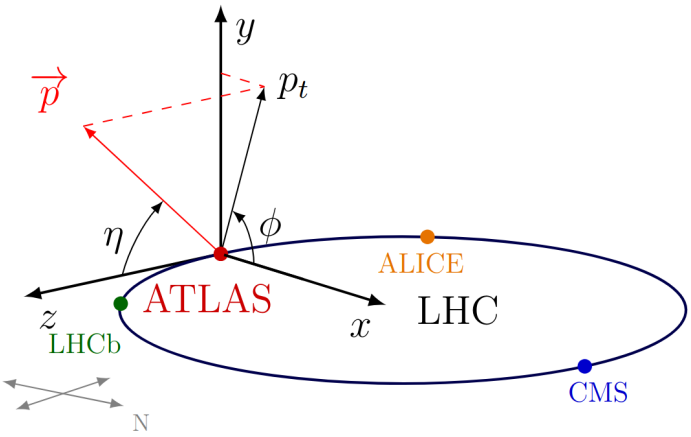


Figure 2.3: Illustration of the ATLAS coordinate system [106].

2.2.2 Inner Detector

The Inner Detector (ID) [107, 108] is the innermost subsystem of the ATLAS detector, and is a precise tracking system designed to measure the properties of

charged particles produced in the collisions. The ID is immersed in a 2 T magnetic field produced by a superconducting solenoid, and is composed of three subdetectors: the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The first two subdetectors are silicon-based detectors, while the TRT is gas straw-tube chamber filled with transition radiation material.

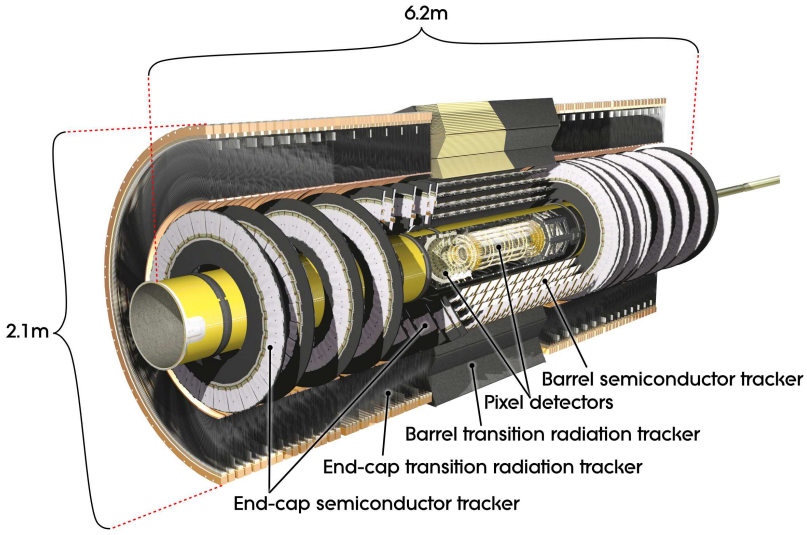


Figure 2.4: Cut-away view of the ATLAS Inner Detector [107].

The three subdetectors together form a system with a length of 6.2 m and a diameter of 2.1 m, covering a pseudorapidity range of $|\eta| < 2.5$. The ID itself is subdivided into a barrel region, covering the range $|\eta| < 1.4$, and two end-cap regions, covering the range $1.4 < |\eta| < 2.5$. When particles traverse the ID, they leave hits in the subdetectors that are used to reconstruct the tracks of the particles, bent by the magnetic field. Using the track information, it is possible to measure the momentum of the particles and determine their charge. The reconstructed tracks can be used in order to determine the primary vertex of the collision, and to identify other vertices that are displaced from the primary vertex and that may originate from the decay of heavy-flavor hadrons that travel a measurable distance before decaying. Schematic views of the ATLAS Inner Detector are shown in Figures 2.4 and 2.5.

The Pixel Detector consists of three layers of silicon pixel sensors in the barrel region and three disks in each end-cap region. The pixel sensors have a size of $50 \times 400 \mu\text{m}^2$ and a thickness of 250 μm. The three layers in the barrel area are located at radii of 50.5, 88.5 and 122.5 mm, while the disks in the end-cap region are located at z-positions of 495, 580 and 650 mm. This structure leads to a resolution of 10 μm in the $r - \phi$ plane and 115 μm in the z-direction for the barrel region, and 115 μm in the $r - \phi$ plane and 10 μm in the z-direction for the end-cap region. An additional layer of pixel sensors, the Insertable B-Layer (IBL) [109], is located at a radius of 33.25 mm in the barrel region, and contains planar and 3D pixel sensors with a size of $50 \times 250 \mu\text{m}^2$ and a thickness of 200 μm and 250

μm respectively. The finer granularity of the IBL improves the impact parameter resolution, especially for low-p_{T} tracks, due to the closer proximity of the sensors to the interaction point and the smaller pixel size. This is particularly important for the identification of jets originating from heavy-flavor quarks.

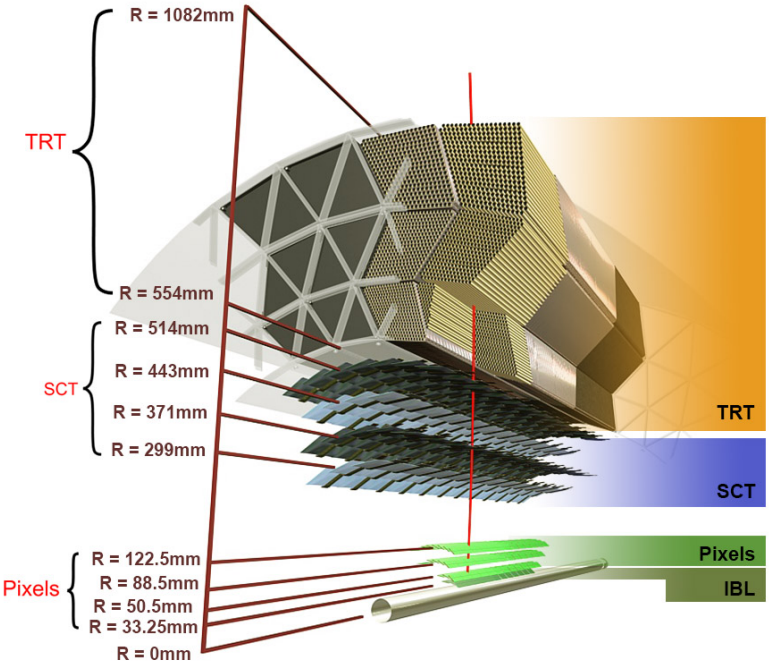


Figure 2.5: Schematic view of the ATLAS Inner Detector showing the Pixel Detector, the Semiconductor Tracker and the Insertable B-Layer [110].

The SCT is composed of four layers of silicon microstrip sensors in the barrel region and nine disks in each end-cap region. The detector consists of 4088 modules, (2112 of them in the barrel) each containing two strip sensors with a pitch of 80 μm that are mounted back-to-back but with an angle of 40 mrad. This configuration allows to measure the tracks with a resolution of 17 μm in the transverse plane and 580 μm in the longitudinal direction.

The TRT is composed of 52544 straw tubes in the barrel region and 122880 in the end-cap region. The tubes have a diameter of 4 mm and are filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2. When a charged particle traverses the tube, it ionizes the gas. The high negative voltage applied to the straw tubes allows the electrons to drift towards the anode wire, producing a signal that is read out by the electronics. The TRT provides a resolution of 130 μm in the $r-\phi$ plane covering the range $|\eta| < 2.0$.

Combining the information from the three subdetectors, it is possible to measure the curvature of the tracks and infer the charge and the momentum of the particles. The target resolution for the momentum measurement is $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 0.05\% p_{\mathrm{T}}/\mathrm{GeV} \oplus 1\%$, where p_{T} is the transverse momentum of the particle.

2.2.3 Electromagnetic calorimeter

Outside the ID, the electromagnetic calorimeter is designed to measure the energy of electrons and photons produced in the collisions. The ATLAS electromagnetic calorimeter [111] is a sampling calorimeter composed of liquid argon (LAr) as the active medium and lead as the absorber. It is divided in a barrel covering the range $|\eta| < 1.475$ and two end-cap regions covering the range $1.375 < |\eta| < 3.2$. The absorber plates and the electrodes are arranged in a structure of accordion-shaped layers, that allows to achieve a full azimuthal coverage without cracks. High voltage is applied to the electrodes to collect the ionization electrons produced by the particles in the liquid argon.

In the barrel region, the electromagnetic calorimeter is divided into three cylindrical layers. The first layer has a depth of 4.3 radiation lengths (denoted as X_0 and defined as the distance in which an electron looses $1/e$ of its energy), and a granularity of $\Delta\eta \times \Delta\phi = 0.0031 \times 0.0982$. The second layer has a depth of 16 radiation lengths and a granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.0245$. The third layer has a depth of 2 radiation lengths and a granularity of $\Delta\eta \times \Delta\phi = 0.05 \times 0.0245$. The thickness of the calorimeter is 22 radiation lengths in the barrel region at $\eta = 0$, but it increases to 33 radiation lengths at $|\eta| = 1.3$. A presampler is located in front of the calorimeter to correct for energy losses in the solenoid before the particles reach the calorimeter. The presampler consists of a LAr detector with a thickness of 1.1 cm in the barrel region and 0.5 cm in the end-cap region. The sketch of a barrel LAr calorimeter module is shown in Figure 2.6. The total radiation length of the electromagnetic calorimeter is about 22 X_0 in the barrel region and 24 X_0 in the end-cap region. The energy resolution of the electromagnetic calorimeter increases with the energy of the incident particle due to the stochastic nature of the electromagnetic showers. The energy resolution is parametrized as $\sigma_E/E = 10\%/\sqrt{E/\text{GeV}} \oplus 0.7\%$.

2.2.4 Hadronic calorimeter

The ATLAS electromagnetic calorimeter is surrounded by the hadronic calorimeter, designed to absorb the hadrons produced in the collisions. The hadronic calorimeter is a sampling calorimeter composed of scintillating tiles in the barrel region and liquid argon in the end-cap and forward regions.

2.2.4.1 Tile Calorimeter

The Tile Calorimeter (TileCal) [112] is a sampling calorimeter covering the central region of the ATLAS detector, with a pseudorapidity range of $|\eta| < 1.7$. The TileCal is composed of a central long barrel (LB) of 5.8 m in length, covering the range $|\eta| < 1.0$, and two extended barrels (EB) with 2.6 m length, covering the range $0.8 < |\eta| < 1.7$. The TileCal as an inner radius of 2.28 m and an outer radius of 4.25 m, accounting for a total thickness of around 7.4 interaction lengths (mean

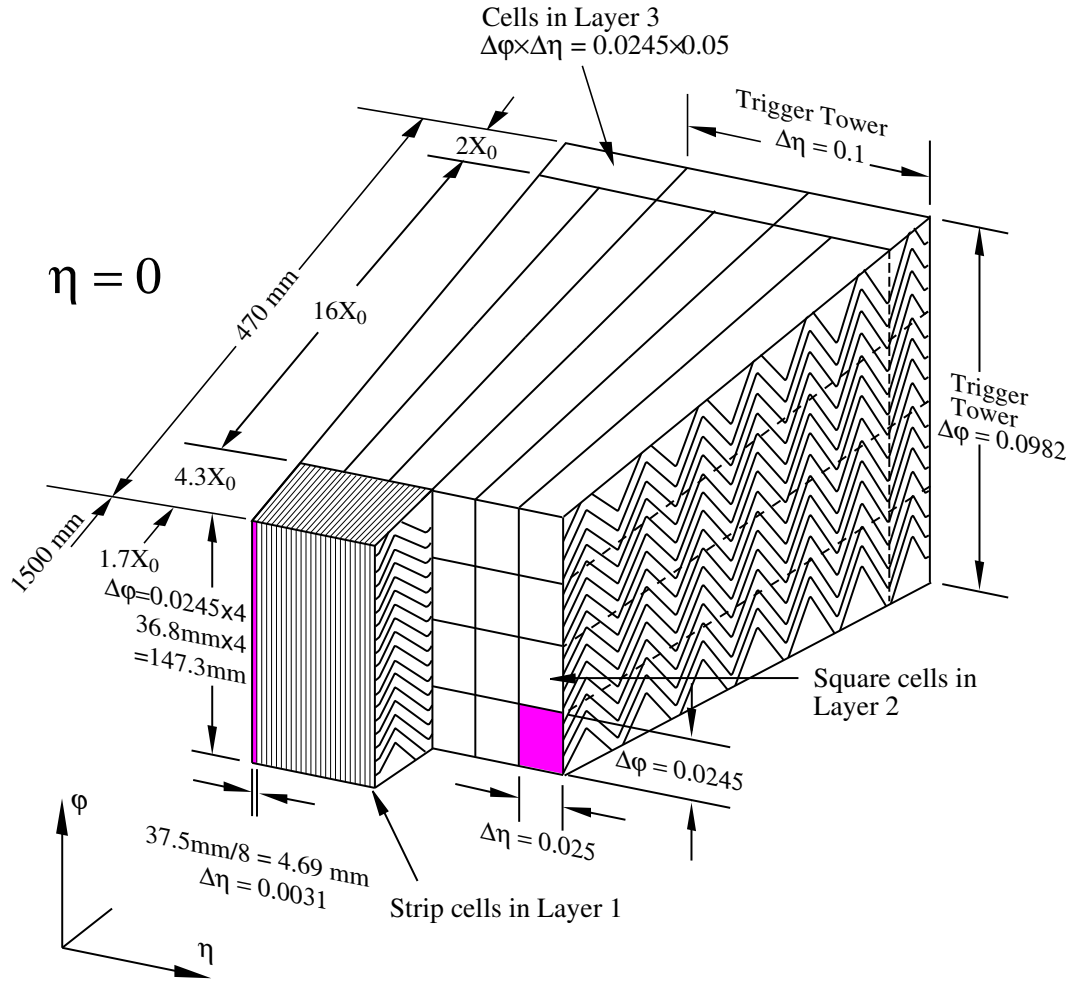


Figure 2.6: Sketch of a barrel LAr calorimeter module [102].

distance travelled by a hadronic particle before undergoing an inelastic nuclear interaction). The calorimeter is composed of steel as the absorber and scintillating tiles made of polystyrene as the active material. Each barrel is divided azymuthally into 64 wedge-shaped modules, each one covering a range of $\Delta\phi \approx 0.1$. Each module is divided into three layers, with a thickness of 1.5, 4.1 and 1.8 interaction lengths in the barrel region and 1.5, 2.6 and 3.3 interaction lengths in the extended barrels.

The modules have a periodic structure, with scintillating tiles of 3 mm thickness alternating with steel absorber plates of 4 mm thickness. The scintillating tiles are coupled to wavelength-shifting fibers that collect the light produced by the passage of the particles through the tiles. The grouping of the fibers defines the readout cells, with a size of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the barrel region and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ in the extended barrels. The geometry of the cells has a pseudo-projective structure in pseudorapidity.

The fibers associated with a single cell are connected to two photomultiplier tubes, housed in the outermost part of each module. This does not only provide redundancy in the readout, but also allows reducing non-uniformities in the response

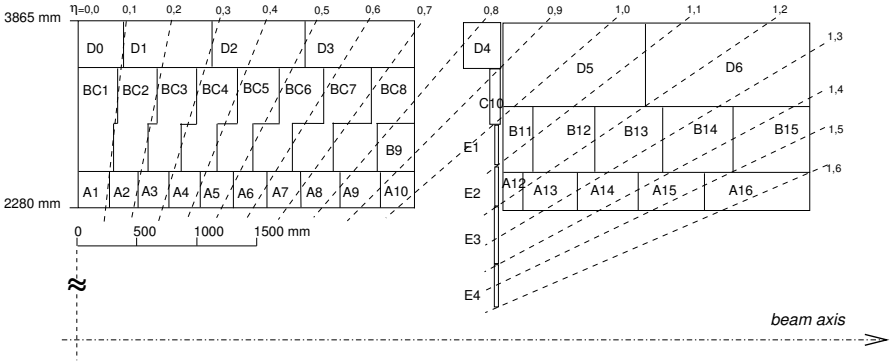


Figure 2.7: Segmentation of half of the long barrel and an extended barrel in the different readout cells in the TileCal. The calorimeter is symmetric around the interaction point [113].

of the calorimeter that depend on the position of the particles that traverse the tile. The PMTs and the on-detector electronics are housed in 1.4 m long aluminum drawers that can be extracted from the calorimeter for maintenance. The long barrel is read out by the electronics held in two drawers, while the extended barrels are read out by the electronics held in a single drawer. This structure defines four readout partitions, called LBA, LBC, EBA and EBC. Each drawer contains 42 channels (readout of a PMT) in the long barrel and 32 in the extended barrels. There are 9852 channels in total in the TileCal, reading 5182 cells.

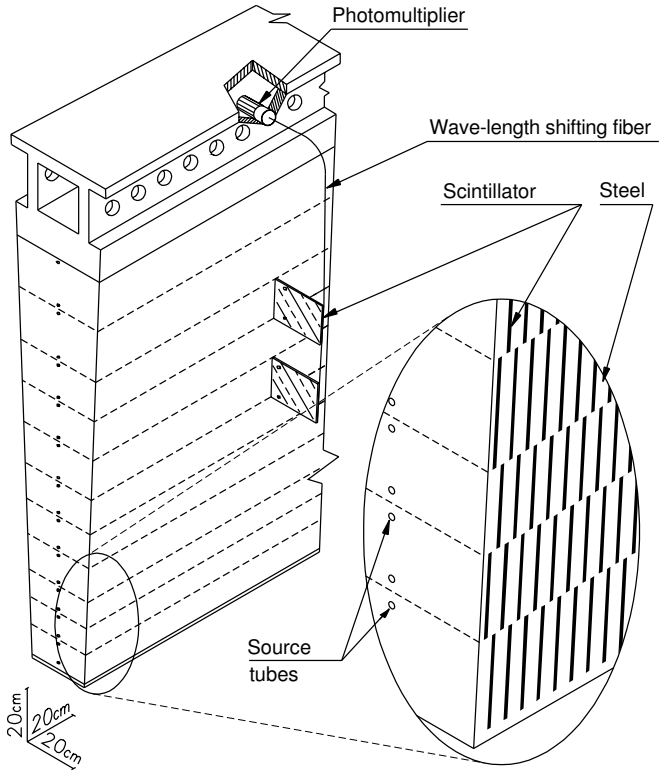


Figure 2.8: Schematic of a module of the TileCal [113].

The signal produced in the PMTs by the light collected by the fibers is shaped by a passive circuit that produces a signal with a width of around 50 ns. The signal is amplified with two different gains, the high gain (HG) and the low gain (LG), with a gain ratio of 64. The analog signals are sampled and digitized at the frequency of 40 MHz using 10-bit Analog to Digital Converters (ADC). The digitized signals are stored in pipeline memories, awaiting for a trigger decision, after which the data is sent to the off-detector electronics. The sampling time of the ADCs can be adjusted to ensure that the peak of the signal is sampled. Once the trigger accepts an event produced in one bunch cross, seven samples around the accepted event are read out and sent to the off-detector electronics, where the Read-Out Driver (ROD) boards reconstruct the energy and time of the signals.

The PMT block is responsible for converting light produced in the scintillating tiles into analog signals. Each PMT block comprises a light mixer, a Hamamatsu R5900 photomultiplier tube [114], a high-voltage divider, and a 3-in-1 card. The light mixer, placed between the fiber bundle and the PMT's photocathode, ensures uniform light distribution, removing any correlation between fiber position and photocathode area. The Hamamatsu R5900 PMT is a compact device ($28 \times 28 \times 28\text{mm}^3$) with 8 to 10 amplification stages. It operates at a voltage of approximately 800 V, and features a rise time of around 1.4 ns, allowing rapid response to incident light. Magnetic sensitivity is reduced by placing the PMT in a μ-metal cylinder, providing shielding up to 200 G. The high-voltage divider is a printed circuit attached to the 3-in-1 card, distributing high voltage to the PMT. Initially, a 1 kV high-voltage power supply provides power to the super-drawer, with adjustable voltages applied to individual PMTs via optocontroller boards. A scheme of a TileCal PMT block is shown in Figure 2.9.

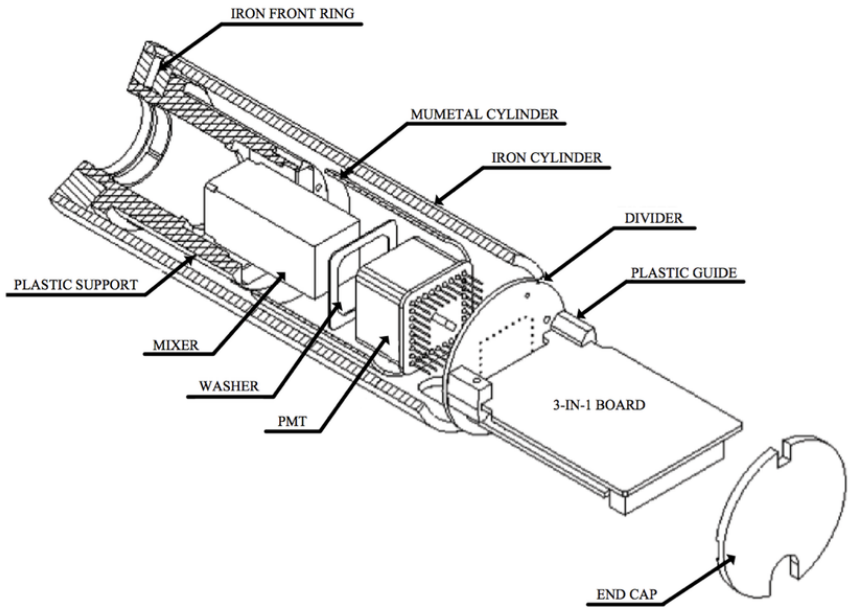


Figure 2.9: Schematic of a TileCal PMT block [115].

The 3-in-1 card [116], measuring 7 cm by 4.7 cm, performs the primary analog signal processing for the front-end electronics. It shapes the raw PMT signal to standardize the pulse shape and minimize fluctuations between pulses. The shaped signal is amplified through two operational amplifiers, producing low-gain and high-gain signals with a fixed gain ratio of 64. These amplified signals are sent to the digitizer boards. Additionally, the 3-in-1 card generates a fast differential trigger signal, derived from the low-gain output, which is forwarded to trigger sum boards on the motherboards for trigger tower summation. For calibration, the 3-in-1 card includes a Charge Injection System (CIS), which injects a known charge into the shaping circuit. Two capacitors (5.1 pF and 100 pF) are used, with the smaller capacitor allowing injections of up to 40 pC for the high-gain ADC, and the larger capacitor supporting up to 800 pC, covering the full dynamic range. The card also features a slow integrator, averaging the PMT signal over a 10 ms period, which is multiplexed and sent to the integrator card via a CANBUS interface. The integrator card, equipped with an ADC and microprocessor, digitizes the signal and transmits the data to the counting room for calibration and Minimum Bias monitoring.

The motherboard provides the structural and electrical support for the electronics in a drawer. Each super-drawer contains two motherboards, which distribute power and Timing, Trigger, and Control (TTC) commands to the 3-in-1 cards. The motherboards support up to four digitizer boards and one optical interface board.

Each digitizer board [117] processes signals from six PMT blocks, receiving both low-gain and high-gain signals from the 3-in-1 cards. These signals are digitized every 25 ns using 10-bit ADCs, with a clock adjustable in steps of 106 ps. Each digitizer board contains two Tile Data Management Units (DMUs), custom ASICs that store data in pipelines. Each DMU handles data from three PMTs, with a programmable pipeline length of up to 256 readout samples, providing a latency of up to 6.4 μs. This latency exceeds the ATLAS Level 1 trigger requirement of 2.5 μs, ensuring adequate time for trigger decisions. Timing and synchronization are managed by a so-called TTCrx chip [118]. In total, eight digitizer boards are installed in each Long Barrel super-drawer, while six digitizer boards are used in each Extended Barrel super-drawer.

The optical interface board [119], located at the center of the super-drawer, distributes TTC optical signals and collects data from the 16 Tile DMUs in the super-drawer. Data from each DMU is transmitted aligned to a common clock, packed into 32-bit words, and temporarily stored in event frames. These frames are sent to the back-end electronics through two redundant optical links operating at 640 Mbps, using G-link transmitters. Data integrity is verified through Cyclic Redundancy Checks (CRC), performed between the digitizers and the interface board, as well as on the event frames before transmission. The optical interface board can store up to 16 physics event frames, ensuring efficient data transmission during regular operation.

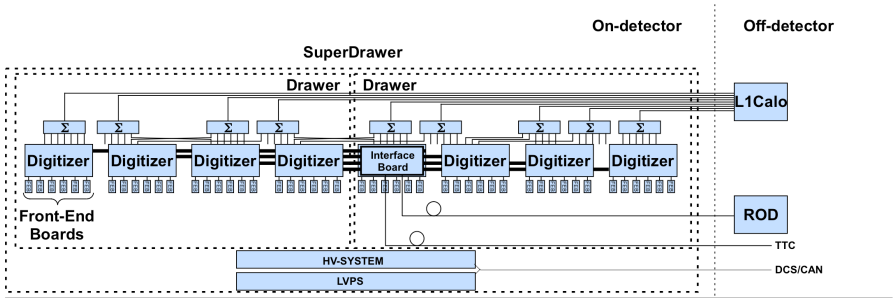


Figure 2.10: The front-end electronics consists of eight identical Digitizer Boards, each reading signals from six PMTs. The analog signals are digitized every 25 ns and stored in pipeline memories within the DMUs on the boards. Upon receiving a Level 1 accept, selected event samples are stored in a buffer. The Interface Card retrieves the event frames from the DMUs, formats the data, and transmits it to the off-detector Read-Out Drivers via optical fibers. The digitizer boards are chained, with data flowing from the outer boards to the Interface Card at the center of the superdrawer. In parallel, low-gain signals are summed by adder boards and sent to the Level 1 Calorimeter Trigger. The Trigger, Timing, and Control (TTC) signals are transmitted to the front-end through optical links and distributed to the Digitizer Boards via the Interface Board. [113].

The back-end electronics system of the Tile Calorimeter is located in the counting rooms, approximately 70 meters away from the detector. The back-end electronics is organized into four partitions, each responsible for the readout of both the Long and Extended Barrels. These partitions are physically distributed across different crates: a 6U Versa Module Eurocard (VME) TTC crate and a 9U VME ROD crate.

The Read-Out Driver (ROD) is the central of the back-end electronics, responsible for receiving data from the super-drawers via optical links. The system comprises a total of 32 RODs, organized into four VME crates, with each crate dedicated to one barrel (Long or Extended). Each crate contains eight RODs.

Each ROD module consists of a ROD motherboard and four Processing Units (PUs), each equipped with two Digital Signal Processors (DSPs). These DSPs handle the processing of data before it is transmitted to the ATLAS Data Acquisition system. The DSPs perform tasks such as zero suppression, data formatting, and error checking. The data from eight super-drawers is processed by a single ROD, and the processed event fragments are transmitted to the DAQ system at a rate determined by the trigger rate.

2.2.4.2 End-cap and forward calorimeters

The hadronic end-cap calorimeter is located behind the electromagnetic LAr end-caps and cover the range $1.5 < |\eta| < 3.2$. This calorimeter is composed of two wheels in each end-cap, with copper as the absorber and liquid argon as the active medium. Each of the wheels is composed of 32 wedge-shaped modules, with 24 copper plates of 25 mm thick in the first wheel and 16 modules with 50 mm thick in the rear wheel. The size of the readout cells is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ on the

region of $1.5 < |\eta| < 2.5$ and $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ on the region of $2.5 < |\eta| < 3.2$. The use of liquid argon as active material in the forward region is motivated by the high radiation levels in this region, since it is more radiation hard than the plastic scintillators used in the TileCal.

Finally, the forward calorimeter is located in the region $3.1 < |\eta| < 4.9$ and is composed of copper and tungsten as the absorber and liquid argon as the active material. The particle flux is very high in this region, and the calorimeter is designed to be very radiation hard. The calorimeter is divided into three layers in each end-cap, the first one used for the measurement of electromagnetic showers and the other two for hadronic calorimetry. In each module, metal rods filled with liquid argon are inserted inside holes in the absorber material. The small LAr-filled gaps allow fast signal collection to deal with the high particle flux.

2.2.5 Muon spectrometer

Muons are minimum-ionizing particles at the energy range of $\mathcal{O}(1)$ GeV to $\mathcal{O}(1)$ TeV, and they are the only particles (apart from neutrinos) that can traverse the calorimeter system and reach the muon spectrometer.

The muon spectrometer [120] is the outermost region of the ATLAS detector, located in a 4 T magnetic field produced by a toroidal magnet system. It consists of four different types of muon chambers, two dedicated to the precise measurement of the muon momentum and two dedicated to the triggering of muons. A scheme of the muon spectrometer is shown in Figure 2.11.

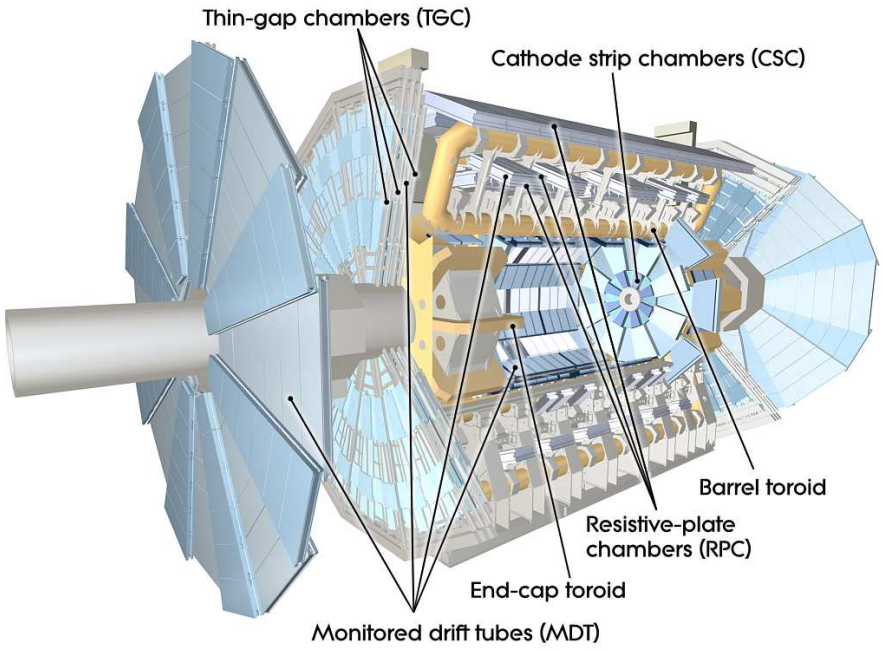


Figure 2.11: Cut away of the ATLAS Muon Spectrometer [102].

The precision chambers are the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC). The MDTs are located in the barrel and end-cap

regions, and cover the range $|\eta| < 2.7$. The MDTs are composed of gas-filled aluminum drift tubes with a diameter of 30 mm and a length between 0.7 and 6.3 m, with a cathode wire made of tungsten and rhenium. The chambers are arranged in three cylindrical layers in the barrel region, each of them at a radius of 5, 7.5 and 10 m. In the end-cap the chambers are arranged in four layers that are perpendicular to the beam axis, at distances of 7.4, 10.8, 14 and 21.5 m from the nominal interaction point.

Since the MDTs have a limited rate capability and the flux of particles is very high in the forward region, the CSCs are used in the range $2 < |\eta| < 2.7$. The cascadas are multiwire proportional chambers with copper cathode strips and anode wires made of the same material as the MDTs. The anode wires are arranged radially, perpendicular to the beam axis, and the two cathodes are segmented, but in different directions (one of them is segmented perpendicular to the anode wires and the other one is segmented parallel to them) to provide two-dimensional information.

The resolution of the MDTs is 35 μm, and they have a slow maximum drift time of 700 ns, while the resolution of the CSCs is 40 μm in the bending plane and 5 mm in the perpendicular plane, but they have a faster maximum drift time of 20 ns.

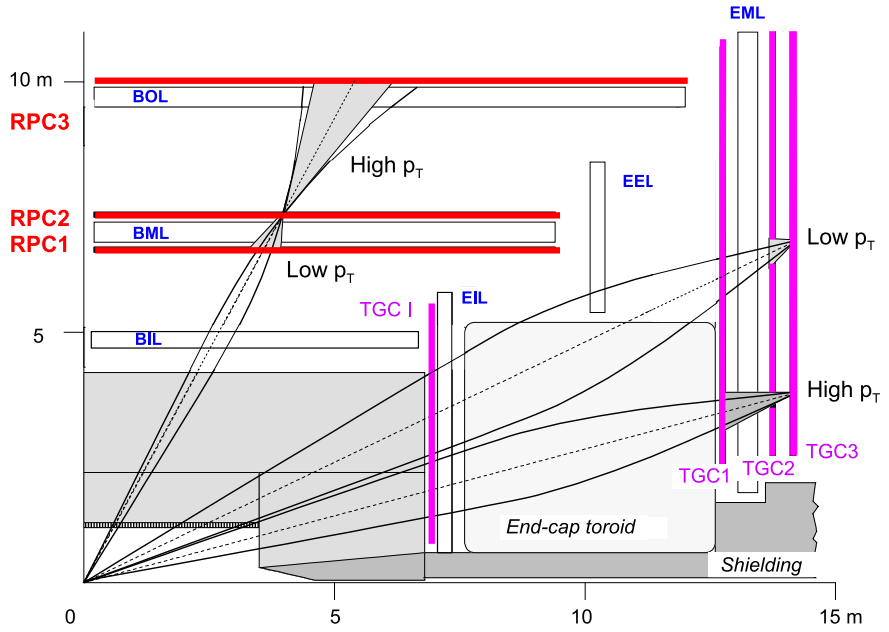


Figure 2.12: Schematics of the muon trigger system [102].

Two types of chambers are used for the triggering of muons within the time constraints of the LHC bunch crossing. The Resistive Plate Chambers (RPC) are used in the barrel. They consist of two parallel electrode plates made of bakelite with a gas gap of 2 mm. There are three layers of RPC, two of them surrounding the intermediate layer of the MDTs and the third one above or below the third layer of MDTs. High voltage is applied to the plates, that are segmented orthogonally

to provide two-dimensional information. This allows to measure the ϕ coordinate of the muons in the barrel, not accessible by the MDT. The RPC have a time resolution of 1.5 ns and a spatial resolution of 10 mm. In the end-cap region, the Thin Gap Chambers (TGC) are used. The TGC are multiwire proportional chambers with a wire to cathode distance of 1.4 mm. The chambers have a double purpose: they are used for the triggering in the range $1.05 < |\eta| < 2.4$ and for muon tracking measurements in the range $1.05 < |\eta| < 2.7$. The TGC have a time resolution similar to the RPC. A scheme of the muon trigger system is shown in Figure 2.12.

The relative momentum resolution of the muon spectrometer is $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 10\%$ for a muon with $p_{\mathrm{T}} = 1$ TeV. As in the ID, the resolution degrades with the momentum since the curvature of the tracks is smaller.

2.2.6 Magnet system

As said before, the ID and the muon spectrometer are immersed in two different magnetic fields, produced by a superconducting solenoid and a toroidal magnet system respectively. The 2 T solenoidal field produces a magnetic field with field lines in the z-direction, deflecting particles in the $r - \phi$ plane. The solenoid is 5.3 m long and has a diameter of 2.5 m. Its thickness was minimized to 4.5 cm to reduce the energy loss of the particles before entering the calorimeter system.

The 4 T toroidal field embedded in the muon spectrometer produces an azimuthal magnetic field, with field lines following a circle in a transverse plane and deflecting particles in the $r - z$ plane. A solenoidal magnetic field would not be perpendicular to muons produced in the forward region. The toroidal configuration provides a solution to this problem, and furthermore minimizes the amount of material in the muon spectrometer, reducing the multiple scattering of the muons, that can limit the momentum resolution. The toroidal magnet system is composed of eight coils in the barrel region and two end-cap toroids, each of them with eight coils.

2.2.7 Forward detectors

The ATLAS Forward Detectors (LUCID, AFP, ALFA, and ZDC) are specialized components positioned at various locations along the beam line, outside the central ATLAS detector, and are designed to measure particles produced in high-energy proton-proton collisions that travel close to the beam line. These detectors have a wide range of applications, from monitoring the luminosity of the LHC to studying elastic pp scattering or heavy-ion collisions.

LUCID (Luminosity Cherenkov Integrating Detector) [121] is used primarily for the measurement of the luminosity at ATLAS. It detects Cherenkov radiation produced by particles passing through its detector material, providing real-time monitoring of the instantaneous luminosity.

The AFP (ATLAS Forward Proton) [122] detectors are designed to study protons that remain intact after collisions and are scattered at very small angles relative to the beam. These detectors allow for precise measurements of the protons' momentum, which can be used to study diffractive processes and search for new physics phenomena that involve forward protons.

ALFA (Absolute Luminosity For ATLAS) [123] is another detector system dedicated to luminosity measurements, but it works by detecting elastically scattered protons at very small angles. ALFA is used in conjunction with other systems to calibrate the luminosity measurements more precisely, ensuring accurate determination of the interaction rates at the LHC.

The ZDC (Zero Degree Calorimeter) [124] is positioned to detect neutral particles, such as neutrons and photons, produced in the very forward direction of the collision. The ZDC is particularly useful in heavy-ion collisions and for studying phenomena like ultra-peripheral collisions, where particles are emitted at angles very close to zero degrees relative to the beam line.

2.2.8 Trigger and Data Acquisition system

During the LHC operations, bunches of protons collide every 25 ns, corresponding to a rate of 40 MHz. Each of the collisions produces a large number of particles that are detected by the ATLAS detector, and the information of a single event has a size of about 1.5 MB. Due to bandwidth and storage limitations, it is not possible to store all the events produced by the LHC, and a trigger system is used to select events with interesting signatures. A diagram of the ATLAS Trigger and Data Acquisition (TDAQ) system during the Run 2 is shown in Figure 2.13.

The trigger system is divided into two levels: the Level-1 trigger and the High-Level Trigger (HLT) [125]. The Level-1 trigger is a hardware-based system utilizing custom and commercial electronics that is designed to reduce the event rate from 40 MHz to 100 kHz within a latency requirement of 2.5 μs. All the information from the front-end electronics of the sub-detectors is stored in pipeline memories at the rate of 40 MHz. The buffers store the information of the events for 2.5 μs, fixing the latency of the system. The trigger decision is based on the information of the calorimeter and the muon spectrometer. The Level-1 Calorimeter (L1Calo) trigger uses reduced information from the calorimeter system (trigger towers) to identify regions of high energy deposition. The Level-1 Muon (L1Muon) trigger uses the information from the RPCs and TGCs to identify muon candidates, measuring the transverse momentum of the muons and assigning them to the correct bunch crossing. The information from the L1Calo and the L1Muon is combined in the Central Trigger Processor (CTP) to make the final decision. If an event is accepted, a signal (L1 accept, L1A) is sent to the front-end electronics of the sub-detectors to read out the full event information corresponding to the event and stored in the pipeline memories. The maximum rate of the L1A signal in ATLAS is 100 kHz.

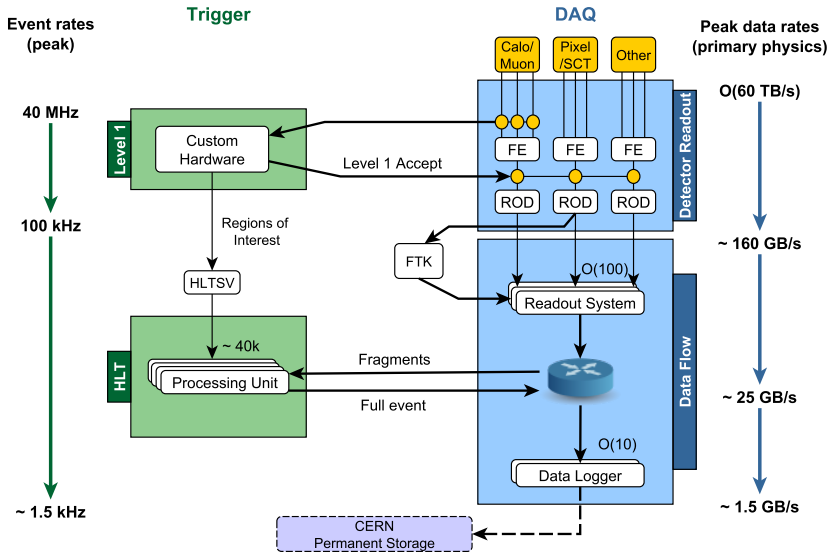


Figure 2.13: Diagram of the ATLAS Trigger and Data Acquisition system during Run 2 [126].

The data from the sub-detectors is processed by the sub-detector Read-Out Drivers (RODs) and sent to the High-Level Trigger (HLT) system. The HLT is a software-based system that runs on a farm of computers where events accepted by the L1 trigger are partially reconstructed and analyzed. The HLT reduces the event rate from 100 kHz to 1 kHz, and the events that pass the HLT are stored in the permanent storage system and sent to the CERN Tier-0 computing center for offline reconstruction. While waiting for the HLT decision, the data from the sub-detectors is stored in the Read-Out Buffers (ROBs) of the Read-Out System (ROS).

2.2.9 The LHC computing grid

The LHC computing Grid, also known as the Worldwide LHC Computing Grid (WLCG) [127], is a global infrastructure designed to provide the necessary computing resources for storing, distributing, and analyzing the vast amounts of data generated by the Large Hadron Collider (LHC). With experiments such as ATLAS producing several petabytes of data annually, a distributed computing system is essential for managing this data and enabling global collaboration.

The WLCG operates through a hierarchical structure, divided into four tiers: Tier-0 (CERN), Tier-1 (national-level centers), Tier-2 (regional and university clusters), and Tier-3 (local clusters). CERN, as Tier-0, handles the initial data storage and processing before distributing it to 13 Tier-1 centers worldwide. These Tier-1 centers are responsible for large-scale reprocessing and archival storage, and they distribute processed data to approximately 170 Tier-2 sites. Tier-2 sites,

typically associated with universities or research institutions, provide the computing power needed for analyses and Monte Carlo simulations. Tier-3 sites serve local computing needs. Figure 2.14 shows a schematic view of the WLCG structure.

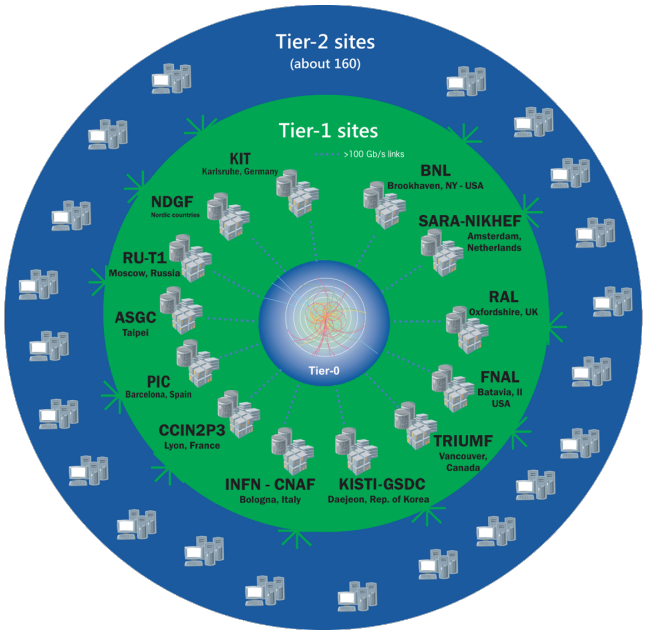


Figure 2.14: The Worldwide LHC Computing Grid hierarchical structure [128].

WLCG integrates over 1.4 million computer cores and provides 1.5 exabytes of storage, spread across 170 computing centers in 42 countries. This infrastructure supports over 12,000 physicists globally, enabling them to access and analyze LHC data almost in real time. On average, the grid runs over 2 million computing tasks per day, and by the end of the second LHC Long Shutdown (LS2), global data transfer rates surpassed 260 GB/s.

Advanced middleware solutions manage the allocation of resources, ensuring the efficient sharing of computing power and storage across different organizations. The WLCG enables the distribution of raw and processed data from CERN to institutions worldwide, facilitating large-scale data analysis and simulations.

Chapter 3

The Phase II upgrade of the ATLAS Tile Calorimeter and the simulation of its read-out electronics

3.1 The upgrade of the ATLAS detector towards the HL-LHC

The High Luminosity LHC [129] is a project that aims to increase the instantaneous luminosity of the LHC by a factor of 5 to 7 with respect to the design value of the LHC. This will allow reaching an integrated luminosity of 3000 fb^{-1} by the end of the HL-LHC era. In this way, the machine will operate at the energy frontier, with a center-of-mass energy of 14 TeV, but it will also provide large datasets for precision measurements and searches for new physics.

The ATLAS detector will have to cope with an instantaneous luminosity of up to 7.5×10^{34} cm^{-2} s^{-1}, which might lead to up to 200 simultaneous proton-proton interactions (pile-up) per bunch crossing. The detector needs to be able to operate under conditions of high occupancy and intense particle fluxes, while maintaining the performance of the experiment.

At the end of the LHC period, the silicon detectors in the inner detector will be close to their radiation tolerance limits. Furthermore, the TRT will be unable to operate under the increased luminosity and pile-up conditions of the HL-LHC, since the occupancy levels would compromise the tracking performance. To guarantee the performance of the vertex and tracking reconstruction, the inner detector will be fully replaced by a new all-silicon tracker, the Inner Tracker. Furthermore, a High Granularity Timing Detector will be installed in the forward region in order to tag tracks and calorimeter clusters with precise timing information. This will mitigate the performance degradation due to the high pile-up conditions [130].

The trigger will also be upgraded to cope with an increased trigger rate of up to 1 MHz, and also to be able to increase the granularity of the detector information. More sophisticated algorithms will be used to keep a high efficiency for

the physics signals, while improving the background rejection capabilities, crucial in high pile-up conditions [131].

Finally, the increased radiation levels require particular attention in the design of new readout electronics for the detectors, to protect them from the effects of single event upsets and other radiation-induced effects.

In order to cope with the increased radiation levels and to be compatible with the new trigger architecture, several systems of the ATLAS detector will be upgraded. The Liquid Argon and Tile Calorimeters will upgrade their on-detector and off-detector electronics. The Muon detectors will install new chambers in the barrel (RPCs and MDTs), replace the TGCs in the gap between the barrel and the end-caps and upgrade the readout electronics for the already installed RPCs, TGCs and MDTs to make them fully compatible with the trigger architecture.

The design of the upgrades of the ATLAS detectors subsystems is already finished, and the different components are in pre-production or production phase. The installation of the new components will take place during the Long Shutdown 3 (LS3) of the LHC, which will take place after 2026. The new components will be commissioned during the LS3 and the LHC Run 4, which is expected to start in the early 2030s.

3.2 The Tile Calorimeter upgrade

The TileCal electronic components installed on the detector will be exposed to a total ionizing dose of up to 1000 Gy and a non-ionizing dose of 1-MeV neutron equivalent fluence of up to 10^{13} cm^{-2}. These radiation levels justify the complete replacement of the TileCal readout electronics, together with the replacement of the 10% of the PMTs that are most exposed to radiation. Apart from the readout electronics, the low voltage and high voltage systems will be substituted by new systems that are more radiation tolerant [132].

The readout electronics can be divided into two main parts: the on-detector electronics, which are installed in the detector, and the off-detector electronics, which are installed in the counting room (USA15). Radiation hardness is not required for the off-detector electronics, but their replacement is needed to be able to process the signals from the detector at a rate of 40 MHz and send them to the trigger system. A scheme of the TileCal readout electronics for the HL-LHC phase is shown in Figure 3.1.

Not all the PMTs will be replaced, but all of them will be equipped with High Voltage Active Dividers, to distribute the power among the dynodes of the PMTs independently of the current passing through the PMT base and maintain the non-linearities of the response below the 1% level, guaranteeing the accuracy of the energy measurements. The readout of the analog signals from the PMTs will be done by the FENICS boards. These boards will provide signal shaping and amplification through two readout paths, similarly to the current system. The fast

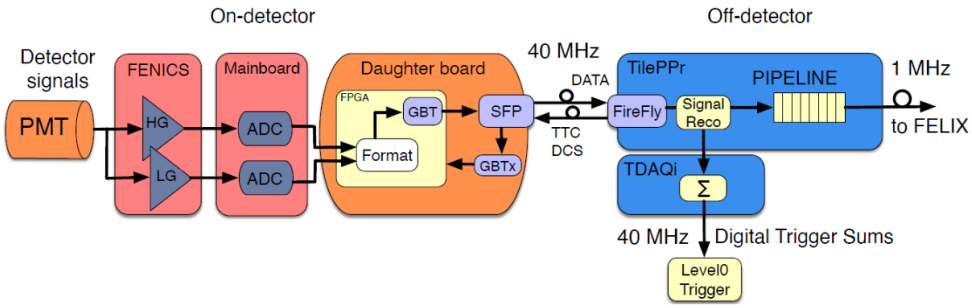


Figure 3.1: Scheme of the TileCal readout electronics for the HL-LHC phase.

readout, used for physics data taking, will amplify the signals with two gains with an approximate ratio of 1:40, covering signals from 200 fC to up to 1000 pC. The slow readout, used for calibration and monitoring, will integrate the PMT currents over a period of time of milliseconds, and will be able to cover over eight orders of magnitude in luminosity measurements thanks to the six different gains.

The signals from up to 12 FENICS boards will be digitized by the Mainboards. The fast readout will be digitized by 24 ADCs operating at 40 Msps with 12 bits of resolution, while the slow readout will use 12 16-bit ADCs [133]. Each Mainboard hosts a Daughterboard, which will be responsible for the high speed communication with the off-detector electronics through two 9.6 Gbps optical links. The Daughterboards will receive and distribute the information of the LHC clock and the configuration commands using two 4.6 Gbps downlinks [134]. Both the Mainboard and the Daughterboard are designed with two independent halves with different power lines to guarantee the redundancy of the system. The two PMTs that are responsible for reading one cell of the TileCal will be connected to different halves of the Mainboard, so that the failure of one half of the Mainboard does not compromise the readout of the cell.

The Daughterboard will send the data at a rate of 40 MHz to the Tile PrePro-cessor (PPr), hosted in an Advanced Telecommunications Computing Architecture shelf and containing four Compact Processing Modules (CPMs), designed with a double Advanced Mezzanine Card (AMC) with Kintex FPGAs. Each PPr will process the data from 2 TileCal modules. The PPr will be responsible for the online reconstruction of the pulse amplitude and time [135]. The Tile PreProcessor will also buffer the data in pipeline memories awaiting the Level 0 trigger decision. Using the reconstructed data, the TDAQ interface (TDAQi) will build trigger objects (energy measured in pseudo-projective towers and cells) to feed the trigger systems, that will be responsible for the trigger decision [136]. Once an event is accepted, the data will be sent to the Front End Link eXchange (FELIX) system, which will manage data aggregation and storage [137]. This readout principle is very different from the current system, where the data is sent to the Read Out Drivers only after the Level 1 trigger decision, and the reconstruction of the pulse amplitude and time is only done for the events that are accepted by the trigger

system. The upgrade of the back-end electronics is fundamental to be compatible with the new trigger architecture and to be able to process the signals from the detector at the rate of the LHC collisions of 40 MHz.

The possibility of performing precise energy reconstruction before the Level 0 trigger decision opens new possibilities to improve the trigger performance. Previous studies have shown that the turn-on curve of the trigger efficiency for jets with a transverse momentum above 70 GeV can be narrowed by more than 10% if the energy is reconstructed with a precision around the cell noise precision (20 MeV), instead of the current precision of $\mathcal{O}(1)$ GeV per trigger tower [132]. However, the studies were performed assuming less than 50 pile-up interactions per bunch crossing. The TileCal signal reconstruction algorithms employed to extract the pulse amplitude from the digitized samples degrade in performance with increasing pile-up. This motivates the necessity of developing new light-weight algorithms that can be executed in the PPr to reconstruct the pulse amplitude and time with high precision in high pile-up conditions.

3.3 Reconstruction of the Tile Calorimeter signal from Run 1 to Run 3

In the current system, once an event is accepted by the trigger system, seven digital samples surrounding the selected bunch crossing are read out. This information can be used in order to reconstruct the amplitude and time of the original pulse, thanks to the prior knowledge of the pulse shape. The amplitude of the pulse is proportional to the energy deposited in the cell after the subtraction of the pedestal, and therefore a precise reconstruction of the amplitude is crucial for the energy measurement. The time or phase is defined as the time difference between the peak of the pulse and the expected arrival of the pulse.

There are two kinds of noise sources that can affect the signal reconstruction. The first one is the electronic or thermal noise, intrinsic to the electronics and that can make the samples fluctuate around their expected value. The electronic noise is larger in the high gain than in the low gain, and it can be characterized by measuring the RMS of the readout samples in pedestal events. In the context of the TileCal readout, a sample is defined as the discrete time-sampled digitized values of the analog pulse produced by the PMT and shaped and amplified by the front-end cards. A pedestal event is an event where no charge is injected in the PMTs, and the readout samples are only affected by the electronic noise.

The second noise source is the so-called pile-up noise, where two different components contribute to the distortion of the physics signal: the in-time pile-up noise, caused by the presence of other proton-proton interactions in the same bunch crossing leading to an overlay of signals, and the out-of-time pile-up noise, caused by the overlay of the signals from other bunch crossings, but that can contribute to the readout samples since the pulse shape is broader than the bunch spacing.

This kind of noise is not independent of the LHC conditions, it scales with the average number of interactions per bunch crossing, and consequently with the instantaneous luminosity [138]. An illustration of a pulse affected by out-of-time pile-up noise is shown in Figure 3.3.

The signal reconstruction in the TileCal is currently performed with a digital filter called Optimal Filtering (OF) [139]. The algorithm aims to reconstruct the amplitude and time of the pulse (and in some variants, the pedestal) performing a linear combination of the digitized samples. The weights used in the linear combination are obtained by minimizing the variance of the reconstructed magnitudes caused by the noise contributions, and taking into account the pulse shape. The algorithm is also very simple, making it suitable for real-time processing in the ROD DSPs. The reconstructed quantities can be written as:

$$A = \sum_{i=1}^{N} a_i S_i, \qquad A\tau = \sum_{i=1}^{N} b_i S_i, \qquad p = \sum_{i=1}^{N} c_i S_i, \tag{3.1}$$

where A is the amplitude, τ is the time or phase, p is the pedestal, S_i are the digitized samples, and a_i, b_i, and c_i are the weights calculated using the pulse shape and the noise covariance matrix to minimize the variance of the reconstructed quantities due to noise. These weights are computed separately for the high and low gain (since the noise differs between the two gains) and are stored in the TileCal database to enable online signal reconstruction in the ROD DSPs and offline event reconstruction. The weights depend on the pulse shape and the offset

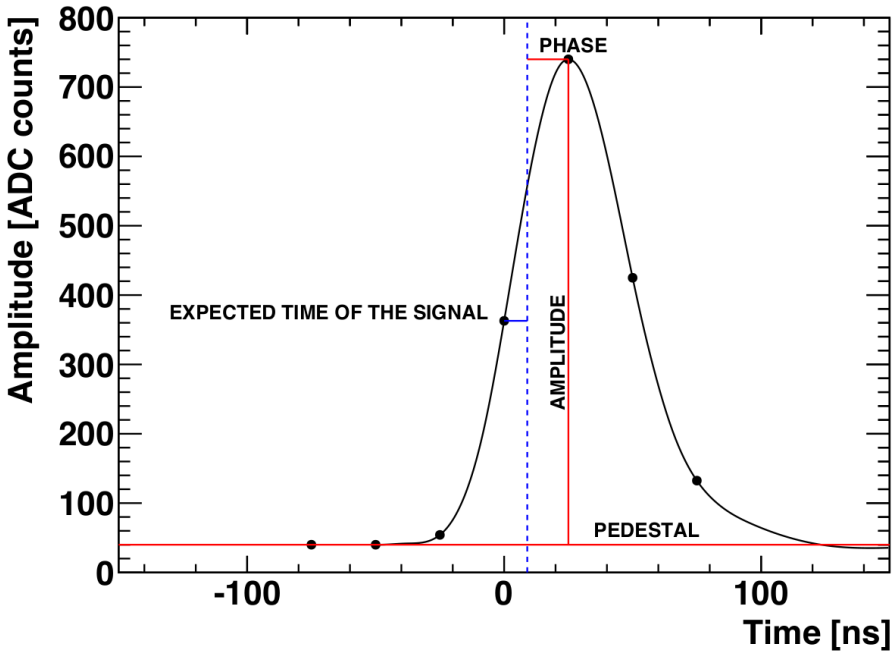


Figure 3.2: Pulse in a physics event indicating the definition of the amplitude, phase and pedestal.

between the true peak of the pulse and the LHC clock. This offset is approximately known in physics collisions, making it possible to retrieve the correct set of weights from the database for signal reconstruction. When the optimal filtering algorithm is applied just once, it is referred to as non-iterative optimal filtering. However, in cases where the offset is not known (e.g. cosmic muon signals), the optimal filtering algorithm is applied iteratively. In the iterative approach, the largest sample is initially chosen as an ansatz for the phase, and the corresponding weights are used to reconstruct the amplitude and time. The reconstructed time is then used to retrieve a new set of weights, and a new set of reconstructed quantities is obtained. This process is repeated until the values converge to a stable solution. This iterative method is slower and is used only for low-rate data taking.

In calibration runs and test beam campaigns, where latency is not an issue, the optimal filtering algorithm is replaced by a more complex technique called the fit method. In this approach, the pulse shape is fitted to the samples using a χ^2 minimization, from which the amplitude, time, and pedestal are derived. While the fit method is more precise than optimal filtering, it is also more computationally intensive, making it unsuitable for real-time processing in the ROD DSPs.

In the HL-LHC conditions, the distortion of the pulses by the pile-up noise will significantly increase, and degrade the performance of the OF algorithm if it is used without modifications. As pointed out before, the signal reconstruction performance will be even more crucial in the HL-LHC phase, since it will impact the trigger decisions. This motivates the development of new signal reconstruction

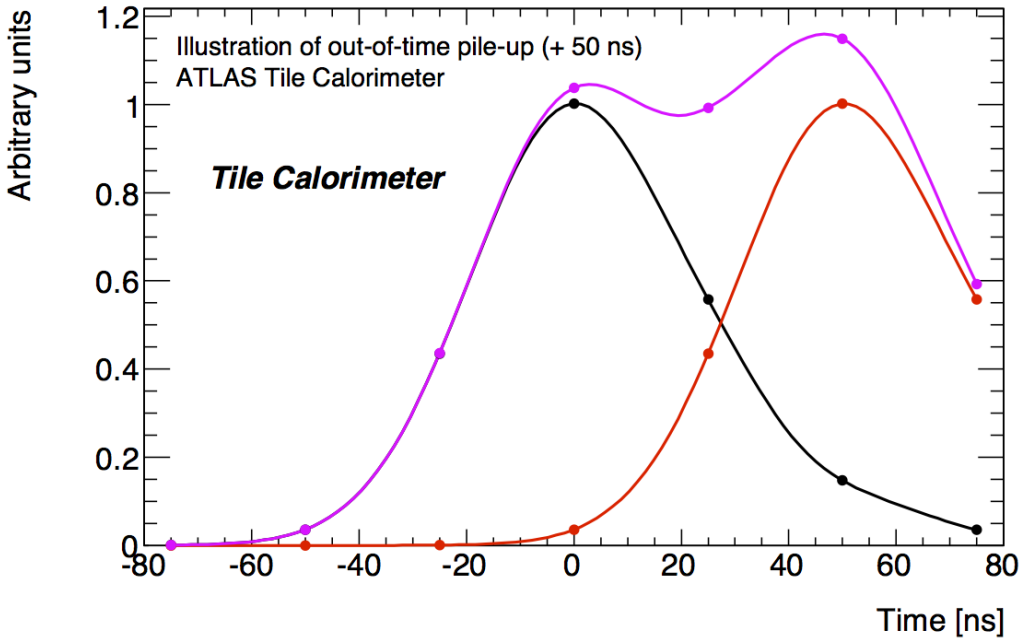


Figure 3.3: Illustration of a pulse affected by out-of-time pile-up noise. The in-time pulse is shown in black, while the out-of-time signal is shown in red. The sum of the two pulses is shown in purple, and the purple dots represent the value of the samples that are read out [113].

algorithms that can be executed in the PPr, and that can provide a high precision in the amplitude and time reconstruction in high pile-up conditions. To do that, it is necessary to be able to simulate the signals in the TileCal channels to be able to reproduce the digitized samples with realistic parameters and noise conditions, having control over the input signals injected in the TileCal digitization chain.

3.4 The Pulse Simulator tool

The Pulse Simulator is a tool integrated into Athena [140–142], the ATLAS offline reconstruction software framework, that allows to simulate physics signals with different parameters and noise conditions and inject them into the TileCal digitization chain in order to simulate the output samples of the readout electronics. This allows studying the performance of different signal reconstruction algorithms under controlled input signals and conditions. With the Pulse Simulator, it is possible to evaluate the impact on the signal reconstruction of various conditions or parameters, such as electronic noise, in-time and out-of-time pile-up, pulse shape changes, phase shifts, and other factors.

The Pulse Simulator was upgraded in order to realistically simulate the signals in the TileCal channels in the HL-LHC conditions, using measurements of the TileCal upgraded electronics from test beams and laboratory data, or simulations of the TileCal response with the HL-LHC conditions. A new version of the Pulse Simulator was developed, and it is currently being used to study the performance of the signal reconstruction algorithms in the TileCal in the HL-LHC conditions. The different updates in the Pulse Simulator are described in the following sections. After the updates, the Pulse Simulator tool has the following configurable options:

- Amplitude of the in-time signal. The amplitude can be set to a fixed value or determined by a user-defined distribution.
- Phase of the in-time pulse with respect to the expected arrival time. The phase, relative to the expected arrival time, can be either fixed or randomly selected from a user-defined distribution.
- Amplitude of an out-of-time signal. This amplitude can also be fixed or randomly sampled from a user-defined distribution.
- Phase of the out-of-time pulse with respect to the expected arrival time. Similar to the in-time pulse phase, this phase can be fixed or derived from a user-defined distribution relative to the expected arrival time.
- Gain switch. Both high and low gains can be recorded, or a gain switch can be simulated. In the latter case, the high gain is saved if the pulse is not saturated; otherwise, the low gain is saved.

- Electronic Noise Simulation. Noise can be simulated using either a single Gaussian distribution (suitable for HL-LHC electronics) or a double Gaussian distribution (for legacy electronics). For the single Gaussian, the user can define the noise RMS. For the double Gaussian, the user can define the RMS for both Gaussians and their respective contributions. These parameters can also be sourced from a database to simulate noise for each individual TileCal channel.

- Pulse-shape imperfections. Imperfections in the pulse shape can be simulated, broadening the pulse distribution to test the robustness of the signal reconstruction algorithms.

- Pedestal configuration. The pedestal can be uniformly configured across all channels or set individually per channel using database values. Configuration is available for both high and low gain settings.

- Pulse shape. The pulse shape is user-configurable. For HL-LHC electronics, a long tail (on the order of microseconds) can be added to the simulation.

- Number of samples. The number of digitized samples is user-configurable.

- Bunch Spacing configuration. This can be set to simulate out-of-time pile-up accurately with different bunch-spacing options (e.g. 25 ns, 50 ns, 75 ns).

- Pile-up simulation. An average interaction level per bunch crossing can be configured to simulate both in-time and out-of-time pile-up. The out-of-time amplitude distribution for pile-up events can be loaded from a configurable file, allowing for different distributions per channel.

- Number of bunch crossings taken into account in the simulation. The user can configure the number of bunch crossings to account for when quantifying the effect of out-of-time pile-up on individual samples.

- Users can choose to simulate the legacy paradigm (where a set number of samples around the triggered bunch crossing are read out) or the new paradigm, where continuous sample readout allows for online reconstruction of pulse amplitude and time per bunch crossing before the Level 0 trigger decision.

3.5 Updates in the Pulse Simulator for the Phase II upgrade

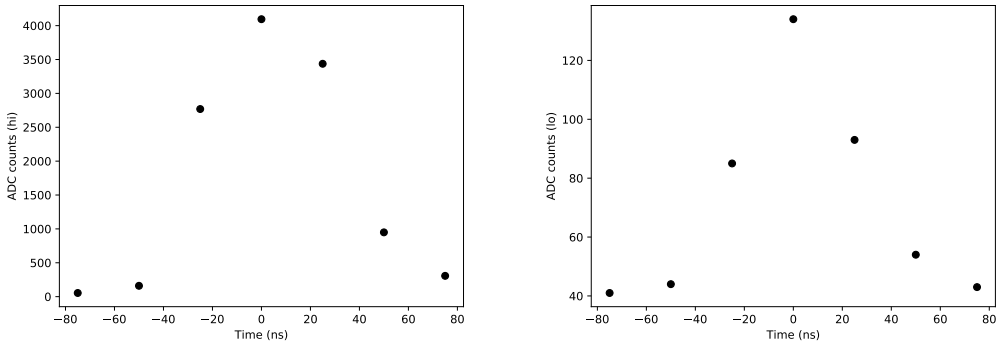
The following sections describe the updates to the Pulse Simulator required for simulating signals in the TileCal channels under the HL-LHC conditions.

3.5.1 Bit range and gains

In the legacy TileCal electronics, the original system currently used for signal processing before the planned upgrades for the High Luminosity LHC, the shaped pulses are amplified in separate high and low gain paths on the front-end cards with a gain ratio of 64:1. The two signals are digitized by two 10-bit ADCs. The electronics can be operated in a bi-gain mode, where both gains are sent to the off-detector electronics and later saved, or in a single-gain mode, where the high gain is saved if the pulse is not saturated and if the measured amplitude is above a certain threshold (i.e. if all the samples are below 1023 counts, but none of the samples is below 2 ADC counts), and the low gain is saved otherwise. If the system is operated in single-gain mode, up to 16 samples can be saved for each event accepted by the trigger system.

In the HL-LHC electronics, the pulses are amplified in two paths with a gain ratio of approximately 40:1, and the signals are digitized by two 12-bit ADCs. The Pulse Simulator was updated to be able to configure 12-bit ADCs for the digitization, and to be able to select if the system should be operated in bi-gain or single-gain mode. An example of the output of the Pulse Simulator with 12-bit ADCs is shown in Figure 3.4, where a pulse saturates the high gain. The data in the two gains is saved in the output file, since the tool was configured to operate in bi-gain mode.

In the FENICS card, the output of the shaper is separated in two branches. In one of them, the pulse is reduced with a passive voltage divider with a factor 0.4. In the other branch, an operational amplifier is used to amplify the signal by a factor of 16. This gives a gain ratio of around 40, allowing to measure signals up to 1020 pC with a pedestal value of 100 ADC counts. The Pulse Simulator was updated to be able to use the gain ratio of 40:1 if the system is operated in the HL-LHC mode. Figure 3.5 shows the ratio of the reconstructed amplitude in the



(a) (b)

Figure 3.4: Pulse simulated with the Pulse Simulator using 12-bit ADCs and the HL-LHC electronics. (a) The pulse saturates the high gain, and the data is also saved in the (b) low gain, since the Pulse Simulator was configured to operate in bi-gain mode.

high gain to the reconstructed amplitude in the low gain for a pulse of 800 ADC counts in the high gain. The ratio is around 40, as expected. The small deviations are due to the pulse amplitude reconstruction by the optimal filtering algorithm and the electronic noise introduced in the simulation.

3.5.2 Number of readout samples

The standard readout window of the legacy electronics contains 7 readout samples around the triggered bunch crossing. However, this value is not necessarily fixed. The electronics can be configured to send up to 16 readout samples to the off-detector electronics (if the system is operated in a single-gain mode). The upgraded electronics will also be able to send more than 7 readout samples to the off-detector electronics, and further studies are needed to determine the optimal number of samples to be sent. The Pulse Simulator was updated to be able to configure the number of samples in the simulation. The optimal filtering coefficients that are stored in the TileCal database are calculated for 7 readout samples, but the Pulse Simulator can calculate the optimal filtering coefficients for any number of samples on-the-fly assuming a noise auto-correlation matrix that is diagonal (i.e. the noise in the samples is uncorrelated, due to the absence of pile-up signals). The user can therefore simulate signals with any number of samples, and the Pulse Simulator will provide the reconstructed amplitude and time using the optimal filtering algorithm. Examples of pulses simulated with different numbers of samples are shown in Figure 3.6.

The Pulse Simulator allows the simulation of an arbitrary large number of samples. Examples of the amplitude, time and pedestal reconstructed with the

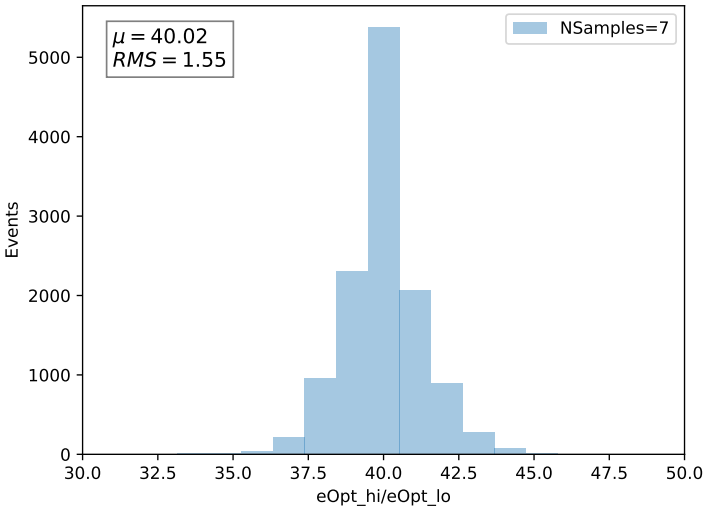
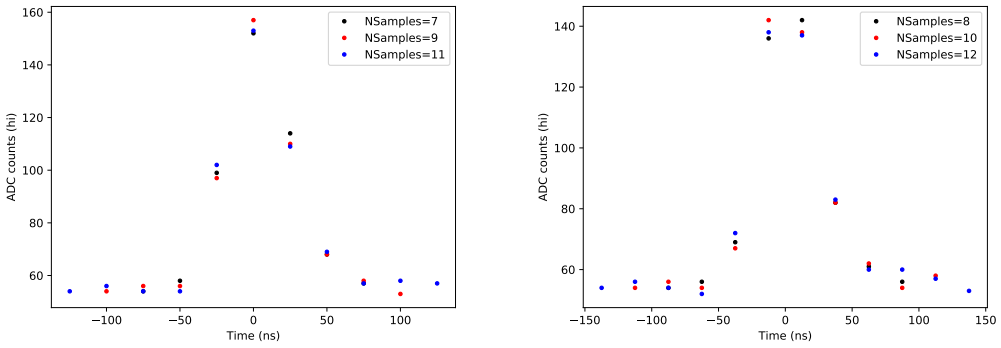


Figure 3.5: Ratio of the reconstructed amplitude in the high gain to the reconstructed amplitude in the low gain using the optimal filtering algorithm on a pulse simulated with the Pulse Simulator using the gain configuration of the HL-LHC electronics.



(a) (b)

Figure 3.6: Pulses with an amplitude of 100 ADC counts simulated with (a) different odd number of readout samples and (b) different even number of readout samples.

optimal filtering algorithm for high gain pulses of 500 ADC counts and a pedestal of 50 ADC are shown in Figure 3.7 for 31 and 63 readout samples. In each of the cases the reconstructed values are compared with the ones obtained with 7 readout samples.

3.5.3 Pedestal and noise values

In the HL-LHC electronics, the pedestal in the fast readout can be individually adjusted for each channel and gain, thanks to the presence of dedicated Digital-to-Analog Converters (DACs) in the Mainboards that are specifically designed to set the baseline for each ADC. The Pulse Simulator was updated to be able to configure the pedestal independently for the high and low gain. The pedestal can be set to a fixed value, or it can be taken from a database that contains the pedestal values for each channel.

During the operation of TileCal in the HL-LHC, the charge values that will be read out will range from 200 fC to 1000 pC. This fact imposes requirements on the noise levels that are acceptable in the system, especially in order to measure low energy deposits with the required precision and guarantee meaningful measurements. There are different ways to define the noise in the system. The electronic noise in TileCal has been measured traditionally in terms of two components: the high frequency noise, that can be measured as the RMS of the samples from pedestal events when no charge is injected in the PMTs, and the low frequency noise, that can be measured as the RMS of one of the samples (typically the first one) used to reconstruct the pulse amplitude. The high frequency noise in the FENICS boards follows a Gaussian distribution with a standard deviation of 2.5 ADC counts for the high gain and 1.0 ADC counts for the low gain. These values correspond to 16 fC and 260 fC, respectively. For an energy deposit of 200 fC, the signal in the high gain is 25 times larger than the noise. The high gain saturates at around 25 pC, which is 25 times the noise level in the low gain. This means that if a noise level of

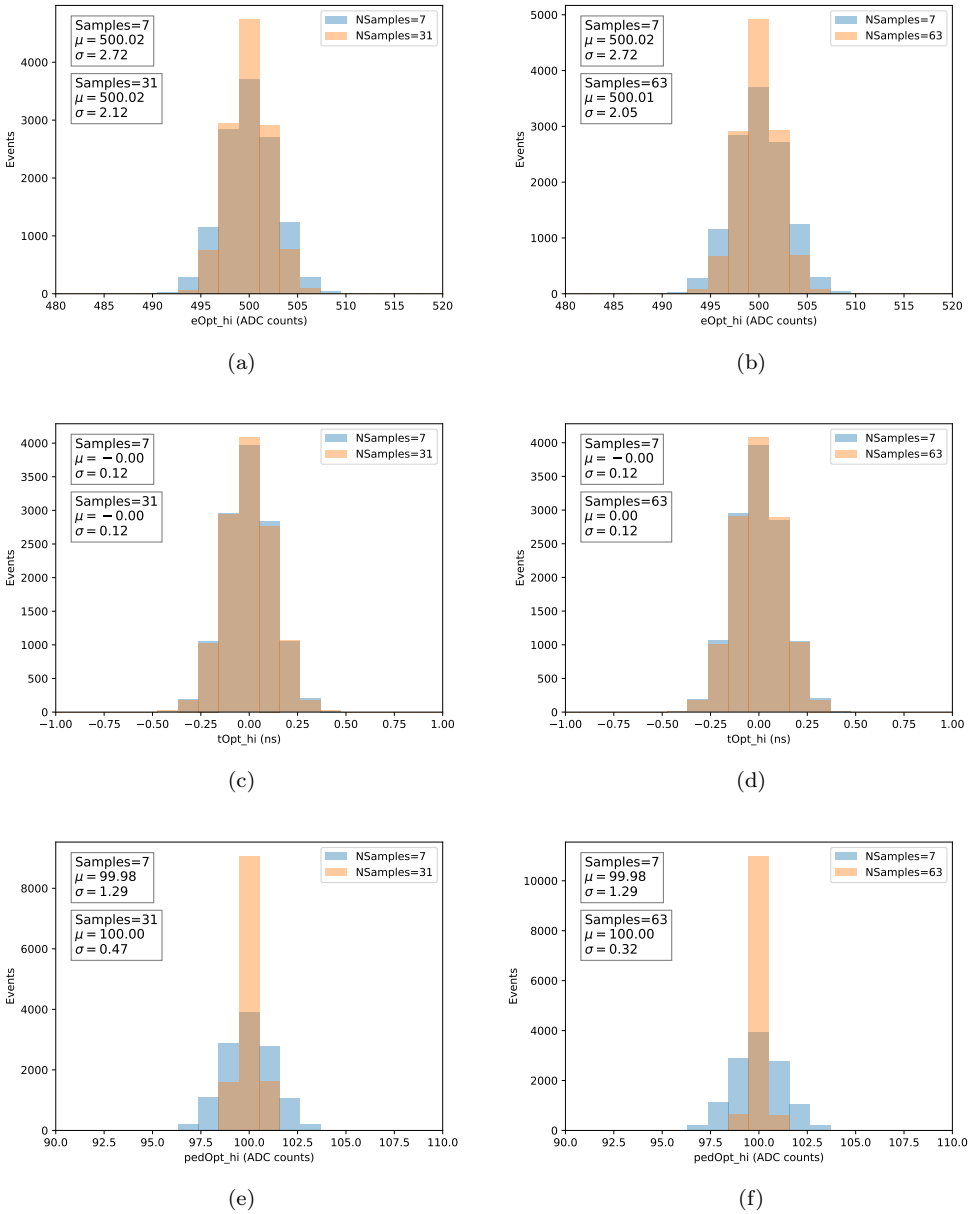


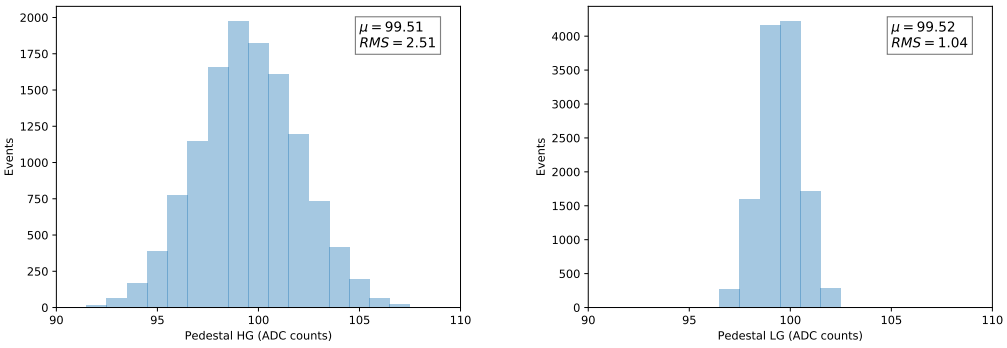
Figure 3.7: Amplitude (a, b), time (c, d) and pedestal (e, f) reconstructed with the optimal filtering algorithm for high gain pulses of 500 ADC counts and a pedestal of 50 ADC counts with 31 (a, c, e) and 63 (b, d, f) readout samples.

5% is considered acceptable in the low gain, one can measure with the low gain charges as low as 5 pC, having a considerably large overlap with the high gain.

The Pulse Simulator is able to introduced high frequency noise in the samples, and the user can configure the RMS of the noise. By default, if the user does not configure the noise in the Pulse Simulator, the tool will use the values of 2.5 ADC counts for the high gain and 1.0 ADC counts for the low gain if the HL-LHC mode is selected. An example of the distribution of the first sample in the readout window for a pedestal of 99.5 ADC counts using events simulated with the Pulse Simulator is shown in Figure 3.8. The distribution is shown for high and low gain, and the RMS of the distributions shows the typical values of the electronic noise in the HL-LHC electronics.

3.5.4 Pulse shape

To accurately simulate the signals in the TileCal channels, a precise description of the pulse shape after amplification and shaping in the front-end electronics is required. The Pulse Simulator was initially developed using the pulse shape characteristic of the legacy electronics, derived from test beam data collected before ATLAS began operations. However, the upgrade of the electronics includes the replacement of the legacy 3-in-1 cards by the FENICS boards (which feature a different shaping circuit). The pulse shape is different, and therefore it is necessary to have a new description of the pulse shape, both to be able to simulate the signals in the TileCal channels but also to develop and implement signal reconstruction algorithms that rely on the knowledge of the pulse shape. The following sections describe the studies that were performed to obtain the pulse shape of the TileCal upgraded electronics using data from test beam campaigns and laboratory measurements. The modifications that were done to the Pulse Simulator to be able to simulate the signals in the TileCal channels with the new pulse shape are also described.



(a) (b)

Figure 3.8: Pedestal computed as the mean of the first sample in the readout window for high (a) and low (b) gain using events from the Pulse Simulator.

3.5.4.1 The TileCal test beam experimental setup

The pulse shape measurements discussed in the next sections were performed using spare modules of the ATLAS TileCal detector located in the H8 beam line in the North Area of CERN. The modules were exposed to beams of different particles, and the signals were read out using the upgraded electronics.

In the line, beams of 400 GeV protons from the SPS are made to collide with a beryllium target to produce a secondary beam line. Colliding this secondary beam with different types of targets allows producing different types of particle beams. The measurements presented here were performed using data from electron and hadron beams. To produce hadrons, the secondary beam hits a 300 mm long copper target, and a 6 mm thick lead target positioned 270 meters downstream that absorbs the electrons. For the electron beams, the secondary beam collides with a 400 mm aluminum target immediately followed by a lead absorber.

Before reaching the TileCal, the particles pass through several detectors that are positioned in the beam line for triggering and particle identification. These beam line detectors include a wire chamber to monitor the beam profile, two scintillating counters that are used to trigger the data acquisition system through a coincidence signal and three Cherenkov detectors that are used to distinguish between electrons and pions, kaons and protons.

The test beam setup includes three spare ATLAS detector modules: two long-barrel modules and one extended-barrel module. These modules are placed on a scanning table capable of adjusting their positions and angles relative to the incoming beam particles. The extended-barrel module consists of one drawer, while each long-barrel module comprises two drawers. In the setup, the modules are labeled as M0 A and M0 C (bottom), LBA65 and LBC65 (middle), and EBC65 (top).

The drawers in modules LBA65 and LBC65 are equipped with upgraded front-end electronics systems proposed for the HL-LHC. In contrast, the super-drawers in modules EBC65 and M0 C use the electronics currently installed in the ATLAS detector. The M0 A module is equipped with multianode PMTs for research and development studies. This setup allowed for the evaluation of the new

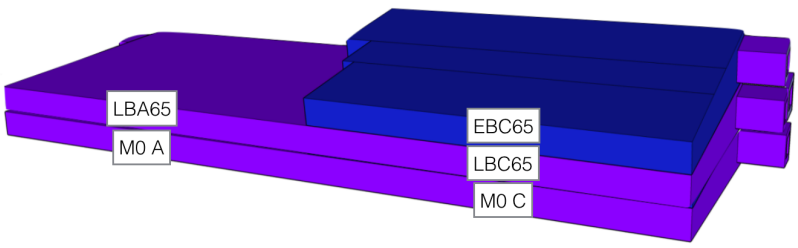
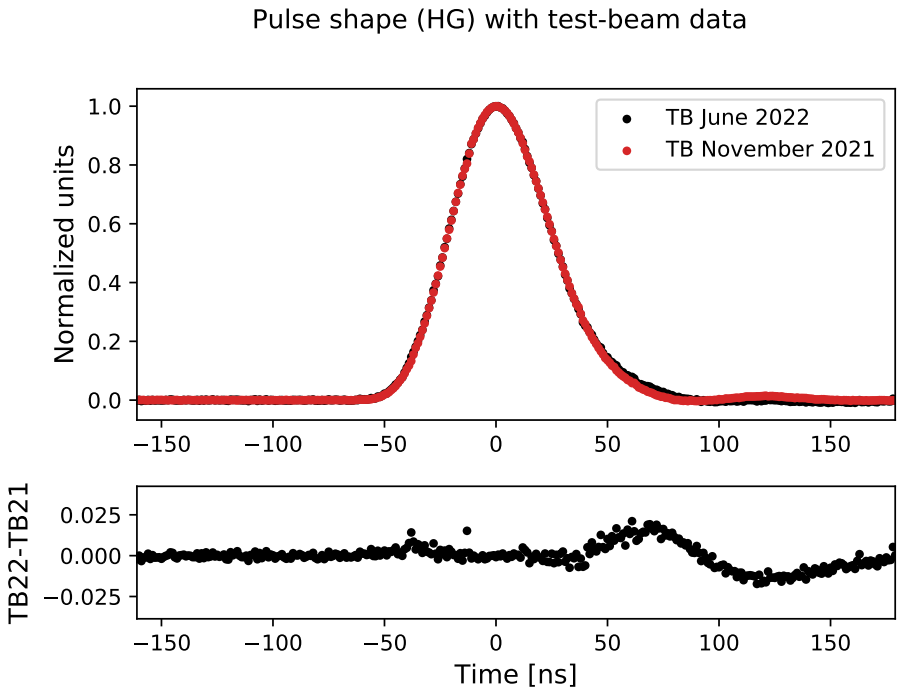


Figure 3.9: Schematic of the test beam setup showing the three modules used in the test beam measurements.

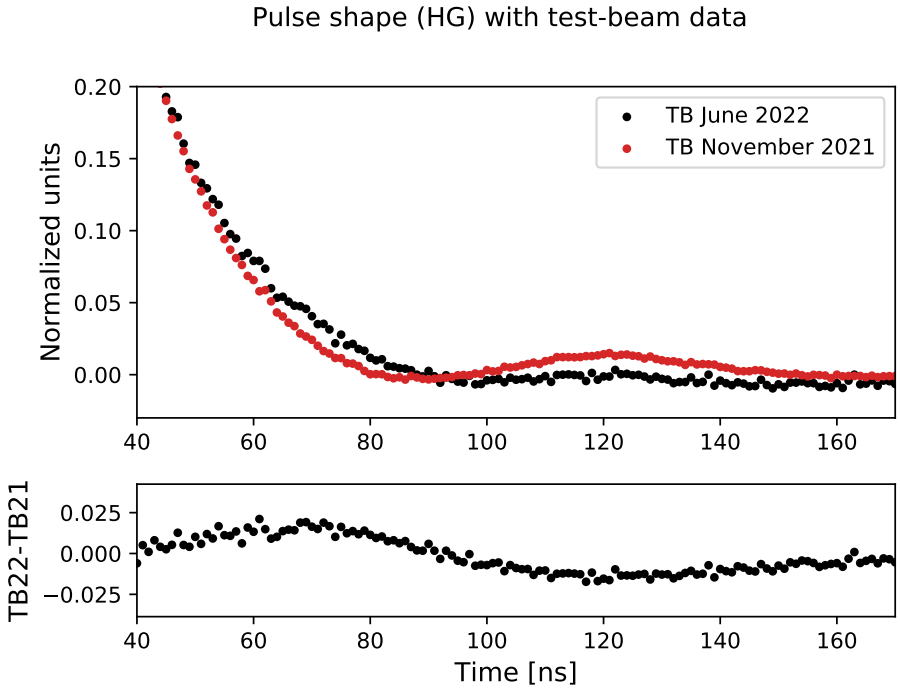
electronics in comparison with the existing systems. A schematic of the test beam setup is shown in Figure 3.9.

3.5.4.2 Pulse shape measurement with test beam data

In the test beam setup, the data acquisition clock is not perfectly synchronized with the particle arrival time, even though the trigger chambers positioned upstream of the TileCal module are used to emit trigger signals when particles are detected, allowing for data collection. However, unlike in the ATLAS setup, where the LHC clock tightly controls the timing and ensures that the readout window is centered around the expected arrival time of particles, the analog pulse in the test beam setup is not always sampled at the same time position. This results in the readout samples capturing different points along the full pulse shape, offering a complete view of the pulse's evolution. In contrast, in ATLAS, the synchronization if the detector with the LHC clock ensures that the pulse is consistently sampled at its peak, assuming proper timing of the channels, limiting the sampling of the pulse shape to a few concrete points.



(a)



(b)

Figure 3.10: Pulse shape comparison between the data from the November 2021 and June 2022 test beam campaigns. (a) Full pulse shape. (b) Zoom-in of the pulse shape in the right tail.

Data from electron beams at different energies were used to estimate the pulse shape in the TileCal upgraded electronics. The data was taken during test

beam campaigns in November 2021 and June 2022. The data was taken storing 15 readout samples for each event to be able to study the pulse shape in a wide time window. The upgraded electronics are capable of reading and processing more than 16 readout samples, but the test beam setup limits the number of samples. The CPMs of the upgraded electronics send the data to the legacy RODs, which can read only up to 16 samples per event. The dataset used to obtain these pulse shape measurements uses the information from the ROD, which was configured to store 15 samples when the data was taken.

The amplitude of the signals in all the channels was reconstructed using the fit method, assuming the pulse shape of the legacy electronics. The pulse shape was obtained for each channel instrumented with the upgraded electronics. To do that, for each of the channels, events in which the amplitude of the signal was above 100 MeV were selected, and the time of the peak obtained from the fit was used to shift the measured samples by that time and center the pulse with respect to a reference time. The pedestal was subtracted from the samples in each event using the average ADC counts in the samples before the arrival of the pulse. The shifted samples were normalized to the reconstructed amplitude to be able to combine the samples from different events and channels. The measurements from the different channels were combined to obtain the average pulse shape in the TileCal upgraded electronics. The study was performed originally using the data from the November 2021 test beam campaign. The obtained pulse shape had

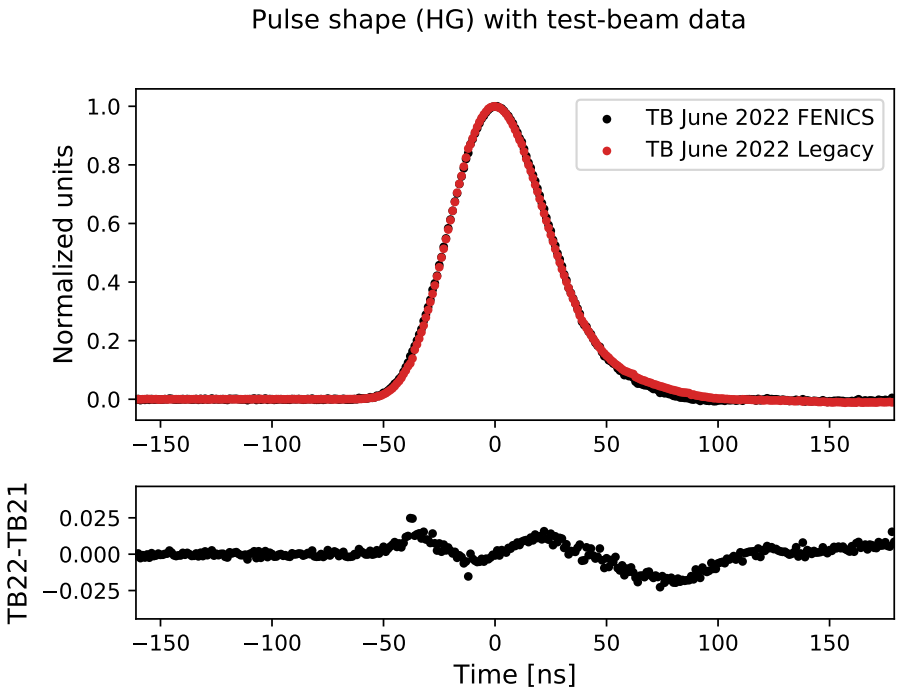


Figure 3.11: Comparison of the pulse shape obtained with the data from the June 2022 test beam campaign with the pulse shape obtained with the legacy electronics.

an overshoot in the right tail of the pulse, visible in 3.10. This overshoot was not present in the pulse shape of the legacy electronics and would have complicated the design of the signal reconstruction algorithms. Modifications were performed in order to minimize the overshoot, and the pulse shape was measured again using the data from the June 2022 test beam campaign. The comparison of the pulse shapes obtained in the two test beam campaigns is shown in Figure 3.10.

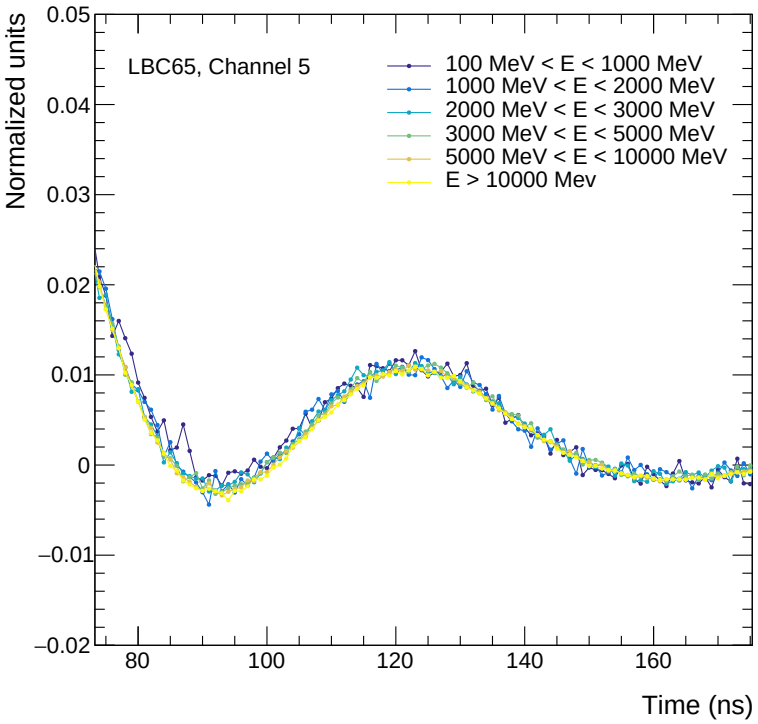
A comparison of the pulse shape obtained with the data from the June 2022 test beam campaign with the pulse shape obtained with the legacy electronics is shown in Figure 3.11. The differences are small, and the final pulse shape is considered to be suitable for the design of the signal reconstruction algorithms. The difference between the pulse shapes in high and low gain were considered negligible, since their difference was smaller than the channel-to-channel differences due to the tolerances in the electronic components.

The pulse shape was also studied as a function of the measured amplitude using the data from the November 2021 test beam campaign. No evidence of a dependence of the pulse shape on the amplitude was found in the range between 100 MeV and 150 GeV. This can be seen in Figures 3.12, where the pulse shape for different measured energies is shown with a zoomed-in view of the right tail of the pulse. The pulse shape is shown for high and low gain. The legacy pulse shape is known to have a small energy dependence, introducing a bias below the percent level in the amplitude reconstruction if an energy-independent pulse shape is used [143]. The absence of an energy dependence in the pulse shape of the upgraded electronics simplifies the design of the signal reconstruction algorithms.

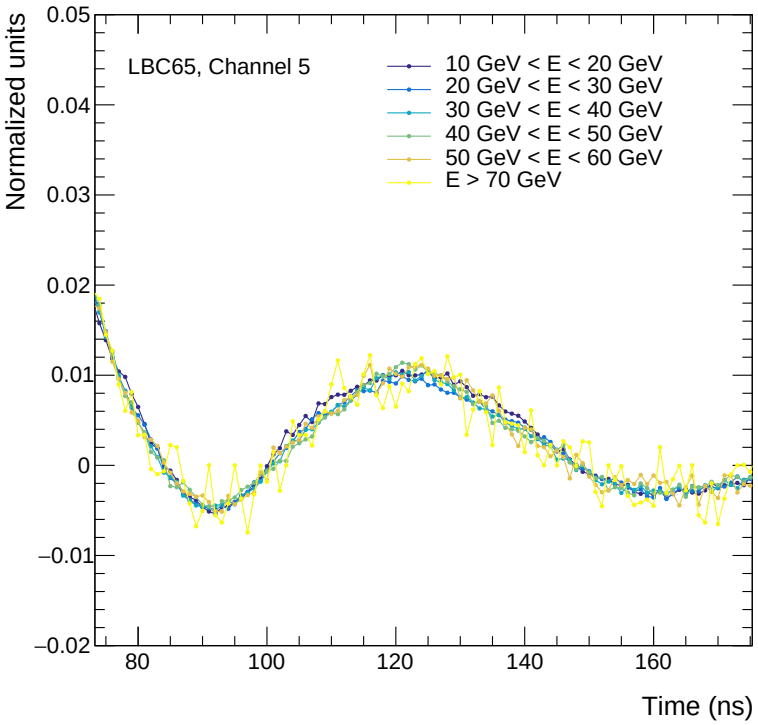
To be able to use the pulse shape for signal reconstruction in following test beam campaigns and simulations of the TileCal response in the HL-LHC conditions, a smoothing algorithm was applied to the measured pulse shape.

3.5.4.3 The pulse shape undershoot using laboratory data

The pulse shape obtained from test beam data was compared with that derived from laboratory measurements performed using the FENICS boards. The laboratory measurements were conducted using the charge injection system integrated into the TileCal electronics. This system allows the injection of a known charge into the readout chain through 220 pF and 5.2 pF capacitors. The amount of charge injected into the circuit can be configured to cover the entire range of charges measurable by the TileCal. While the pulse shape measured with the charge injection system is narrower than the pulse shape generated by physics signals under actual detector conditions, both shapes are consistent in their tails, exhibiting a long tail of approximately microseconds.

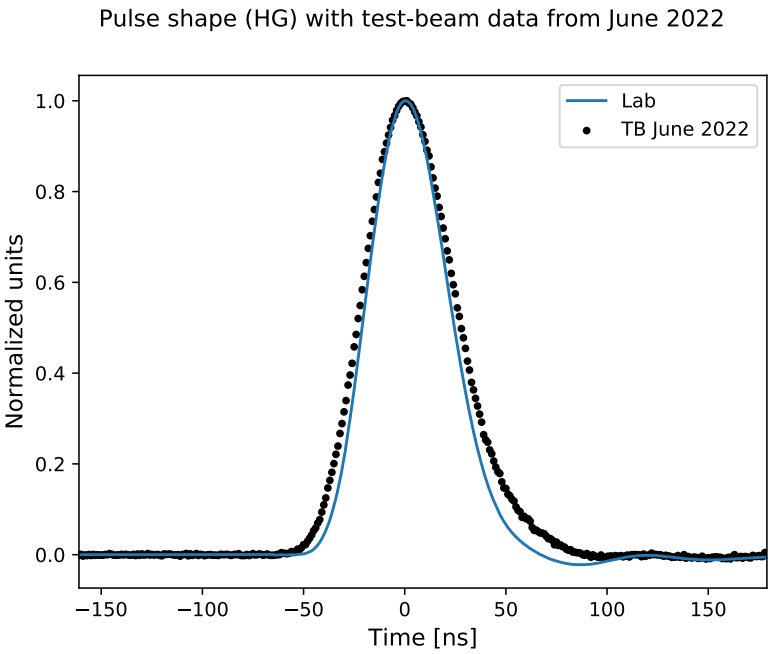


(a)

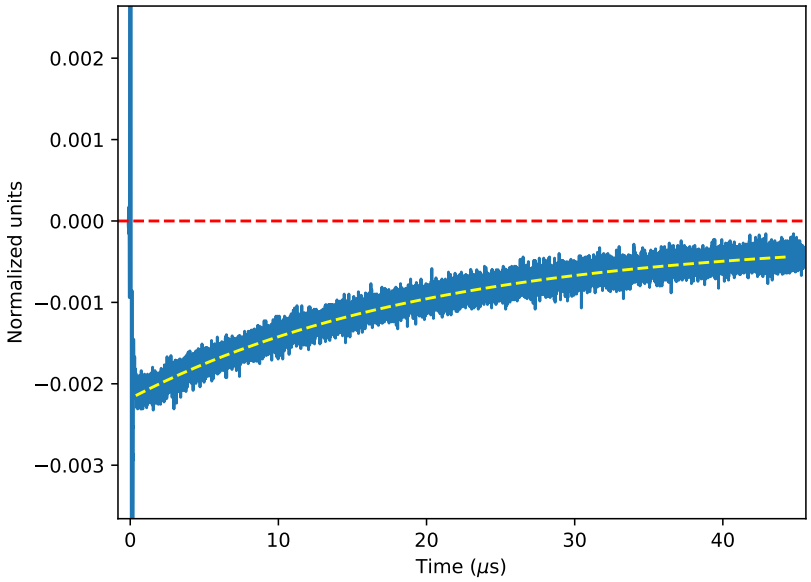


(b)

Figure 3.12: Pulse shape obtained with the data from the November 2021 test beam campaign for different measured energies for (a) high gain and (b) low gain.



(a)



(b)

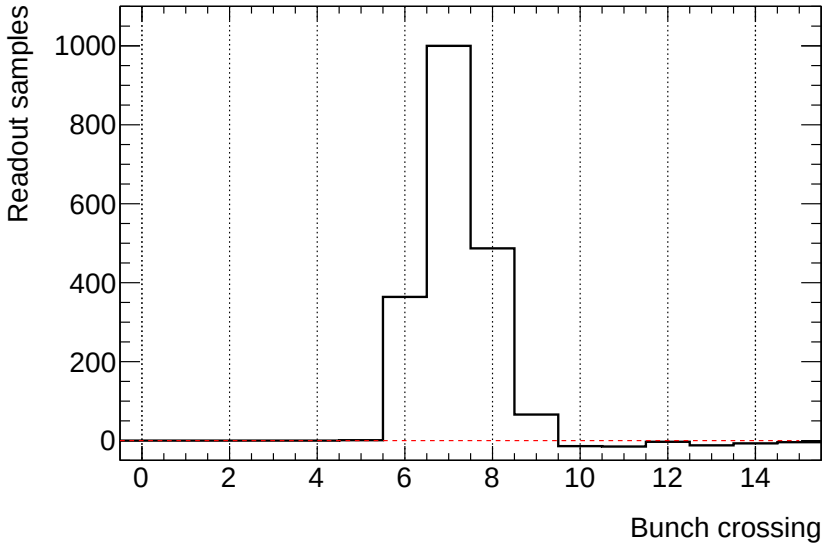
Figure 3.13: (a) Comparison of the pulse shape obtained with the test beam data and the pulse shape obtained with the laboratory data using the charge injection system. (b) Pulse shape tail extending across several microseconds. The fit to the pulse tail is shown in yellow. A red line is drawn at zero to indicate the pedestal level.

The laboratory setup permits a detailed scan of the pulse shape over a range of microseconds, whereas test beam data measurements are limited to the range of hundreds of nanoseconds due to constraints on the number of samples that can be recorded. Figure 3.13a presents a comparison between the pulse shapes obtained

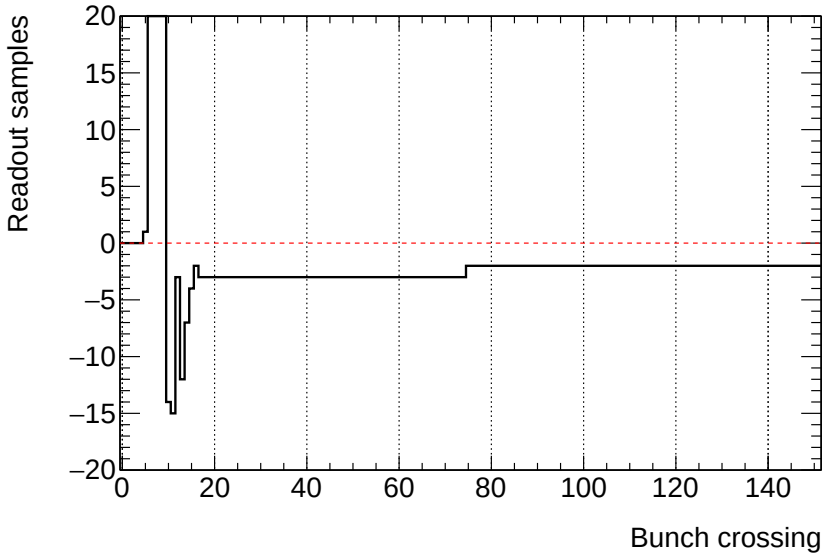
from test beam data and those from laboratory measurements performed using the charge injection system.

Measuring the long tails of the pulse shape is essential for understanding how these tails influence the digitized samples several bunch crossings after the pulse is generated. Although the effect of these tails is small, their extended duration means that contributions from multiple bunch crossings can accumulate and significantly impact the digitized samples. In the context of the HL-LHC, where pile-up noise will be a major concern, it is essential to account for the effect of the long undershoot in simulations. Due to size constraints, the full pulse shape information over several microseconds cannot be stored in the TileCal database. However, the pulse undershoot can be effectively parametrized using a simple functional form and incorporated into the simulation. Figure 3.13b illustrates the pulse tail over several microseconds, along with its corresponding fit. The best fit to the pulse tail was obtained using the function $f(t) = A\left(1 - e^{-t/\tau}\right) + b$, where $A = 0.002(0.002)$, $\tau = 20.462(20.465)\mu$s and $b = -0.002(-0.002)$ for the high (low) gain pulses. The differences observed between the high and low gain pulses were smaller than the differences observed by varying by a few nanoseconds the range of the fit, and therefore were considered negligible for the purpose of the simulation.

The Pulse Simulator was updated to include the possibility of simulating data using the pulse shape of the upgraded electronics and the pulse undershoot. An example of a pulse with these two features is shown in Figure 3.14. The simulation was run assuming no electronic noise to show explicitly the pulse shape produced by the FENICS boards without distortions introduced by the noise (the electronic noise can be greater than the undershoot effect for a single pulse). Since the output of the simulation are the readout samples (integer values), the tail of the pulse is discretized, and the undershoot is not as smooth as the one shown in Figure 3.13b.



(a)



(b)

Figure 3.14: Example of a simulated pulse with an amplitude of 1000 ADC counts using the pulse shape of the upgraded electronics and the pulse undershoot. The two plots correspond to the same pulse, with (a) showing the main part of the pulse, while (b) shows the tail across 150 bunch crossings. The simulation was run without electronic noise to better show the pulse shape produced by the FENICS boards. The pedestal is subtracted from the readout samples. A red line has been drawn at zero ADC counts to indicate the pedestal level and show the readout samples that are below the pedestal.

3.5.5 Pile-up simulation

There are various features in the Pulse Simulator related to the simulation of pile-up, a crucial aspect since pile-up noise will be a significant source of noise under HL-LHC conditions and will greatly impact the performance of signal reconstruction algorithms. Prior to the updates described in this thesis, the Pulse Simulator could inject both in-time and out-of-time pile-up signals with different bunch spacing options. For a bunch spacing of 25 ns, the simulation would inject pile-up signals in each of the 7 bunch crossings surrounding the triggered bunch crossing that were read out. This feature has been extended to simulate an arbitrary number of bunch crossings surrounding the triggered bunch crossing, ensuring compatibility with different readout windows and varying numbers of samples.

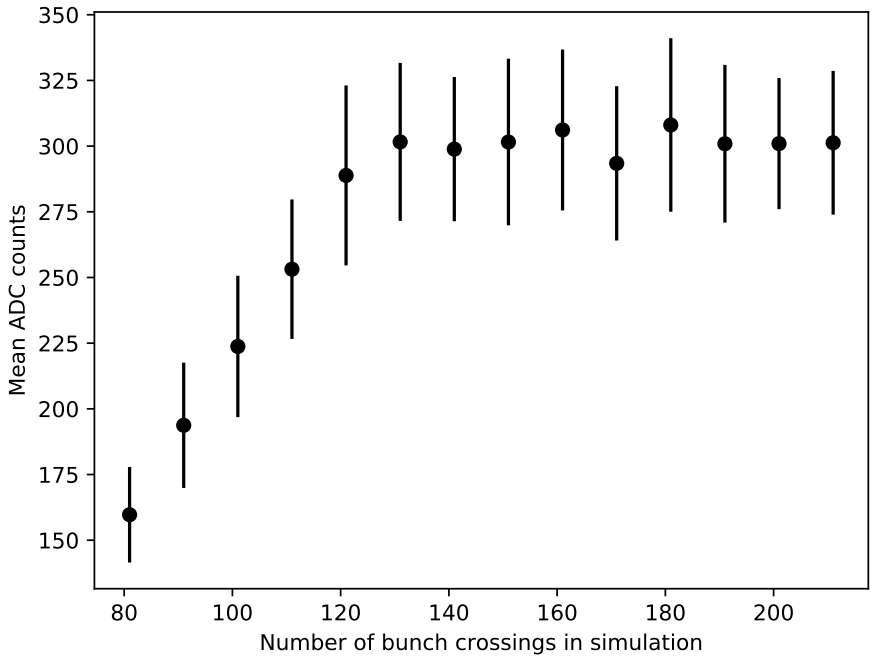


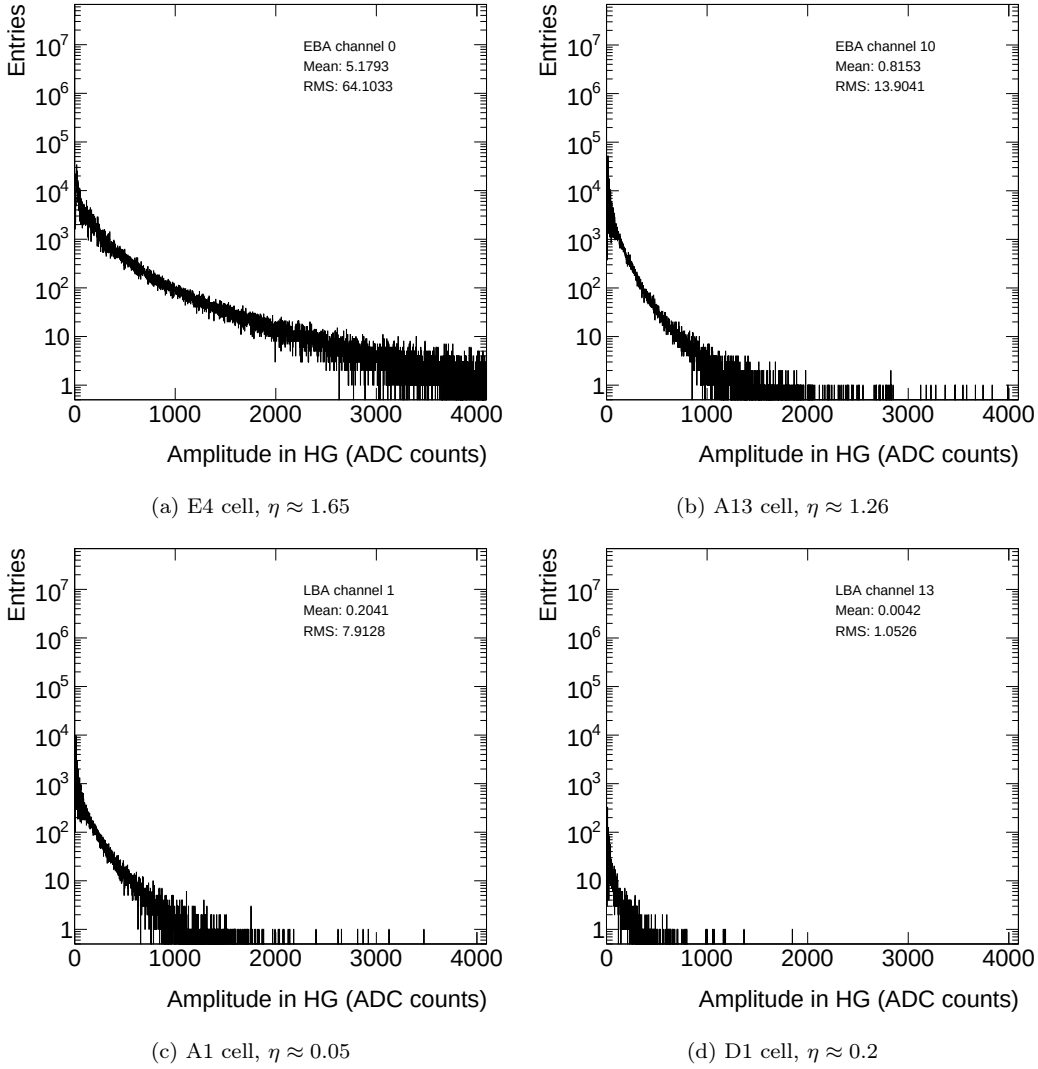
Figure 3.15: Mean ADC counts in events read from a channel connected to an A1 cell. The events do not contain an in-time pulse but are affected by out-of-time pile-up signals, with an average of 200 *pp* interactions per bunch crossing. The electronics pedestal was set to 100 ADC counts. The error bars represent the standard deviation of the readout samples. The number of samples in the readout window was set to 99.

The simulation of the legacy electronics did not include a pulse shape where the amplitude becomes negative for a certain time. The pulse shape was assumed to be zero outside the time range where the pulse was measured, between -75 ns and 75 ns with respect to the triggered bunch crossing. The samples that are read out within this range are not only affected by the in-time signal and the pile-up signals produced in the bunch crossings covered by the readout window, but also by the pile-up signals (or any signal) surrounding the readout window. To account for this effect, the Pulse Simulator was simulating 21 bunch crossings around the

simulated signal and convolving the amplitudes of all the signals in these 21 bunch crossings to obtain the digitized samples within the readout window. However, once the number of samples became a configurable parameter, and furthermore, once the effect of the pulse shape was extended to the range of microseconds, the fixed number of 21 bunch crossings was insufficient to account for the effect of pile-up signals surrounding the readout window. The Pulse Simulator was updated to simulate an arbitrary number of bunch crossings surrounding the triggered bunch crossing and to account for as many bunch crossings as needed to simulate the effect of those signals in the digitized samples.

Figure 3.15 shows the mean of the ADC counts in the high gain for a channel connected to an A1 cell without an in-time pulse, a pedestal of 100 ADC counts, and out-of-time pile-up signals corresponding to an average of 200 *pp* interactions per bunch crossing, as a function of the number of bunch crossings taken into account in the simulation. The number of samples in the readout window was set to 99, and the number of bunch crossings taken into account in the simulation was varied from 81 to 201. The simulation was performed with the pulse shape of the legacy electronics, and no negative tail was included in the simulation. The pile-up signals shift the ADC counts towards higher values, but the shift is underestimated when the number of bunch crossings taken into account by the Pulse Simulator is small, since the pile-up outside the readout window is not accounted for. The standard deviation of the readout samples is also underestimated when the number of bunch crossings taken into account is small, since the effect of pile-up noise is not fully simulated. Once the number of bunch crossings taken into account is large enough and all the samples in the readout window are affected by the pile-up amplitudes convolved with the front-end electronics pulse shape, the mean and standard deviation of the readout ADC counts stabilize, and increasing the number of simulated bunch crossings out of the readout window does not affect the results.

With the original implementation of the Pulse Simulator, each pile-up signal was simulated with an amplitude randomly sampled from a distribution provided by the user. This meant that if the user wanted to simulate a specific pile-up level, they first needed to derive the distribution of the pile-up signals for each specific pile-up level (i.e., for each average number of interactions per bunch crossing). To simplify the simulation of pile-up signals and allow the user to directly simulate a specific pile-up level, the Pulse Simulator was updated to simulate arbitrary pile-up levels by sampling the amplitudes of the pile-up signals from the distribution of amplitudes of a single minimum bias event. First, a simulation of low-momentum inelastic collisions (minimum bias events) was used to derive probability distributions of the amplitudes of the pile-up signals in each of the TileCal channels. The impact of the pile-up signals on the digitized samples varies significantly for each cell, as it depends on the cell's position in rapidity and also on the layer of the cell (cells closer to the beam pipe are more affected by the pile-up signals). Some examples of the distributions of the amplitudes of the pile-up signals in the high gain for different cells are shown in Figure 3.16. The plots show the significant size of



(a) E4 cell, $\eta \approx 1.65$

(b) A13 cell, $\eta \approx 1.26$

(c) A1 cell, $\eta \approx 0.05$

(d) D1 cell, $\eta \approx 0.2$

Figure 3.16: Distributions of pile-up signal amplitudes in the high gain for various TileCal cells.

the pile-up signals for E4 and A13 cells in the extended barrel. The fluence of particles coming from minimum bias interactions significantly changes from cell to cell, as shown in Figure 3.17. The E cells suffer from the highest fluence of particles, followed by the cells in the A layer, closer to the beam pipe.

From the distributions of the amplitudes of the pile-up signals in the minimum bias events, the Pulse Simulator is able to sample the amplitudes of the pile-up signals for each channel in the simulation of the pile-up using an arbitrary number of interactions per bunch crossing. The user needs to provide the average number of pp interactions per bunch crossing, $< \mu >$, and the Pulse Simulator will randomly generate the number of interactions per bunch crossing following a Poisson distribution with mean $< \mu >$. For each of the interactions, the Pulse Simulator will sample the amplitudes of the pile-up signals from the distribution of amplitudes of the pile-up signals in the minimum bias events. The Pulse Simulator will sum the amplitudes of the pile-up signals in each bunch crossing to obtain

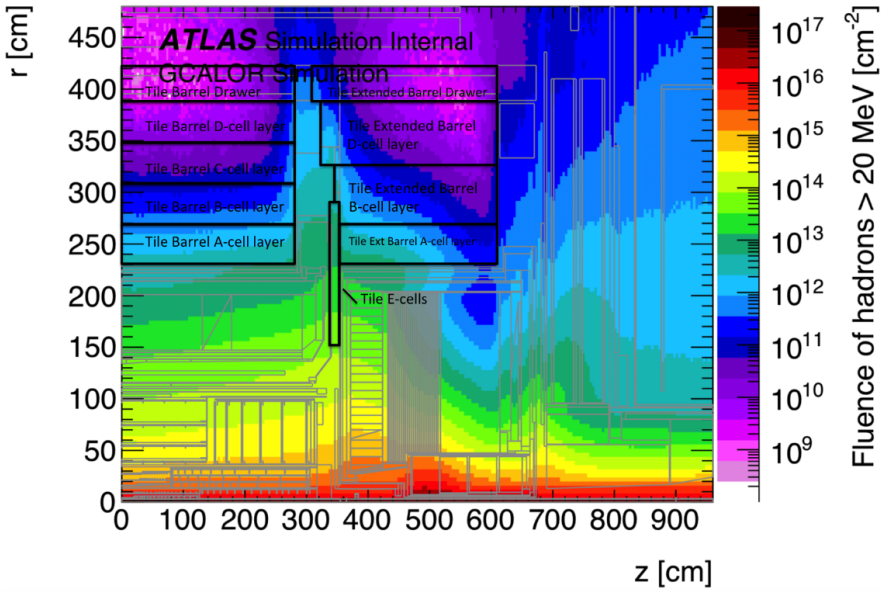
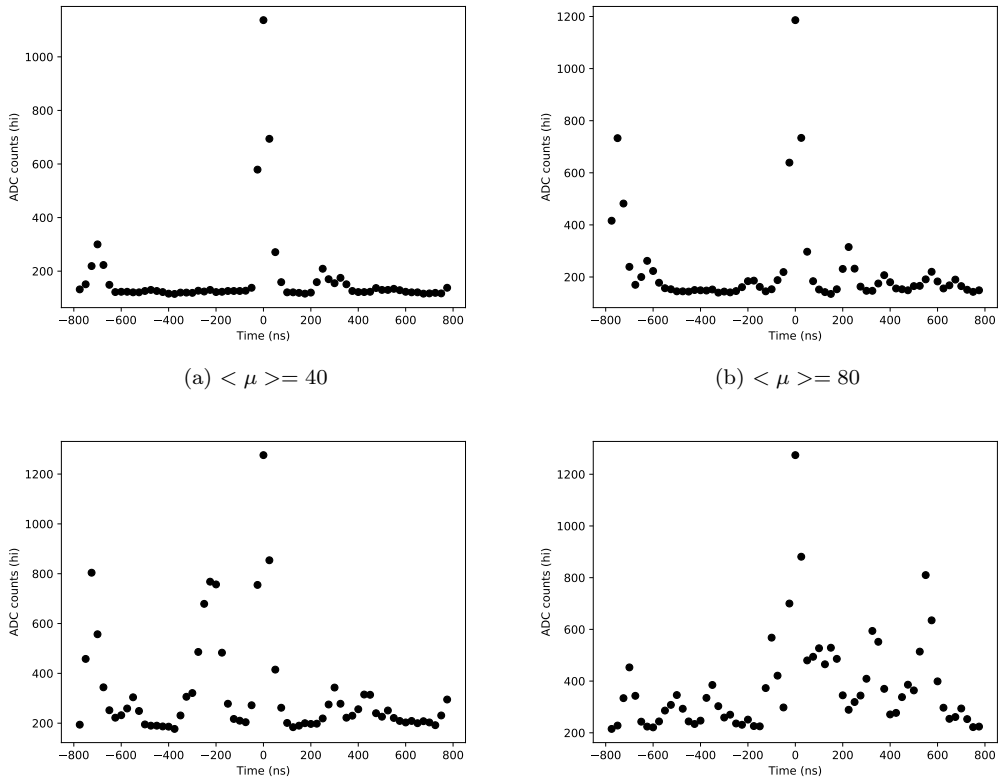


Figure 3.17: Simulation of the total fluence of hadrons with $E > 20$ MeV per cm^2 in TileCal for an integrated luminosity of 4000fb^{-1}. The simulation considers minimum bias proton-proton events at $\sqrt{s} = 14$ TeV, with a cross-section of 80 mb. The interaction of the particles with the detector material is simulated with the GEANT3/GCALOR package [132]. The largest fluence of particles is observed in the E cells, followed by the A cells that are closer to the gap between the barrel and the extended barrel. These cells experience the largest impact from pile-up signals.

the total true amplitude produced by pile-up signals in a concrete bunch crossing. This process is repeated for each of the bunch crossings taken into account in the simulation. Then, each of the true amplitudes is convoluted with the pulse shape of the electronics under consideration and the contributions of all the pulses are summed to obtain the digitized samples. Each sample is varied individually to take into account the effect of the electronic noise. Examples of individual pulses in an A1 cell affected by different levels of pile-up are shown in Figure 3.18. The cell is located in the long barrel, but it is moderately affected by the pile-up signals due to its position in the innermost layer of the detector. The plots show in-time pulses of 1000 ADC counts in high gain, and out-of-time pile-up signals associated with an average of 40, 80, 140 and 200 *pp* interactions per bunch crossing. The pulses were simulated reading out 63 samples in each event, and the number of bunch crossings taken into account in the simulation was 201. The effect of the pile-up signals is visible in the digitized samples, especially for the pulses affected by the highest pile-up levels.

3.5.6 Continuous read-out simulation

One of the primary motivations for upgrading the TileCal electronics, beyond the need for radiation-hard components, is to ensure compatibility with the new trigger architecture and with the higher readout and processing rates. In the legacy system,



(a) $< \mu >= 40$ (b) $< \mu >= 80$

(c) $< \mu >= 140$ (d) $< \mu >= 200$

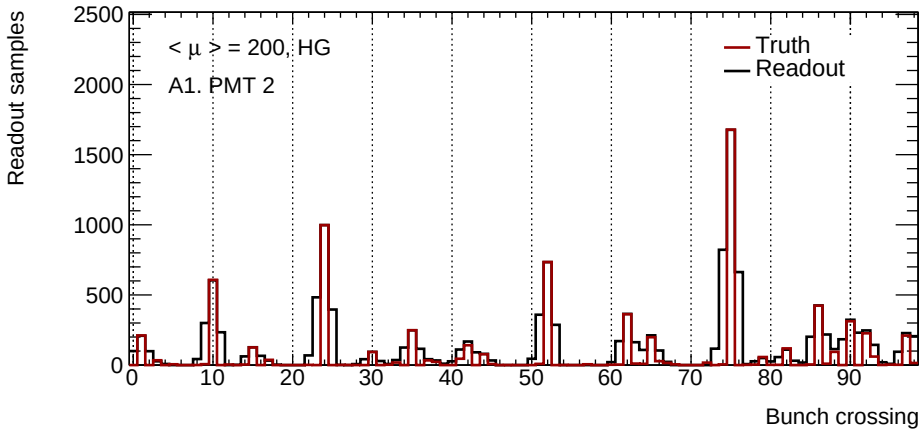
Figure 3.18: Pulses with an amplitude of 1000 ADC counts in high gain in an A1 cell affected by different levels of pile-up.

the analog signals are summed in projective towers in pseudorapidity and sent to the Level 1 trigger system. While the trigger system processes the decision, the digitized samples are stored in pipeline memories within the on-detector electronics. Once the trigger decision is made and communicated to the on-detector electronics, the samples associated with the triggered bunch crossing are read out and sent to the off-detector electronics for signal reconstruction. In contrast, the new architecture reads out samples continuously and transmits them to the off-detector electronics, where signal reconstruction is performed in real-time at 40 MHz. There is no triggered bunch crossing before the signal reconstruction: all the bunch crossings are read out and reconstructed. Trigger primitives are then generated in the off-detector electronics and sent to the Level 0 trigger system, where the decision is made, while the readout samples and reconstructed information are held in pipeline memories within the off-detector electronics. The Pulse Simulator was updated to generate a continuous pipeline of readout windows, according to the HL-LHC paradigm, as the previous implementation was only capable of simulating independent pulses, following the legacy paradigm.

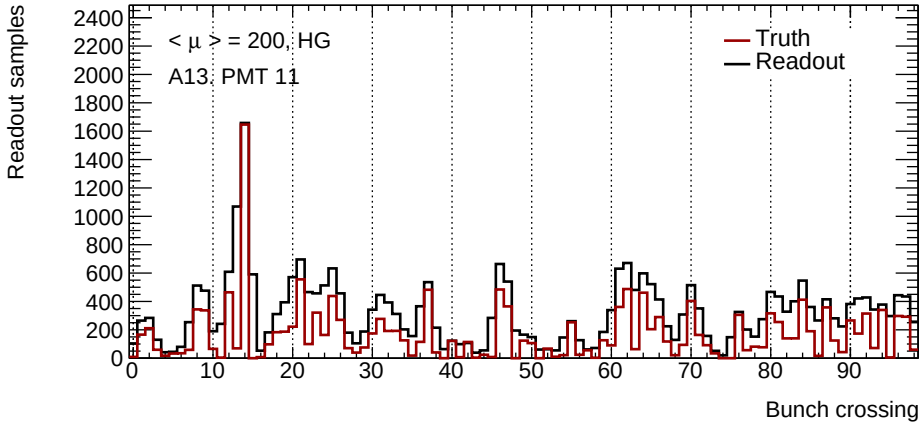
The continuous read-out mode in the Pulse Simulator is implemented using a sliding window approach. This window typically spans 10000 bunch crossings by default, corresponding to 250 μs, which is significantly longer than a single

LHC orbit, lasting 89.3 μs. True amplitudes of the pile-up and in-time signals are generated for each of the bunch crossings in the window, following the same procedure as described in the simulation of pile-up in the previous section. These true amplitudes are then convolved with the pulse shape of the desired electronics. This information is used to generate the digitized sample associated with the central bunch crossing in the window. Subsequently, the window is shifted by one bunch crossing (the oldest bunch crossing is removed from the window, and a new one is added, properly simulating the pile-up signals), and the process is repeated to generate the digitized sample associated with the new central bunch crossing. This process is repeated as many times as specified by the user, and the digitized samples are stored in the output file. In this way, the Pulse Simulator can simulate a continuous pipeline of readout samples. Examples of the simulated pulse sequences in continuous read-out mode for different cells in the TileCal detector are shown in Figure 3.19. The simulation was performed with a bunch spacing of 25 ns and an average of 200 *pp* interactions per bunch crossing. The plots show the digitized samples of 100 bunch crossings with the high gain readout of a channel of an A1 cell, an A13 cell, and an E4 cell. In the E4 cell, some samples are saturated due to the high amplitude produced by the pile-up signals.

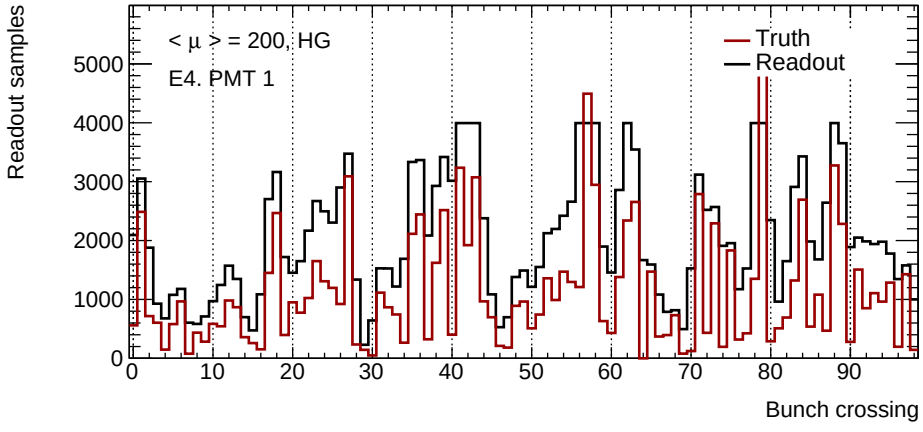
The ability to simulate the continuous read-out mode by the Pulse Simulator opens up new possibilities for studying the performance of signal reconstruction algorithms for TileCal under HL-LHC conditions. However, there are still some limitations in the simulation of the continuous read-out mode. The Pulse Simulator cannot currently simulate the train structure of the LHC. While the main part of the pulse shape has a limited duration, and the effect of pile-up signals is confined to a few bunch crossings, the tails of the pulse shape can extend for several microseconds, and the impact of pile-up signals can be significant in the digitized samples many bunch crossings after the pile-up signals were produced. The train structure may influence the digitized samples, particularly considering the so-called abort gap, a 3 μs gap without particle bunches required by the extraction kicker system of the LHC beam. Since no particles are present in the abort gap, there will be no signal pulses in the TileCal detector, and the baseline fluctuation in the digitized samples close to the abort gap caused by the negative tails of previous pulses will differ with respect to the rest of bunch crossings. Incorporating the train structure into the simulation of the continuous read-out mode would allow for more realistic conditions, as the absence of the abort-gap structure in the simulation currently overestimates the effect of the pile-up for some bunch crossings.



(a)



(b)



(c)

Figure 3.19: Simulated pulse sequences in the continuous read-out mode for different cells in the TileCal detector. The simulation was done with a bunch spacing of 25 ns and an average of 200 *pp* interactions per bunch crossing. The plots show the digitized samples in the high gain for 100 bunch crossings in a (a) A1 cell, (b) A13 cell and (c) E4 cell.

Chapter 4

Data and simulated samples

4.1 Proton-proton interactions at the LHC

The protons collided at the LHC are not elementary particles, but rather composite particles made of quarks and gluons. In reality, the entities that actually collide within the LHC are proton constituents. While many properties of the proton are determined by its valence quarks (two up quarks and one down quark), these are not its only components. The proton is better described as a dynamic sea of quarks and gluons, continuously being created and annihilated. The composition of the proton varies depending on the energy scale at which it is examined. At low energy scales, many of the physical processes inside the proton are dominated by non-perturbative effects. However, the high-energy collisions at the LHC, which involve large momentum transfers, allow the interactions between quarks and gluons to be described using perturbative calculations, thanks to the phenomenon of asymptotic freedom in QCD [9].

To compute the cross-section for a given process after a proton-proton collision, $\sigma(pp \to X)$, it is necessary to convolve the parton distribution functions (PDFs) of the partons (the probability of finding a parton with a certain momentum fraction x inside the proton) with the partonic cross-section $\hat{\sigma}(ij \to X)$ for the process $ij \to X$. The PDFs encode the low-energy, non-perturbative physics of the proton at a given energy scale, which are determined by global fits to experimental data [144]. The partonic cross-sections, on the other hand, are calculated using perturbative QCD, since the momentum transfer in the hard scattering process is large enough to allow the use of perturbation theory. The partonic cross-sections are then summed over all possible parton combinations in the protons, weighted by the PDFs, to obtain the total cross-section for the process. The computation of $\sigma(pp \to X)$ can be expressed as:

$$\sigma(pp \to X) = \sum_{ij} \int dx_1 dx_2 f_i(x_1, \mu_F) f_j(x_2, \mu_F) \hat{\sigma}(ij \to X) \left(x_1, x_2, \mu_F, \mu_R, \alpha_s(\mu_R)\right), \tag{4.1}$$

where $f_i(x, \mu_F)$ ($f_j(x, \mu_F)$) are the PDFs of parton i (j) in the proton, μ_F is the factorization scale, μ_R is the renormalization scale, and $\alpha_s(\mu_R)$ is the strong coupling constant evaluated at the scale μ_R.

The factorization scale is an arbitrary parameter that is introduced to separate the long-distance, non-perturbative physics of the proton from the short-distance, perturbative physics of the hard scattering process. This separation allows the calculation of the cross-section to be factorized into two parts through the so-called factorization theorem: the PDFs, which encode the non-perturbative physics of the proton, and the partonic cross-sections, which can be calculated using perturbation theory [145]. This scale is usually set to the transverse momentum of the hard scattering process.

The renormalization scale, on the other hand, is introduced to absorb the divergences that arise in perturbative calculations due to the self-energy of the particles involved in the process [146]. These divergences are absorbed into the coupling constant, which becomes a function of the renormalization scale (i.e. the coupling constant becomes scale-dependent, and the strength of the interaction changes with the energy scale at which it is probed). The renormalization scale is usually set to the energy scale of the hard scattering process. The partonic cross-sections are calculated using perturbative QCD, and are usually computed to a given order in the strong coupling constant α_s. The higher the order of the calculation, the more terms are included in the perturbative expansion, and the more accurate the prediction of the cross-section becomes.

On the other hand, the PDFs are determined by global fits to experimental data at a given energy scale. The PDFs are scale-dependent, and evolve with the energy scale at which they are probed. The evolution of the PDFs is described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations, which describe how the PDFs change as the energy scale of the process increases [147–149]. The equations rely on the splitting functions, which describe the probability of a parton splitting into two partons as a function of the energy scale of the process. The splitting functions are calculated using perturbative QCD, and are usually computed to a given order in α_s.

4.2 Event simulation

The simulation of proton-proton collisions at the LHC involves several stages [150]. The process begins with the calculation of the cross-section for the hard scattering, which models the interaction between partons within the protons. Next, parton showering is simulated, describing the emission of gluons and the splitting of partons. This is followed by the hadronization phase, where partons combine to form color-neutral hadrons, and the decay of unstable particles is simulated. Additionally, the underlying event and pile-up are incorporated, accounting for additional proton-proton interactions within the same bunch crossing as the primary

scattering. The event is then processed through a detector simulation, modeling the interactions of particles with the detector, leading to their reconstruction as physics objects. Finally, the simulated detector response is digitized, and the signals undergo the same reconstruction algorithms applied to real data. An illustration of a proton-proton collision at the LHC is shown in Figure 4.1.

4.2.1 Matrix element and parton showering

The partonic cross-sections for the hard scattering process are calculated using perturbative QCD, since it involves large momentum transfers that allow the use of perturbation theory. At these energy scales, the partons can radiate gluons and split into other partons, which leads to the emission of additional partons in the initial and final states. A full calculation of the cross-section would require the inclusion of all possible parton emissions [152]. The partonic cross-sections including all possible terms in the perturbative expansion can be expressed as:

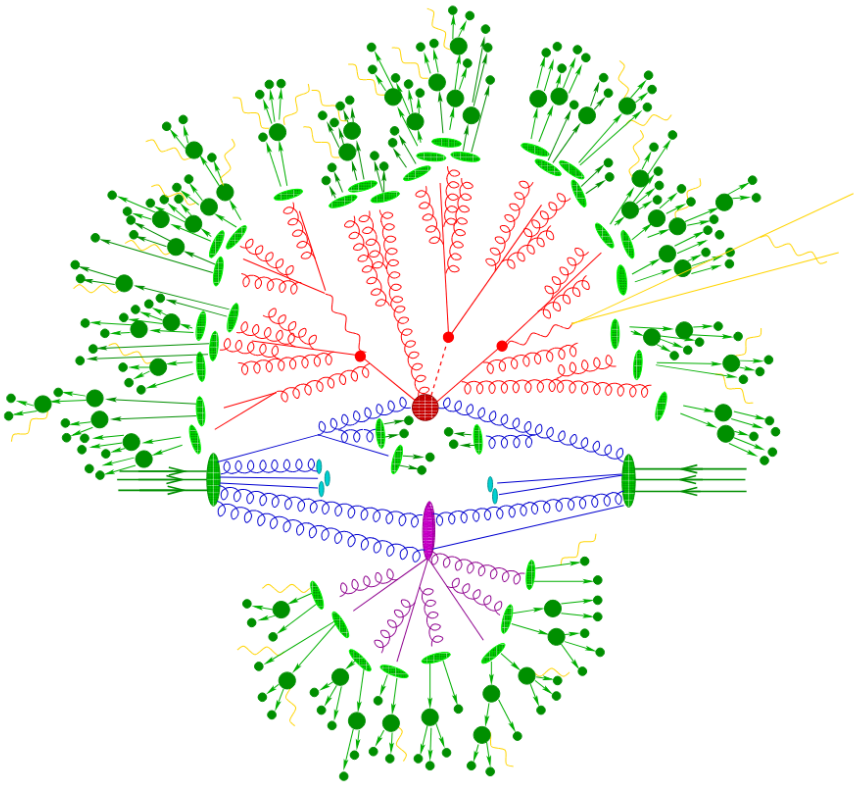


Figure 4.1: Diagram illustrating the components relevant for simulating a proton-proton collision event, excluding pile-up [151]. The initial partons from the colliding protons, shown in blue, contribute to the hard interaction (large dark red region). Additional hard QCD radiation, outgoing partons, and subsequent decays are also represented in red. The hadronization of the final state partons is indicated by light green ellipses, while the decay of the resulting hadrons is represented by dark green regions with arrows. Partons that do not participate in the primary scattering make up the underlying event, shown in purple. Photon radiation, which can occur at any stage, is illustrated in yellow.

$$\hat{\sigma}(ij \to X) = \sum_{n=0}^{\infty} \int \mathrm{d}\Phi_{X+n} \left| \sum_{k=0}^{\infty} \mathcal{M}_{X+n}^{(k)} \right|^2 , \tag{4.2}$$

where $\mathcal{M}_k^{(n)}$ is the matrix element for the process $ij \to X + n$, being n the number of additional partons in the final state, and $\mathrm{d}\Phi_{X+n}$ is the phase space element for the process $ij \to X + n$. The number of virtual loops included in the calculation is denoted by k. The matrix elements are computed up to a given order in the strong coupling constant α_s. For example, the leading order (LO) calculation includes only the tree-level matrix elements, and corresponds to $n = 0$ and $k = 0$. If the process to be studied involves the production of N partons in the final state, then a computation with $n = N$ and $k = 0$ would be referred as a leading order calculation for the process $ij \to X + N$. A matrix element calculation with $k + n = m$ is referred to as a calculation at $\mathrm{N^mLO}$ for the process $ij \to X$, but just at $\mathrm{N^{m-k}LO}$ for the process $ij \to X + n$.

This is the exact approach to compute the cross-section for a given process involving the production of N partons in the final state. However, the calculation of the matrix elements for processes with a large number of final state partons becomes computationally expensive, since the number of Feynman diagrams grows factorially with the number of partons. To overcome this limitation, the matrix elements are calculated up to a given order in the strong coupling constant, and the missing terms are approximated using the parton showering algorithm, where the splitting of partons into other partons is simulated with approximate matrix elements [153]. The parton showering approximation allows the simulation of the cross-section for the production of $n + 1$ partons starting from the cross-section for the production of n partons and taking into account the parton splitting functions, that describe the distribution of the energy and momentum of the radiated partons. The showering starts at the scale of the hard scattering process, and is applied subsequently to all the partons simulated in the hard scattering process until the scale of the showering is below a certain cutoff scale, usually set to the hadronization scale around 1 GeV.

4.2.2 Matching

The hard scattering process is described by full matrix-element calculations, while the subsequent evolution of hard partons into jets is handled through QCD parton cascades and phenomenological hadronization models. However, ensuring consistency between these different components is complex due to the multiple scales involved in an event, making it difficult to cleanly separate the hard process from the parton evolution.

In events with multiple jets, soft and collinear parton emissions typically do not alter the jet structure, but the emission of a high-energy, large-angle parton can add a new jet, transforming an n-jet event into an $n + 1$ jet event. To resolve

this, a merging scheme is needed to decide, on an event-by-event basis, whether a given emission should be modeled by the matrix element calculation or the parton shower. The primary goal is to prevent double-counting, ensuring that an event does not appear multiple times under different modeling schemes.

There is an inherent inconsistency between the matrix element computation and the parton showering for calculating the full cross-section at orders $n > 1$, due to possible overlap in the phase space of additional partons included in the matrix element at order n and those from the splitting at order $n - 1$. Various methods, known as matching, address this issue of double counting. A common solution is slicing, which introduces a matching scale: the higher energy region is described exclusively by the matrix element, while any additional partons with lower energy are handled by the parton shower algorithm, ensuring each energy range is optimally modeled by the appropriate method.

Two commonly used approaches are the MLM [154] and CKKW [155] schemes. The MLM method employs a geometric approach, using cone jets to analyze radiation patterns, while the CKKW scheme relies on analytical reweighting of matrix elements combined with constrained parton-shower evolution. Both approaches introduce a threshold that dictates which part of the event is described by the parton shower and which by the matrix element. As a result, tree-level multileg matrix elements with increasing jet multiplicity are merged with parton showers to create an inclusive sample, avoiding double-counting while maintaining accuracy at leading-log order. The MLM prescription at NLO is also known as the FxFx scheme [156].

4.2.3 Hadronization

When the parton shower falls below the cutoff scale, around 1 GeV, the partons are combined into color-neutral hadrons through a process known as hadronization. Since the energy scale is low, QCD interactions become non-perturbative, and the hadronization cannot be described using perturbative techniques. Instead, phenomenological models, such as the Lund string model and the cluster model, are used to simulate this stage. In both cases, partons are assigned a momentum, flavor, and color after the shower, and these models provide a way to reconstruct the hadronization process by grouping the partons into hadrons.

The Lund string model [157] simulates the hadronization process as the formation of a string between quarks and antiquarks, where the potential energy grows with distance, mimicking the QCD potential. If the energy in the string exceeds the total mass of a quark-antiquark pair, the string breaks, creating new quark-antiquark pairs and producing additional hadrons. The momentum of these hadrons is determined by a fragmentation function that connects it to the original partons.

In the cluster model [158] the final-state gluons are split into quark-antiquark pairs. These quarks are then grouped into color-neutral clusters. The clusters

undergo a series of decays and fragmentations, breaking into smaller clusters or directly forming hadrons. This process continues iteratively in both models until all partons are confined into stable hadrons.

4.2.4 Underlying event and pile-up

The underlying event refers to the additional activity in a proton-proton collision that is not directly associated with the hard scattering process. It typically consists of soft interactions between partons and is modeled using phenomenological approaches due to the non-perturbative nature of these interactions. The partons involved in these soft interactions are connected to those participating in the hard scattering process through color connections. The description of the underlying event is composed of several components. The first is multiple parton interactions, which describe the semi-hard interactions between different partons within the proton. These interactions may produce additional jets in the final state with low transverse momentum, often oriented back-to-back in the azimuthal angle, unlike jets from initial-state radiation, which align with the parton that radiated them.

The second component of the underlying event includes the beam remnants—those portions of the protons that do not take part in either the hard scattering process or the multiple parton interactions. This component is soft and is also modeled using phenomenological methods.

Pile-up and non-collision background simulation refers to the simulation of extra proton-proton interactions beyond the one containing the hard scattering event [159]. These backgrounds consist of several elements. The first is in-time and out-of-time pile-up, referring to additional interactions occurring either within the same bunch crossing or in nearby crossings. Another component is the beam halo, which arises from protons in the beam that do not focus on the center of the detector and may interact with detector or accelerator components, such as scraping the collimators before the interaction point, producing muons that travel parallel to the beamline. The simulation also accounts for interactions between protons and residual gas in the beam pipe. Additionally, background originating from the LHC cavern is simulated, which mostly affects the muon spectrometer.

To simulate these effects, each of these components is generated separately and combined with the hard scattering process before being passed through the detector simulation. The pile-up events are soft QCD interactions that are overlaid, and the number of pile-up events is taken from the distribution of the number of interactions per bunch crossing in the data.

4.2.5 Detector simulation

The generated particles need to be passed through a simulation of the ATLAS detector in order to model how the particles interact with the detector and are reconstructed into the physics objects that are used in the analyses. The simulation

of the detector response is done using the GEANT4 toolkit, which simulates the passage of particles through the detector material and the response of the detector to the particles [160]. The energy deposits in the detector are then converted into detector signals, which have the same output format as the real data. The simulated signals are then passed through the same reconstruction algorithms as the data, and the reconstructed objects are used in the same way as the data in the analyses.

The simulation of the complete geometry of the sub-detector systems using GEANT4 is referred to as *Full Simulation*. While this approach ensures the highest level of precision, it demands considerable computational resources, with each event often requiring several minutes to process. Given that over 90% of the CPU resources are allocated to calorimeter simulations, more efficient alternatives are frequently employed. The AtlFast-II method (frequently known as *Fast Simulation*) incorporates accelerated techniques for both calorimeter and inner detector simulations [161, 162] and is used to simulate these systems. The remaining sub-detectors continue to be simulated with Geant4, contributing significantly less to the overall CPU time while still achieving a satisfactory level of accuracy.

The detector description relies on a detailed simulation of the geometry of the detector, containing information about the dimensions of the detector components and their material composition. However, this is not the only information that is needed to simulate the detector response. The simulation also needs to take into account the detector conditions, such as the dead regions of the detector, the noise in the readout electronics, the misalignment of the components of the detector or the calibration of the detector, among other aspects. The simulation of Monte Carlo events is done using the conditions at the time of the simulation production, and this implies that the simulation of the detector always contains some level of mismatch with the real data, since the conditions of the detector change constantly during the data taking.

4.2.6 Monte Carlo generators

Monte Carlo generators are dedicated software tools designed to perform event simulations based on pseudorandom number generation to produce events with predicted distributions. These generators are typically classified into general-purpose and specialized types: general-purpose generators can reproduce all steps of event generation, while specialized generators focus on specific functionalities, such as matrix element calculations or parton shower modeling.

Several MC generators are commonly employed to simulate the full event process, including hard scattering, parton showering, hadronization, and the underlying event. PYTHIA [163] is a general-purpose generator to simulate high-energy proton-proton collisions. It utilizes LO matrix elements for $2 \to n$ (where $n \leq 3$) processes and incorporates a p_{T}-ordered parton shower, based on the Lund string

model for hadronization. While Pythia effectively captures soft and hard interactions, including initial-state radiation and final-state radiation, its LO cross-section calculations may not always suffice for precise analyses, making it frequently utilized as a parton shower generator in conjunction with dedicated matrix element generators.

Herwig [164] offers similar capabilities but emphasizes angular-ordered parton showers. It computes $2 \to 2$ LO matrix elements and addresses gluon splitting processes in a manner that fully incorporates spin correlations, a feature that differentiates it from Pythia. Herwig is capable of simulating a wide range of processes with NLO precision for the matrix element calculation, but leading to a large fraction of events with negative weights, that need to be handled carefully in different stages of the analysis. For that reason, the generator is often used in conjunction with other generators that provide the matrix element calculation at higher orders, but that do not include the parton showering and hadronization steps. Herwig employs a cluster model for hadronization and is often interfaced with frameworks that simulate the underlying event.

Sherpa [165] is another versatile MC generator that can provide both LO and NLO matrix elements for processes involving multiple partons, applying the CKKW matching procedure to transition between ME and parton showers. Its cluster model for hadronization allows for effective simulations of complex final states, particularly those involving multiple jets or electroweak bosons.

For more precise high-order matrix element calculations, MadGraph5_aMC@NLO [166] is frequently employed, implementing the MC@NLO method to interface with parton showers using MLM and FxFx matching prescriptions. It enhances the accuracy of event kinematics and cross-section predictions, often used in conjunction with Pythia or Herwig for a comprehensive simulation.

The PowhegBox [167, 168] framework also stands out for its ability to provide high-order matrix element calculations. It implements a 5-flavor scheme [169] and employs the POWHEG method [170] for matching, ensuring a consistent treatment of QCD corrections in both the matrix elements and parton showers.

4.3 Data and simulated samples

The data used for the analyses in this thesis were collected from proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV by the ATLAS detector during Run 2 of the LHC (2015–2018). Events were selected based on quality criteria, ensuring all relevant detector components were operational [171]. The total integrated luminosity used differs slightly between the two analyses: 138 fb^{-1} for the lepton-flavor violating decay search, and 140.1 fb^{-1} for the $H \to \tau\tau$ analysis, where the latter reflects the latest luminosity calibration [172].

The analyses presented in this thesis rely on simulated Monte Carlo samples for both signal and background processes. These samples are used to predict

the expected event yields and model the detector response. The simulations incorporate up-to-date theoretical predictions for cross-sections, parton distribution functions (PDFs), and advanced event generation techniques, including parton showering and hadronization models. Signal samples focus on the main Higgs boson production mechanisms, while background samples cover a variety of Standard Model processes that can mimic the signal topologies. Table 4.1 provides an overview of the MC generators used for the key signal and background processes, as well as the corresponding cross-section normalizations, calculated at different orders of perturbative QCD and electroweak theory. The following sections detail the specific MC generators and configurations used for the signal and background samples in the analyses.

Process	Generator		PDF set		Tune	Normalization
	ME	PS	ME	PS		
Higgs boson						
ggF	PowhegBox v2	Pythia8	PDF4LHC15nnlo	CTEQ6L1	AZNLO	N^{3}LO QCD + NLO EW
VBF	PowhegBox v2	Pythia8	PDF4LHC15nlo	CTEQ6L1	AZNLO	NNLO QCD + NLO EW
VH	PowhegBox v2	Pythia8	PDF4LHC15nlo	CTEQ6L1	AZNLO	NNLO QCD + NLO EW
$t\bar{t}H$	PowhegBox v2	Pythia8	NNPDF3.0nnlo	NNPDF2.3lo	A14	NLO QCD + NLO EW
tH	MadGraph5_aMC@NLO	Pythia8	CT10	NNPDF2.3lo	A14	NLO
bbH	PowhegBox v2	Pythia8	NNPDF3.0nnlo	NNPDF2.3lo	A14	NLO
Background						
Vjets (QCD/EW)	Sherpa v2.2.1		NNPDF3.0nnlo		Sherpa	NNLO for QCD, LO for EW
ttbar	PowhegBox v2	Pythia8	NNPDF3.0nnlo	NNPDF2.3lo	A14	NNLO + NNLL
Single top	PowhegBox v2	Pythia8	NNPDF3.0nnlo	NNPDF2.3lo	A14	NLO
Diboson	Sherpa v2.2.1		NNPDF3.0nnlo		Sherpa	NLO

Table 4.1: Summary of the Monte Carlo generators employed for the primary signal and background samples. The final column, titled *Normalization* indicates the perturbative order used in the cross-section calculations for normalizing the simulated datasets.

4.3.1 Simulation of Higgs boson samples

The analyses presented in this thesis consider the main production modes of the Higgs boson at the LHC. In the case of the differential cross-section measurements, the considered production modes are the gluon-gluon fusion, vector boson fusion, the associated production with a vector boson, and the associated production with a top quark pair. The smaller contributions from the associated production with a top quark and with two b quarks are also considered. In the case of the search for lepton-flavor-violating Higgs boson decays, the production modes considered are the ggF, VBF, and VH. The setup used for to simulate the production of the Higgs boson is the same in the two analyses.

The ggF signal samples are generated at NNLO in QCD using Powheg NNLOPS. The prediction of the Monte Carlo generator is scaled to the cross-section computed at N^{3}LO in QCD [25, 26, 173–176] including NLO electroweak corrections [27, 177]. The samples are generated using the PDF4LHC15nlo PDF

set [178] with the AZNLO [179] tune for Pythia8 used for the parton showering and hadronization. In order to derive uncertainties on the parton showering and underlying event modeling, an alternative sample is generated using the Powheg NNLOPS generator interfaced with Herwig7 [180] for the parton showering and hadronization. Similarly, to estimate the uncertainties on the choice of the generator for the matrix element calculation, an alternative sample is generated using MadGraph5_aMC@NLO [166] interfaced with Pythia8. This setup makes use of the FxFx merging scheme to combine the matrix element calculation with the parton showering.

The VBF production mode events are generated at NLO using Powheg interfaced with Pythia8. The AZNLO tune is used for the parton showering and hadronization and the PDF4LHC15nlo PDF set is used for the PDFs. The prediction of the Monte Carlo generator is scaled to the cross-section computed at an approximate NNLO in QCD [28], and including NLO electroweak corrections [29, 30]. The calculations include the effects due to finite heavy quark masses and soft-gluon resummation at next-to-next-to-leading-logarithm (NNLL) accuracy. In order to estimate the uncertainties due to the choice of the parton showering and underlying event modeling, an alternative sample is generated using Herwig7 for the parton showering and hadronization, but keeping the matrix element calculation with Powheg. Similarly, to estimate the uncertainties due to the choice of the generator for the matrix element calculation, an alternative sample is generated using MadGraph5_aMC@NLO interfaced with Pythia8.

The production of the Higgs boson in association with a vector boson is generated at NLO with one additional parton using Powheg interfaced with Pythia8. The AZNLO tune is used for the parton showering and hadronization, and the PDF4LHC15nlo PDF set is used for the PDFs. Additionally, the gluon-induced production of the Higgs boson in association with a vector boson is generated at LO with the same setup. The prediction is normalized to the NNLO computation in QCD, including NLO electroweak corrections, for the quark-induced production, and to the NLO computation in QCD for the gluon-induced production [32–38]. As in the case of the ggF and VBF samples, an alternative sample is generated using Herwig 7.2 to estimate the uncertainties due to the choice of the parton showering and underlying event modeling.

The production of the Higgs boson in association with a top quark pair is generated with PowhegBox v2 at NLO in QCD [40–43] using the NNPDF3.0nlo PDF set [181]. The A14 tune of Pythia8.230 is used for the parton shower [182]. The bottom and charm quark decays are generated using EvtGen v1.6.0 [183]. The alternative samples to evaluate the impact of the parton showering and underlying event modeling are generated using Herwig 7.04 using the H7UE tune [184] interfaced with the same PowhegBox v2 setup for the matrix element calculation. To estimate the uncertainties due to the choice of the generator, an alternative sample is generated using MadGraph5_aMC@NLO interfaced with Pythia8, as in the case of the other signal samples.

The production of the Higgs boson in association with a single top quark is modeled at NLO using MadGraph5_aMC@NLO interfaced with Pythia8 using the CT10 PDF set and the A14 tune. The samples are normalized to the cross-section computed at NLO in QCD.

The bbH production mode is generated at NLO using PowhegBox v2 with the NNPDF3.0nnlo PDF set. The parton shower is simulated using Pythia8 with the A14 tune and the NNPDF2.3lo PDF set. The samples are normalized to the cross-section computed at NLO in QCD.

The branching ratios for the Higgs boson decays are computed using the HDECAY [185–187] and PROPHECY4F [188–190] programs. The normalization of the signal samples accounts for the branching ratios of the Higgs boson decays to the final states considered in the analyses. To compute the production cross-sections and branching ratios, the Higgs boson mass is set to 125.09 GeV. The samples that simulate the lepton-flavor-violating Higgs boson decays into $\mu\tau$ and $e\tau$ final states assume a branching ratio of 1% for such decays.

4.3.2 Simulation of background samples

The QCD V+jets production is generated at NLO with up to two additional partons using Sherpa v2.2.1 [165] using the NNPDF3.0nnlo PDF set. The prediction achieves LO accuracy in QCD with up to four additional partons using the Comix [191] and OpenLoops matrix element generators [192–194]. The matching of the matrix element calculation with the parton showering is done using the MEPS@NLO prescription of Sherpa. The predictions are normalized to the cross-sections computed at NNLO [195].

The electroweak V+jets production, giving rise to the $\ell\ell jj$, $\ell\nu jj$, and $\nu\nu jj$ final states, is generated at NLO with up to two additional partons using Sherpa v2.2.1 with the NNPDF3.0nnlo PDF set. The same setup used for the QCD V+jets production is used in this case for the matching of the matrix element calculation with the parton showering. To avoid overlaps with the semileptonic VV production topologies, the samples were generated using the VBF approximation, which requires a t-channel color-singlet exchange.

The production of $t\bar{t}$ pairs is generated at NLO using PowhegBox v2 using the NNPDF3.0nlo PDF set interfaced with Pythia8.230 with the A14 tune. The h_{damp} parameter, a resummation damping factor that controls the matching of the matrix element calculation with the parton showering (and consequently the amount of high-p_{T} radiation against the $t\bar{t}$ system recoil), is set to 1.5 times the top quark mass [196]. The bottom and charm quark decays are generated using EvtGen v1.6.0. The additional samples to estimate a parton-shower uncertainty are generated with the same setup but interfaced with Herwig 7.04 for the parton showering, using the H7UE tune and the MMHT2014nlo PDF set [197]. For the matrix element generator uncertainty, an alternative sample is generated using

MadGraph5_aMC@NLO with the NNPDF3.0nlo PDF set interfaced with Pythia8 using the A14 tune and the NNPDF2.3lo PDF set.

The single top quark production in the s-channel (t-channel) is generated using PowhegBox v2 at NLO in QCD using the NNPDF3.0nlo PDF set and considering a five-flavor (four-flavor) scheme. The parton showering is performed using Pythia8.230 with the A14 tune and the NNPDF2.3lo PDF set. The cross-sections are normalized to the NLO computation in QCD with Hathor 2.1 [198].

The diboson production is simulated using Sherpa v2.2.1 and Sherpa v2.2.2. The matrix element calculation is performed at NLO with up to one additional parton, and to LO accuracy with up to four additional partons. The gluon-induced production of WW, WZ, and ZZ is also simulated at LO with up to one additional parton. The matching and merging of the matrix element calculation with the parton showering is done using the MEPS@NLO prescription of Sherpa. The NNPDF3.0nnlo PDF set is used for the PDFs. The corrections from virtual QCD contributions are included using the OpenLoops library. The samples are normalized to NLO cross-section predictions.

Chapter 5

Object reconstruction and identification

The particles produced in the proton-proton collisions at the interaction point of the ATLAS detector produce signals in the different subdetectors. These signals are combined to reconstruct, calibrate and identify the different objects that are finally used in the physics analyses: tracks, vertices, electrons, photons, muons, jets, hadronic tau leptons and missing transverse momentum. This chapter describes the algorithms used to reconstruct and identify these objects.

5.1 Tracks and vertices

The trajectory of charged particles in the ID is one of the basic inputs that are used to reconstruct many kinds of objects, including electrons, muons or jets. The charged particles leave hits in the different layers of the ID. The position of these hits is used to reconstruct the trajectory of the particle, which is helicoidal due to the presence of the magnetic field produced by the solenoid, has a curvature inversely proportional to the transverse momentum of the particle. The tracks are characterized by five parameters: the transverse momentum, p_{T}, the polar angle, θ, the azimuthal angle, ϕ, the impact parameter in the transverse plane with respect to the nominal interaction point, d_0, and the impact parameter in the longitudinal plane with respect to the nominal interaction point, z_0. These parameters are represented in Figure 5.1.

The reconstruction of the tracks is done in several steps [200]. First, the hits in the different layers of the ID are associated to form track seeds. The hits in the Pixel detector and the SCT that are above a certain energy threshold are used to form the clusters. These clusters are combined in groups of three to form track seeds. Then, additional clusters in the ID are added to the track seeds using a combinatorial Kalman filter, forming track candidates [201]. The Kalman filter is a recursive algorithm that adds the clusters to the track candidate in a step-by-step way, updating in each step the parameters of the track (like the position and the

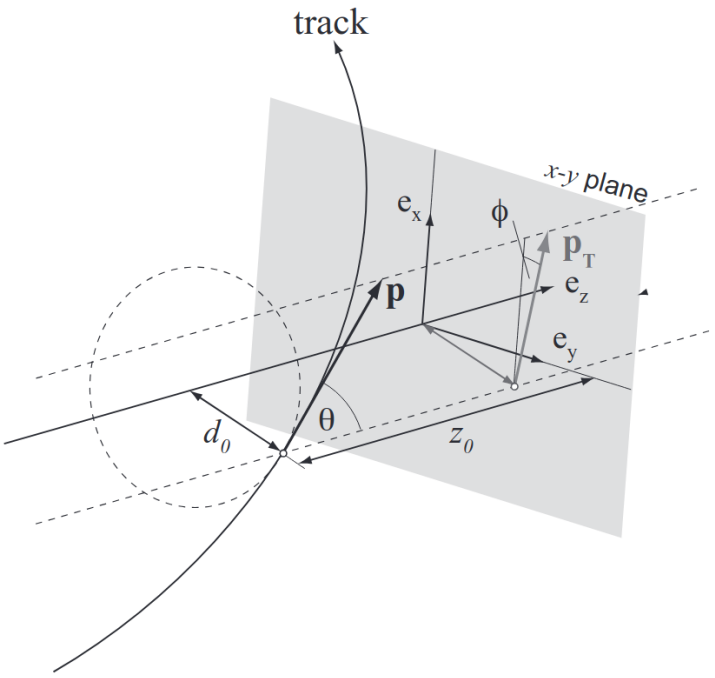


Figure 5.1: Schematic representation of a track and the parameters that are used to characterize a track in ATLAS. The e_z direction follows the beam line [199].

momentum) using the information of the new cluster and predicting the position of the track in the next layer. It is possible to have more than one track candidate for the same cluster, and therefore an ambiguity solving algorithm is used to select the best track candidate based on the quality of the track. The quality of the track is evaluated with a score that takes into account the χ^2 of the track fit, the p_{T} of the track, the number of clusters associated to the track candidate and the number of holes in the track. A hole is defined as a missing hit in the track that is expected to be found in a certain layer. This score penalizes tracks with holes and low p_{T}, since they are more likely to correspond to incorrectly reconstructed tracks. The score is used to select the best track candidate among the possible candidates for the same cluster. Once the ambiguity is solved, a set of requirements is applied to the track: the track is required to have at least 7 clusters between the Pixel and the SCT detectors, less than 2 holes in the Pixel and a maximum of 2 holes in the SCT, a p_{T} greater than 500 MeV and a $|\eta| < 2.5$. Furthermore, the track is required to satisfy $|d_0| < 2$ mm and $|z_0 \sin\theta| < 3$ mm. The tracks that satisfy these requirements are used to perform a high-resolution fit, using all the hits in the ID.

Tracks are then used to reconstruct the primary vertices of the event, defined as the interaction points of the proton-proton collisions [202, 203]. The primary vertices are reconstructed with a vertex finding algorithm, and then a fit is performed in order to determine the position of the vertex with a better precision. Once the vertex position is determined, the tracks that are incompatible with the vertex are removed from the list of tracks used to reconstruct the vertex, and can be

used to reconstruct other vertices. The primary vertex that has the highest sum of squared transverse momenta of the tracks associated to it, $\sum p_{\mathrm{T}}^2$, is selected as the hard-scatter vertex, and the other vertices are considered as pile-up vertices or secondary vertices, specially relevant for b-tagging algorithms.

5.2 Electrons

The electrons that traverse ATLAS leave a track in the ID and deposit their energy mainly in the electromagnetic calorimeter. The electrons produce showers of electromagnetic particles in the calorimeter, depositing a significant fraction of their energy through bremsstrahlung radiation and inducing subsequent conversions of the photons into electron-positron pairs.

5.2.1 Reconstruction

The reconstruction of electrons [204, 205] starts with the identification of the energy clusters in the electromagnetic and hadronic calorimeters. The first step is the identification of proto-clusters, that start from seed cells with an energy above four times the noise level of the calorimeter. Once the seed cells are identified, all the adjacent cells with an energy above twice the noise level are added to the proto-cluster. If two proto-clusters overlap, they are merged into a single cluster. If there are two local maxima of energy in the cluster and the energy in the cells where the maxima are found is above 500 MeV, the cluster is split into two clusters. The resulting clusters are called topo-clusters [206]. Electrons are expected to deposit most of their energy in the electromagnetic calorimeter. Only topo-clusters with an energy in the electromagnetic calorimeter greater than 400 MeV and a fraction of the energy in the electromagnetic calorimeter greater than 50% are considered as electron or photon candidates.

If there are no tracks in the ID compatible with the resulting topo-cluster, a pattern recognition algorithm that takes into account for energy losses due to bremsstrahlung radiation is used to find possible tracks. Furthermore, tracks that might be compatible with the topo-cluster are fitted again using a Gaussian Sum Filter algorithm that takes into account the energy losses due to bremsstrahlung radiation [207].

The topo-clusters are sorted in decreasing order of transverse energy, E_{T}, in the electromagnetic calorimeter, and they are required to have a minimum transverse energy of 1 GeV and an associated track with a minimum of four hits in the Pixel or SCT detectors to form electron seeds. The neighboring topo-clusters, called satellites, are added to the electron seed if they are within an angular distance of $\Delta\eta \times \Delta\phi < 0.0075 \times 0.125$. The window is extended to $\Delta\eta \times \Delta\phi < 0.125 \times 0.3$ if two clusters share the same track. Satellite clusters are the result of the electromagnetic shower, and therefore are removed from the list of

candidate clusters to form electron seeds. The combination of electron seeds and satellite clusters is called super-cluster.

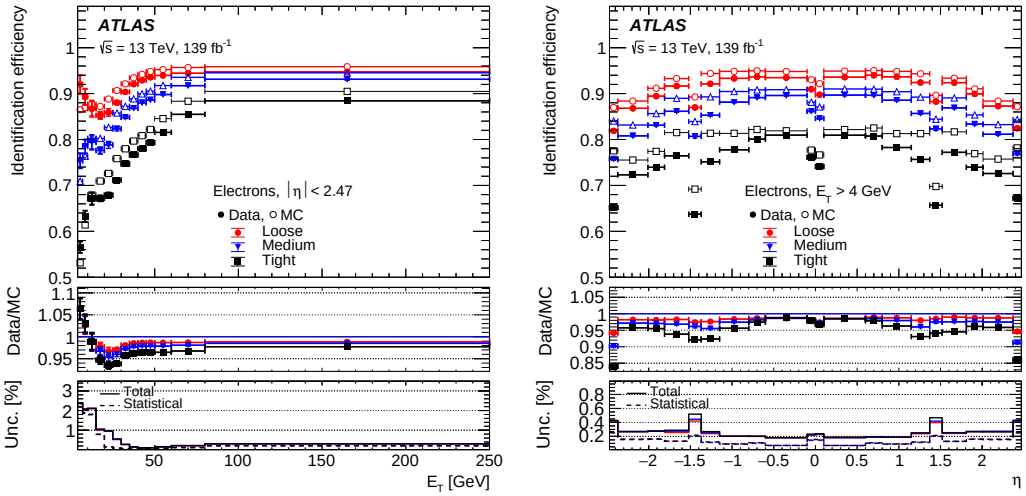
Once tracks are matched to super-clusters, an initial energy calibration is applied. If a cluster lacks a well-matched track, it is identified as a photon. If a good track match is found, the cluster is reconstructed as an electron. In ambiguous cases, both electron and photon objects are created for further analysis. The energies are then recalibrated, taking into account the matched track to more accurately reflect the original electron energy. The electron's p_{T} is derived from the calorimeter energy measurement, while its η and ϕ are obtained from the associated track properties.

5.2.2 Identification

The electron reconstruction algorithm can misidentify background objects, such as hadronic jets, converted photons, and electrons from semileptonic decays of heavy-flavour hadrons, as prompt electrons. To reject such misidentified objects, a likelihood-based identification algorithm is employed in Run 2 analyses [208]. This algorithm is based on a multivariate analysis technique that combines several discriminating variables. These variables capture the electron's track properties in the Inner Detector, the matching between the track and energy clusters, as well as the longitudinal and lateral shapes of the electromagnetic shower in the calorimeters. The likelihood method calculates a discriminant that separates signal electrons from background, based on the probability density functions for each variable. PDFs are derived from histograms in different $|\eta|$ and E_{T} bins, constructed using data from $Z \to ee$ events for signal and primarily dijet events for background.

The algorithm offers three working points (WPs): *Loose*, *Medium* and *Tight*. These points offering different levels of efficiency and background rejection. These WPs use the same discriminant variables, but the cut thresholds are tuned differently for each level, offering a balance between maximizing prompt electron identification and minimizing background contamination. For instance, the *Loose* WP efficiency ranges from 85% at $E_{\mathrm{T}} = 20$ GeV to 96% at $E_{\mathrm{T}} = 100$ GeV, while the *Tight* WP efficiency ranges from 55% at $E_{\mathrm{T}} = 4.5$ GeV to 90% at $E_{\mathrm{T}} = 100$ GeV.

Identification efficiencies are determined from both real data and simulated samples, and are defined as the fraction of reconstructed electrons that pass a given identification criterion. Figure 5.2 illustrates the efficiencies as a function of transverse energy and pseudorapidity, derived from $Z \to ee$ events using the tag-and-probe method. The efficiency varies across detector regions, with a notable drop between the barrel and the endcap regions ($1.37 < |\eta| < 1.52$). Additionally, simulation-to-data scale factors, which correct for discrepancies between the simulated and real data, are applied in bins of p_{T} and η to ensure accurate modeling of electron yields in physics analyses.



(a) (b)

Figure 5.2: The identification efficiencies for electrons originating from $Z \to e^+e^-$ decays are presented as a function of the electron's (a) transverse momentum and (b) pseudorapidity, corresponding to various identification working points. The upper panels illustrate the efficiencies derived from both data and simulation. The middle panels display the ratio of these efficiencies, which serves as a correction factor in analyses. Finally, the lower panels depict the relative statistical and total uncertainties associated with the data-to-MC ratio [205].

5.2.3 Isolation

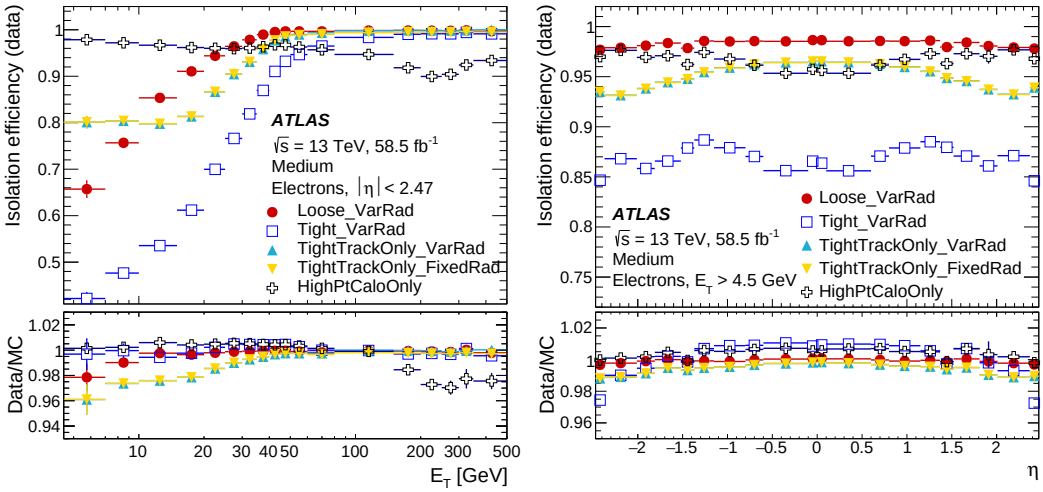
Electron isolation plays a crucial role in distinguishing prompt electrons from background sources, such as hadronic jets, photon conversions, and electrons from semileptonic decays of heavy-flavour hadrons. Prompt electrons are typically surrounded by minimal activity, whereas non-prompt electrons or misidentified objects are often accompanied by additional particles, resulting in more energy deposits around them. Isolation variables are used to quantify this surrounding activity by summing the transverse momenta of tracks or the transverse energies of calorimeter clusters within a cone around the electron candidate.

Two primary types of isolation are employed: calorimeter-based and track-based isolation. The calorimeter-based variable, $E_{\mathrm{T}}^{\mathrm{topocone20}}$, is defined by summing the transverse energy of topological clusters within a cone of $\Delta R = 0.2$ around the electron cluster's barycenter. To isolate the electron's surroundings, the energy deposited by the electron itself is subtracted within a rectangular region of $\Delta\eta \times \Delta\phi = (5 \times 7) \times 0.025$ around the cluster. However, corrections must be applied to account for energy leakage, pile-up, and underlying event effects, as these can influence the measurement, particularly in high-activity environments.

On the track-based side, the isolation variable $p_{\mathrm{T}}^{\mathrm{varcone20}}$ is calculated by summing the transverse momentum of selected tracks within a variable-sized cone centered on the electron track. The cone size decreases for higher-p_{T} electrons, defined as $\Delta R = \min(10/p_{\mathrm{T}}[\mathrm{GeV}], 0.2)$. Tracks within $|\eta| < 2.5$ and $p_{\mathrm{T}} > 1$ GeV are included in the sum, excluding the track associated with the electron itself.

This variable helps to reject nearby tracks from the same vertex, especially when dealing with boosted objects.

Isolation criteria are then defined by calculating the ratio of these isolation variables to the electron's transverse momentum, such as $E_{\mathrm{T}}^{\mathrm{topocone20}}/p_{\mathrm{T}}$ and $p_{\mathrm{T}}^{\mathrm{varcone20}}/p_{\mathrm{T}}$. Working points can be established by either applying fixed cuts or aiming for specific efficiency targets.



(a) (b)

Figure 5.3: The efficiency of various isolation working points for electrons from inclusive 2018 data in $Z \to e^+e^-$ events is shown as a function of (a) the electron's E_{T} and (b) the electron's η. The electrons are required to satisfy the criteria of the *Medium* selection from the likelihood-based electron identification. The lower panel presents the ratio of efficiencies measured in data compared to those in Monte Carlo simulations. Total uncertainties are included, including both statistical and systematic components. [205].

Isolation efficiencies, typically measured using the tag-and-probe method on $Z \to ee$ events, show small deviations between data and simulation, with scale factors close to unity for $p_{\mathrm{T}} > 10$ GeV.

5.3 Muons

The muons produced in the proton-proton collisions in the ATLAS detector are usually minimum-ionizing particles that traverse the detector leaving minimal energy deposits in the subdetectors. The muons leave hits in the ID, small energy deposits in the calorimeters, and tracks in the MS.

5.3.1 Reconstruction

Muon reconstruction begins in the MS with a pattern recognition algorithm that forms local track segments within individual chambers [209]. Hits from the various layers of the MS are combined to create preliminary track candidates. These

candidates are required to point approximately toward the interaction point (IP), with their curvature matching the expected bending in the $R - z$ plane due to the magnetic field. A χ^2 fit is then performed on the candidate, refining the trajectory by including additional hits along the predicted path while excluding hits that deviate from it. This iterative process continues until an optimal track is obtained. Ambiguities between multiple track candidates are resolved based on the track quality, and the final track is re-fitted with a loose IP constraint before being extrapolated to the beamline.

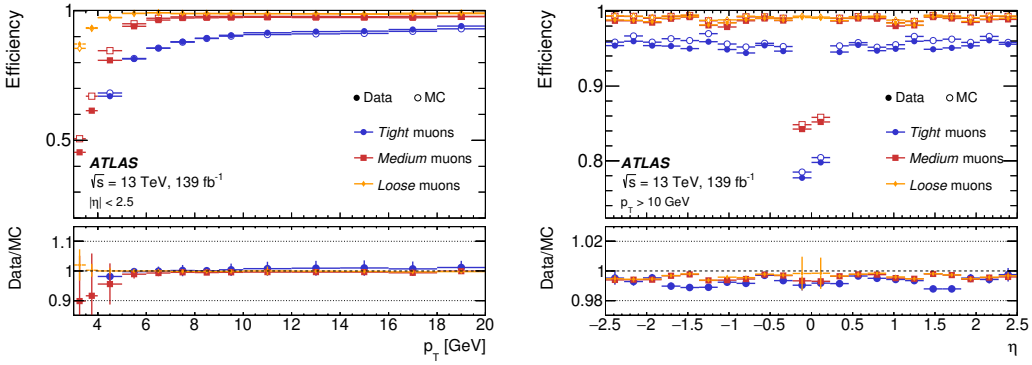
Global muon reconstruction integrates measurements from the ID, MS, and calorimeters. Five primary reconstruction methods are employed, each optimized for different muon candidates. *Combined Muons* (CB) are reconstructed by matching tracks from both the MS and ID, with energy losses in the calorimeters accounted for. A χ^2 fit is performed on the combined hits, iteratively refining the MS track by updating it with additional hits and removing outliers. *Inside-out Muons* (IO) are reconstructed by extrapolating an ID track into the MS, requiring at least three loosely aligned hits in the MS to form a valid candidate. This method does not require a standalone MS track and allows for improved reconstruction when the MS hits are sparse.

Muon Spectrometer Extrapolated Muons are reconstructed when no matching ID track is available, typically in the pseudorapidity region ($2.5 < |\eta| < 2.7$), where the ID coverage is limited. The MS track is extrapolated to the beamline, providing a standalone reconstruction in this high-$|\eta|$ region. *Segment-tagged Muons* (ST) are reconstructed by matching at least one track segment in the MS to an extrapolated ID track. In this case, the track parameters are determined entirely by the ID, with the MS serving to confirm the candidate. *Calorimeter-tagged Muons* (CT) are identified by matching an ID track to energy deposits in the calorimeters that satisfy the requirements for minimally ionizing particles. Due to background contamination, especially at low transverse momenta, a p_{T} threshold of 5 GeV is imposed for CT muons.

Muon candidates, excluding ME muons, must satisfy stringent requirements in the ID to ensure track quality. Each muon candidate must have at least one hit in the Pixel detector, five hits in the SCT, and fewer than three missing hits in the Pixel or SCT layers. Additionally, at least 10% of the Transition Radiation Tracker (TRT) hits initially assigned to the track must be included in the final fit.

5.3.2 Identification

Muon identification in ATLAS applies additional selection criteria to reconstructed muons to enhance the purity and quality of the candidates. These criteria are used to define different working points, each balancing efficiency and background rejection. The three main WPs are *Loose*, *Medium*, and *Tight*, while additional WPs such as *low-*p_{T} and *high-*p_{T} target muons with very low or high transverse momenta, respectively.



(a) (b)

Figure 5.4: Efficiencies for muon reconstruction and identification are determined for the *Loose*, *Medium*, and *Tight* criteria. The left plot shows the efficiency measured in $J/\psi \to \mu\mu$ events as a function of transverse momentum, while the right plot presents the efficiency measured in $Z \to \mu\mu$ events as a function of pseudorapidity (η) for muons with $p_{\mathrm{T}} > 10$ GeV. Open markers represent the predicted efficiencies, while filled markers correspond to measurements from data. Statistical uncertainties are indicated by error bars where relevant. The lower panel compares the measured and predicted efficiencies, including both statistical and systematic uncertainties [209].

The *Loose* WP includes all types of reconstructed muons: *combined*, *inside-out*, *segment-tagged*, and *calorimeter-tagged* candidates. *Loose* muons prioritize high efficiency at the cost of lower background rejection, especially in the pseudorapidity range $|\eta| < 0.1$, where the number of MS chambers is reduced due to structural constraints in the ATLAS detector. These muons have looser requirements on the transverse momentum and allow for lower-quality candidates, such as those with $p_{\mathrm{T}} < 7$ GeV.

The *Medium* WP, used for the analysis with muons in the signal processes, is more restrictive. It includes only CB and IO muon candidates within the pseudorapidity range $|\eta| < 2.47$. Muon candidates must include at least three hits in two or more layers of the MS precision chambers, except in the region $|\eta| < 0.1$, where one layer is sufficient. Additionally, the compatibility between the charge-over-momentum ratio (q/p) measured in the ID and MS is required to be less than 7, requiring loose agreement between the two measurements.

The *Tight* WP imposes the most stringent requirements to achieve maximum background rejection. This WP only includes CB and IO muons and demands that at least one track segment consists of three hits in two different MS stations, specifically in the mediators or the CSCs. The selection criteria also include cuts on the q/p significance, defined as the difference between the measurements of q/p in the ID and MS divided by their combined uncertainties, as well as in variables measuring the consistency of the p_{T} measurements across the two sub-detectors. These cuts are optimized in p_{T} and η bins to balance identification efficiency and background suppression.

Muon identification efficiency depends on both p_{T} and η, with the *Medium*

WP reaching efficiencies of up to 97% for prompt muons in simulated $t\bar{t}$ events. Figure 5.4 shows the efficiencies for muon reconstruction and identification as a function of p_{T} and η for the *Loose*, *Medium*, and *Tight* WPs in prompt muons from $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ events. For background muons, such as those originating from light hadron decays, the efficiency is much lower, at approximately 0.24%. These efficiencies are validated using real data from $Z \to \mu\mu$ events, and the disagreements between simulation and data are corrected through scale factors, which are typically close to 1. Larger deviations are observed in regions such as $|\eta| < 0.1$ or at specific azimuthal angles ($\phi \approx -1.2$ and $\phi \approx -2.0$) due to the presence of detector support structures. The scale factors are binned in η and ϕ to account for detector non-uniformities, ensuring that muon identification is accurately modeled in physics analyses.

5.3.3 Isolation

Muon isolation is used to distinguish prompt muons from backgrounds such as semi-leptonic hadron decays, which are accompanied by nearby particle activity. Isolation is measured by assessing the amount of momentum or energy in a cone around the muon, divided by the muon's p_{T}. Two types of isolation are defined: track-based isolation using the ID tracks and calorimeter-based isolation using the energy deposits in the calorimeter.

For track-based isolation, two observables are commonly used: a fixed cone of radius $\Delta R = 0.2$, defining the $p_{\mathrm{cone20}}/p_{\mathrm{T}}$ variable, or a variable cone radius $\Delta R = \min\left(10\,\mathrm{GeV}/p_{\mathrm{T}}, 0.3\right)$, defining the $p_{\mathrm{varcone30}}/p_{\mathrm{T}}$ variable. The transverse momenta of all ID tracks, excluding the muon, within this cone and associated with the primary vertex are considered for this calculation.

For calorimeter-based isolation, a fixed cone of $\Delta R = 0.2$ is used to define the $E_{\mathrm{T}}^{\mathrm{topocone20}}/p_{\mathrm{T}}$ variable. All energy deposits in the calorimeter, excluding those from the muon itself, are summed, with corrections applied for pile-up effects. Track-based isolation generally offers better resolution and lower pile-up dependence, while calorimeter-based isolation includes contributions from neutral particles and particles with transverse momenta below the threshold for ID detection.

Figure 5.5 illustrates the distributions of the isolation variables, normalized by the muon's p_{T}, for both prompt and non-prompt muons obtained from simulated $t\bar{t}$ events. As anticipated, the distribution for prompt muons exhibits a pronounced peak near zero, whereas non-prompt muons display a more uniform distribution, reflecting the capability of isolation variables to distinguish between the two types of muons.

Different isolation working points (WPs) are defined using these observables. For instance, the *FCTightTrackOnly_FixedRad* WP uses only track-based isolation with the fixed and variable cone size observables: $p_{\mathrm{varcone30}}/p_{\mathrm{T}} < 0.06$ for $p_{\mathrm{T}} \leq 50\,\mathrm{GeV}$ and $p_{\mathrm{cone20}}/p_{\mathrm{T}} < 0.06$ for $p_{\mathrm{T}} > 50\,\mathrm{GeV}$, with a requirement for tight track association to the vertex and p_{T} of tracks greater than 1 GeV.

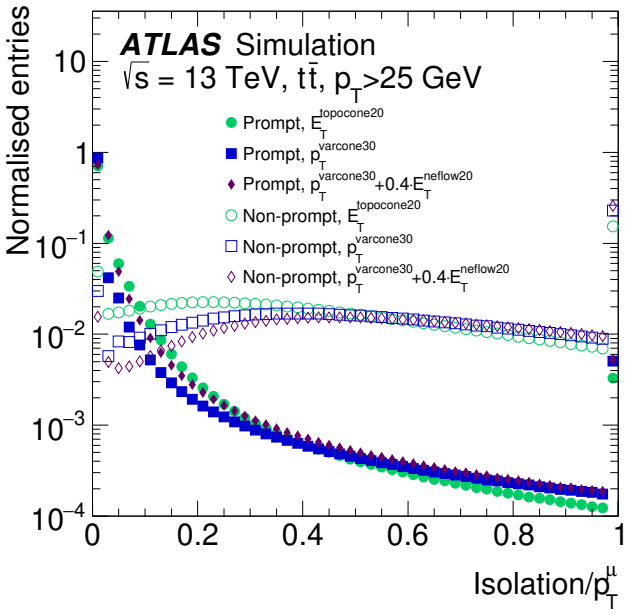


Figure 5.5: Distributions of different variables used to define the muon isolation WPs normalized to the muon p_{T} in simulated $t\bar{t}$ events. The variables are shown for prompt and non-prompt muons [209].

Isolation efficiencies are typically measured using tag-and-probe methods in $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ events. Scale factors to correct simulation to data are provided, with systematic uncertainties associated to them, including those from tag-and-probe methods and background estimation techniques.

5.4 Jets

In proton-proton collisions at the LHC, quarks and gluons produced at the partonic level undergo fragmentation and hadronization, resulting in collimated sprays of hadrons, referred to as jets. These jets are observed in the ATLAS detector through particle tracks reconstructed in the ID and energy deposits in both the electromagnetic and hadronic calorimeters. Dedicated algorithms cluster these objects, and similar algorithms can be applied to stable final-state particles in simulations to model jets.

Jet reconstruction in ATLAS relies heavily on sequential recombination algorithms, with the anti-k_t algorithm being the most commonly employed [210]. This algorithm is designed to be both infrared- and collinear-safe, ensuring stability in the presence of soft and collinear emissions from partons. The anti-k_t algorithm operates by defining a distance measure between two objects i and j (either tracks or topo-clusters) as:

$$d_{ij} = \min(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{5.1}$$

where $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the distance in η-ϕ space between the objects, and R is the radius parameter that defines the size of the jet. The exponent p is typically set to -1 for the anti-k_t algorithm, ensuring that objects with higher transverse momenta dominate the clustering. A beam distance d_{iB} is also computed for each object, defined as:

$$d_{iB} = p_{\mathrm{T},i}^{2p}, \tag{5.2}$$

where the smallest of these distance measures, either d_{ij} or d_{iB}, is selected at each iteration. If d_{ij} is smallest, objects i and j are merged into a new object. If d_{iB} is smallest, object i is classified as a jet and removed from the list of objects. The clustering process continues until no further objects remain. The algorithm's radius parameter R defines the characteristic scale of the jet and determines the extent of clustering in η-ϕ space, with typical values of $R = 0.4$ or $R = 1.0$ for small- and large-radius jets, respectively [211].

Following jet reconstruction, an energy calibration procedure is applied to correct for detector effects and achieve consistency with particle-level jet energies [212]. The first step is the pile-up correction, which accounts for additional interactions in the same or neighboring bunch crossings. This correction is derived using the jet area and the event transverse momentum density. Additionally, residual pile-up corrections are applied, which depend on the number of primary vertices in the event and the average number of interactions per bunch crossing.

The next step in the calibration chain is the jet energy scale (JES) calibration, which adjusts the reconstructed jet energy and pseudorapidity to align with the particle-level truth using scale factors derived from simulation. These corrections are p_{T}- and η-dependent, with specific adjustments made to account for the varying calorimeter response to electromagnetic and hadronic showers. The calibration is designed to ensure that the reconstructed jet energy matches the true energy of the initiating parton at the particle level.

Global Sequential Calibration (GSC) is applied after the JES correction to account for additional effects that depend on the jet's internal structure, flavor, or origin. GSC corrections are multiplicative factors that adjust the jet energy based on properties such as jet width, vertex association, and track information. This step aims to correct for systematic variations in the jet response between quark- and gluon-initiated jets, as well as jets from different fragmentation and hadronization processes.

Finally, in-situ calibration is performed to correct for any residual differences between the simulated and measured detector responses. This calibration step uses well-understood reference objects such as isolated photons or Z bosons decaying to leptons. By comparing the balance in transverse momentum between these reference objects and jets, correction factors are derived and applied only to jets in data, ensuring the jet energy response matches that in simulation. Additionally, multijet balance techniques are used to refine the calibration further.

Jet energy resolution (JER) is evaluated after the energy calibration, using dijet balance techniques and random-cone methods to measure fluctuations in the reconstructed jet energy [213]. The JER is parameterized as a function of p_{T} and η, and simulation-based smearing is applied to reproduce the observed resolution in real data.

To mitigate the impact of pile-up on jet reconstruction, the Jet Vertex Tagger (JVT) algorithm is employed to differentiate between jets originating from the primary vertex of the hard-scatter process and those produced by pile-up interactions [214]. The JVT uses track-based variables, including the fraction of jet-associated tracks that originate from the primary vertex, combined with corrections based on the number of vertices in the event and the reconstructed jet p_{T} after pile-up subtraction. The JVT algorithm evaluates jets within the $|\eta| < 2.4$ region, corresponding to the ID acceptance. To cover the forward region, the forward JVT (fJVT) algorithm is used, which extends the coverage to $|\eta| < 4.5$.

5.4.1 *b*-jet tagging

Identifying jets originating from b-quarks, commonly referred to as b-jets, is an essential task in many physics analyses at the LHC, particularly for processes involving top quarks and Higgs bosons. b-tagging algorithms are designed to distinguish b-jets from jets initiated by lighter quarks or gluons. The characteristic features of b-jets include the relatively long lifetime of B hadrons, their high transverse momentum fraction, and the presence of displaced secondary and sometimes tertiary vertices due to the non-prompt decays of B hadrons.

B hadrons typically travel a few millimeters before decaying, producing secondary vertices and tracks with large impact parameters relative to the primary vertex. This decay topology forms the basis for various tagging techniques. The impact parameter of tracks is an essential variable for separating b-jets from other jets. Track-based taggers like IP2D and IP3D measure the significance of the transverse and longitudinal impact parameters, respectively, to create discriminants that recognize tracks associated with displaced vertices [215, 216].

In addition to impact parameter-based algorithms, secondary vertex reconstruction algorithms play an important role in b-tagging. These methods search for displaced vertices formed by the decay of B hadrons. Features such as the invariant mass of tracks associated with the secondary vertex, vertex flight distance, and multiplicity of associated tracks are used to characterize b-jets. In some cases, B hadron decays involve intermediate c-hadrons, leading to the formation of tertiary vertices, which can be included in more advanced tagging methods.

The low-level taggers such as IP2D, IP3D, and the secondary vertex algorithm are then combined into high-level b-tagging algorithms to improve the overall performance. The DL1r algorithm is based on a deep neural network (DNN) architecture and takes the outputs of several low-level taggers as inputs, including IP2D, IP3D, and the secondary vertex algorithm. It also incorporates additional

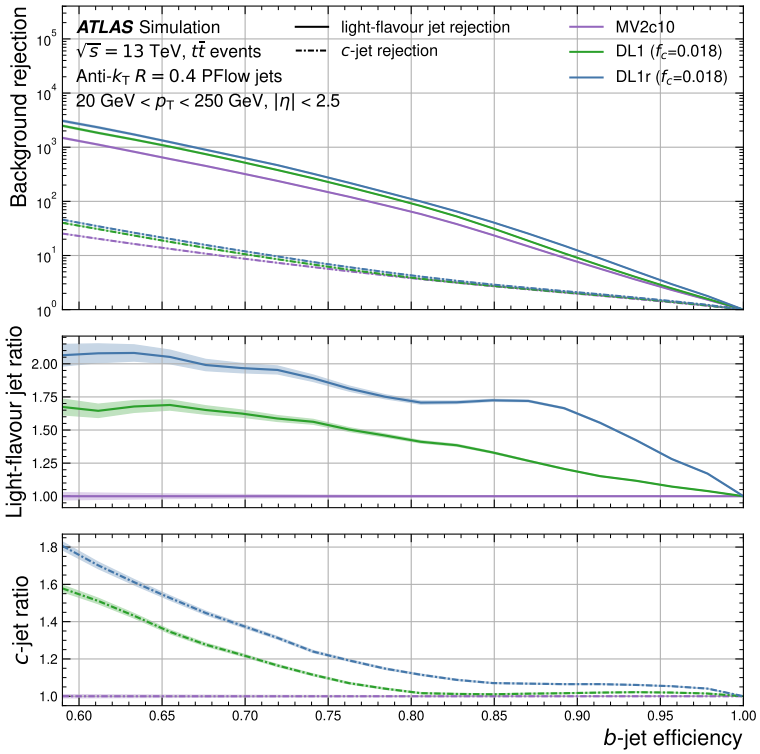


Figure 5.6: Rejection factors for light-flavor jets and c-jets as a function of the b-jet efficiency for three high-level b-taggers, including the DL1r algorithm. The lower two panels display the ratios of the light-flavor jet rejection and c-jet rejection rates of the algorithms compared to MV2c10 algorithm [216]

variables, such as those used in a jet vertex finder algorithm for better c-jet identification and a recursive neural network (RNN) to capture correlations among tracks from the same b-hadron. The DL1r network is trained to classify jets as b-jets, c-jets, or light-flavor jets, and its output consists of three probabilities, one for each jet type. The final discriminant is derived from these probabilities, providing flexibility and improved performance, especially in distinguishing b-jets from c-jets and light-flavor jets.

The performance of b-tagging algorithms is typically quantified by the efficiency for identifying b-jets and the corresponding rejection rates for c-jets and light-flavor jets. The efficiency refers to the fraction of true b-jets correctly tagged as b-jets, while the rejection measures how effectively c- and light-flavor jets are rejected. Different working points are defined based on the b-tagging efficiency, with typical operating points set at 60%, 70%, 77%, and 85% efficiency. These working points represent different trade-offs between b-jet identification and background rejection. The b-tagging performance of three different high-level taggers, including the DL1r algorithm, is shown in Figure 5.6.

To ensure accurate b-tagging in data, scale factors are applied to simulated events. These scale factors account for any discrepancies between the performance of b-tagging algorithms in simulated and real collision data. The scale factors are

derived using control regions in data where the flavor composition of jets is well understood, such as in $t\bar{t}$ events. The scale factors are then applied to simulations to correct the b-jet efficiency and light-flavor jet rejection to match the performance observed in data.

5.5 Hadronic τ-leptons

The τ-leptons can decay either hadronically or leptonically within the ATLAS detector. When they decay leptonically, specialized τ reconstruction is not used, as the decay products appear as electrons or muons, accompanied by missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) due to the presence of neutrinos. A key distinction, however, arises from the finite lifetime of the τ lepton, which causes the τ decay vertex to be displaced from the primary pp collision vertex. This displacement leads to a unique impact parameter distribution for τ leptons compared to prompt electrons or muons. In contrast, hadronic decays, referred to as $\tau_{\text{had-vis}}$, result in jets typically composed of charged and neutral pions.

The reconstruction of $\tau_{\text{had-vis}}$ starts with jets reconstructed using the anti-k_t algorithm with a radius parameter of $R = 0.4$. These jets are required to have transverse momentum $p_{\mathrm{T}} > 10$ GeV and pseudorapidity $|\eta| < 2.5$. Jets in the transition region, $1.37 < |\eta| < 1.52$, are excluded. The energy of the $\tau_{\text{had-vis}}$ candidate is initially determined from the summed energy of topological clusters within a cone of $\Delta R = 0.2$. The jet energy is calibrated using a local hadronic calibration technique [217].

The τ production vertex is selected by summing the transverse momenta of tracks within $\Delta R = 0.2$ around the $\tau_{\text{had-vis}}$ jet. The vertex with the largest momentum sum is chosen. Tracks associated with the $\tau_{\text{had-vis}}$ candidate must have $p_{\mathrm{T}} > 1$ GeV, at least two hits in the Pixel detector, and at least seven hits in the SCT and TRT. Additionally, transverse and longitudinal impact parameters relative to the τ-vertex must satisfy $|d_0| < 1.0$ mm and $|z_0 \sin\theta| < 1.5$ mm.

The number of associated tracks categorizes the $\tau_{\text{had-vis}}$ candidate into either a 1-prong or 3-prong decay of the τ-lepton. After determining the number of tracks, a dedicated energy calibration is applied to the $\tau_{\text{had-vis}}$ candidate to align the reconstructed energy with the generated values from simulations.

To distinguish hadronically decaying τ-leptons from other jets in the detector, an RNN is applied. Separate networks are trained for 1-prong and 3-prong $\tau_{\text{had-vis}}$ decays. The RNN uses a mix of low- and high-level features that include both track and cluster variables, as well as the overall jet shape and information about decay vertex properties. The training is performed using simulated $\gamma^* \to \tau\tau$ samples as signal and dijet events as background. The RNN output is designed to remain stable across varying $\tau_{\text{had-vis}}$ transverse momentum and pileup conditions.

The output of the RNN is used to define four WPs for $\tau_{\text{had-vis}}$ identification: *Tight*, *Medium*, *Loose*, and *VeryLoose*. Each WP offers a different balance of signal

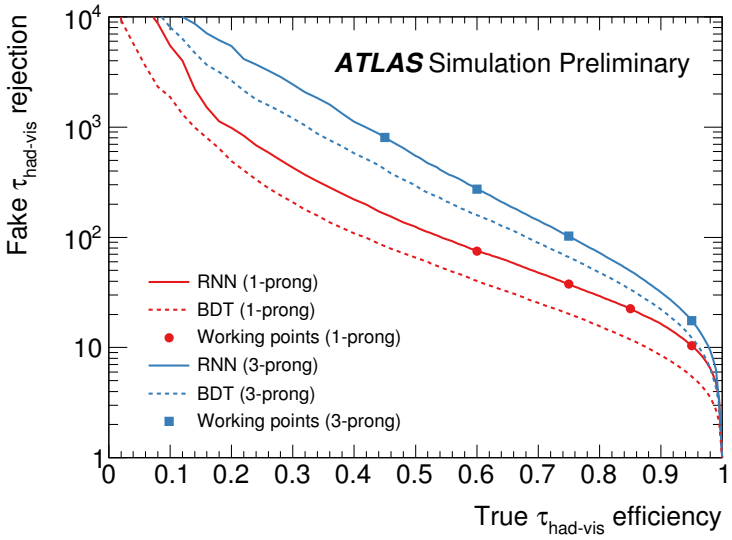


Figure 5.7: The rejection for quark and gluon jets misidentified as $\tau_{\text{had-vis}}$ is presented as a function of the actual $\tau_{\text{had-vis}}$ efficiency. The curves for 1-prong (red) and 3-prong (blue) $\tau_{\text{had-vis}}$ candidates are displayed, using the RNN (solid line) and BDT (dashed line) identification algorithms. The markers represent the four defined working points: *Tight*, *Medium*, *Loose*, and *VeryLoose* [217].

efficiency and background rejection. Efficiencies for 1-prong $\tau_{\text{had-vis}}$ decays are 60%, 75%, 85%, and 95% for the *Tight*, *Medium*, *Loose*, and *VeryLoose* WPs, respectively. Corresponding efficiencies for 3-prong decays are slightly lower, at 45%, 60%, 75%, and 95%. The rejection rates for quark and gluon jets misidentified as $\tau_{\text{had-vis}}$ are shown in Figure 5.7 as a function of the true $\tau_{\text{had-vis}}$ efficiency for the RNN and Boosted Decision Tree (BDT) identification algorithms.

A separate multivariate discriminant based on BDTs is also constructed to reject backgrounds arising from electrons misidentified as 1-prong $\tau_{\text{had-vis}}$ candidates. This discriminant is known as the *electron-veto BDT* (eBDT). This discriminant is built using information from the calorimeter and the tracking detector. Information from the TRT system is critical for optimizing the performance of this discriminant.

The performance of τ-lepton identification, reconstruction, and triggering is relevant for analyses involving τ-lepton decays. These performance metrics are evaluated using a tag-and-probe method in events enriched with $Z \to \tau_\mu \tau_{\text{had}}$ decays. The tag-and- technique allows for a detailed study of the efficiency of τ-lepton identification, triggering, and reconstruction, using events triggered by the muon and containing a hadronically decaying tau in the final state. This method provides a data-driven framework for evaluating both online and offline algorithms and for deriving scale factors to correct for discrepancies between data and simulation.

The identification and energy reconstruction algorithms are studied using the $Z \to \tau\tau$ process. The offline τ-lepton identification efficiency correction factors are measured with uncertainties of approximately 5% for 1-prong decays and 6% for 3-prong decays for $p_{\text{T}} > 20$ GeV and averaging over η. For the online τ-lepton ID efficiency, precision varies between 3% and 20%, depending on the E_{T} of the

τ-lepton candidate identified by the offline algorithms.

In addition to $Z \to \tau\tau$ events, the $t\bar{t} \to b\mu\nu_\mu b\tau\nu_\tau$ process, where one τ-lepton decays hadronically and the other decays to a muon, provides complementary information to evaluate the τ-performance algorithm's efficiencies at higher p_{T} values. The τ-leptons from top quark decays exhibit a harder p_{T} spectrum than those from Z boson decays. The online τ-lepton ID efficiency in $t\bar{t}$ events is measured with a precision ranging from 8% to 14%, depending on the p_{T} of the hadronically decaying τ-lepton.

5.6 Missing transverse momentum

The missing transverse momentum ($\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$) is a vector quantity defined in the transverse plane of the detector, used to account for particles that are not directly detected. Such particles include neutrinos or other weakly interacting particles, which escape the detector without interacting. Due to the conservation of momentum in the transverse plane, the sum of all visible momenta should ideally be zero. Any non-zero value indicates the presence of undetected particles. The magnitude of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is referred to as the missing transverse momentum, $E_{\mathrm{T}}^{\mathrm{miss}}$.

The calculation of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ includes contributions from both the hard and soft terms [218]:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\Bigg(\underbrace{\sum_{\text{electrons}} \vec{p}_{\mathrm{T}}^{\,e} + \sum_{\text{photons}} \vec{p}_{\mathrm{T}}^{\,\gamma} + \sum_{\tau-\text{leptons}} \vec{p}_{\mathrm{T}}^{\,\tau} + \sum_{\mu} \vec{p}_{\mathrm{T}}^{\,\mu} + \sum_{\text{jets}} \vec{p}_{\mathrm{T}}^{\,\text{jet}}}_{\text{hard term}} + \underbrace{\sum_{\substack{\text{unused} \\ \text{tracks}}} \vec{p}_{\mathrm{T}}^{\,\text{track}}}_{\text{soft term}} \Bigg). \quad (5.3)$$

The hard term consists of reconstructed and calibrated objects such as electrons, photons, hadronically decaying τ-leptons, jets, and muons. The soft term accounts for energy deposits not associated with these hard objects and can be constructed using tracks from the inner detector or energy clusters from the calorimeter. In the ATLAS experiment, the track-based soft term is often used due to its robustness against pile-up, as it includes only tracks associated with the hard-scatter vertex.

In simulated events, the performance of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is evaluated by comparing data with MC predictions. For processes like $Z \to \mu^+\mu^- +$ jets, where the true $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is expected to be small, mismodeling can arise due to detector limitations, jet mismeasurements, or pile-up. Studies of these processes allow an assessment of the detector response and resolution of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$.

The resolution of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is characterized by the RMS of its x and y components with respect to their expected values. It is often studied as a function of the scalar

sum of transverse energies (ΣE_{T}) of the reconstructed objects in $Z \to \mu^+\mu^-$ events. At low ΣE_{T}, the resolution is dominated by muon momentum resolution, while at high ΣE_{T}, the resolution is driven by the JER.

Systematic uncertainties on $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ arise from both hard and soft terms. Hard term uncertainties are propagated from the reconstructed objects, while soft term uncertainties, particularly those arising from pile-up, are evaluated by studying soft term components in both data and MC.

Chapter 6

Differential cross-section measurements of the Higgs boson production in the $H \to \tau\tau$ decay channel

The precise measurement of the Higgs boson production cross-section in the $H \to \tau\tau$ decay channel is crucial for testing the Standard Model predictions. This process allows for a direct determination of the Higgs boson's Yukawa coupling to τ leptons, a key parameter in the SM. Given the high precision with which the τ lepton mass is known [219], the corresponding Yukawa coupling can be calculated accurately, and its experimental measurement provides a stringent test of the SM's electroweak sector. Among the lepton species, the τ lepton's mass, significantly larger than that of the electron and muon, makes this channel the most sensitive probe for the Higgs-lepton interaction due to its larger Yukawa coupling. Furthermore, the $H \to \tau\tau$ decay channel provides a unique sensitivity to the measurement of the Higgs boson's vector boson fusion production mechanism, characterized by the distinctive signature of two high-energy forward jets. The $H \to \tau\tau$ decay channel provides the highest sensitivity among Higgs boson decay channels for probing this production mode.

6.1 Analysis motivation and strategy

Since the discovery of the Higgs boson by the ATLAS and CMS collaborations in 2012, studying its properties has been a major priority for LHC experiments. The decay into a $\tau\tau$ pair has the largest branching ratio among the leptonic decay channels, making it a primary mode for investigating the Higgs boson's interactions with leptons.

Early evidence for the $H \to \tau\tau$ decay was established during Run 1. ATLAS achieved a significance of 4.5σ [220], while the combined ATLAS and CMS results

confirmed the decay with a significance of 5.5σ [50]. Subsequent analyses using the Run 2 dataset have shifted focus towards more precise measurements. These include measurements of the total cross-section for the $H \to \tau\tau$ process, cross-sections for the dominant production modes, and differential cross-sections based on key kinematic variables, such as the Higgs transverse momentum (p_{T}^{H}), jet multiplicity, and dijet invariant mass (m_{jj}) [221, 222].

The $H \to \tau\tau$ channel has provided the most precise single measurements of VBF production. In the first measurement using the full Run 2 dataset, ATLAS measured VBF production in this decay channel with an observed significance of 5.3σ, in line with the expected SM value, and determined a cross-section of $0.90^{+0.20}_{-0.17}$ times the SM prediction [222]. Similar results were reported by CMS, which measured the VBF cross-section to be 0.81 ± 0.17 times the SM expectation [223].

The first ATLAS and CMS measurements using the full Run 2 also expanded their scopes to include measurements within the STXS framework. This approach partitions the Higgs boson phase space into multiple regions based on event kinematics, allowing for a more detailed investigation of production modes. Following the STXS framework, the analysis measured the cross-sections in nine fiducial volumes, six of them targeting ggF production at different p_{T}^{H} regimes and with varying jet multiplicities, and three targeting respectively VBF, VH, and $t\bar{t}H$ production. The results of this measurement are shown in Figure 6.1. The measured cross-sections are in good agreement with the SM predictions. In the fiducial volume targeting $qq \to Hqq$ production with $m_{\mathrm{jj}} > 350$ GeV (measuring mainly VBF production assuming SM-like Higgs boson production), the $\sigma \times \mathcal{B}(H \to \tau\tau)$ was measured to be $0.78^{+0.21}_{-0.18}$ times the SM prediction, showing the potential for high precision for differential cross-section measurements in a VBF enriched phase space.

The measurements presented in this chapter, and that constitute one of the main results of this thesis, extend the previous ATLAS analyses expanding the scope of the STXS results to include differential cross-section measurement in more STXS bins. The analysis presented here categorizes VBF events into multiple regions in order to measure the VBF production cross-section as a function of p_{T}^{H} and m_{jj} in eight mutually exclusive bins.

In addition to the detailed study of VBF production, these new results extend the analysis of $t\bar{t}H$ production within the STXS framework. The previous inclusive measurement in the fully hadronic final state has now been refined by measuring cross-sections in multiple p_{T}^{H} bins, which are expected to be more sensitive to potential deviations from the SM in higher momentum regions. The measurement of $t\bar{t}H$ production in the $H \to \tau\tau$ decay where the τ-leptons decay leptonically are not part of this measurement, since they are targeted by other analyses of the ATLAS collaboration [224].

Additionally, fiducial differential cross-section measurements have been performed for the $H \to \tau\tau$ decay channel. Unlike the STXS framework, these fiducial measurements do not distinguish between different Higgs boson production modes.

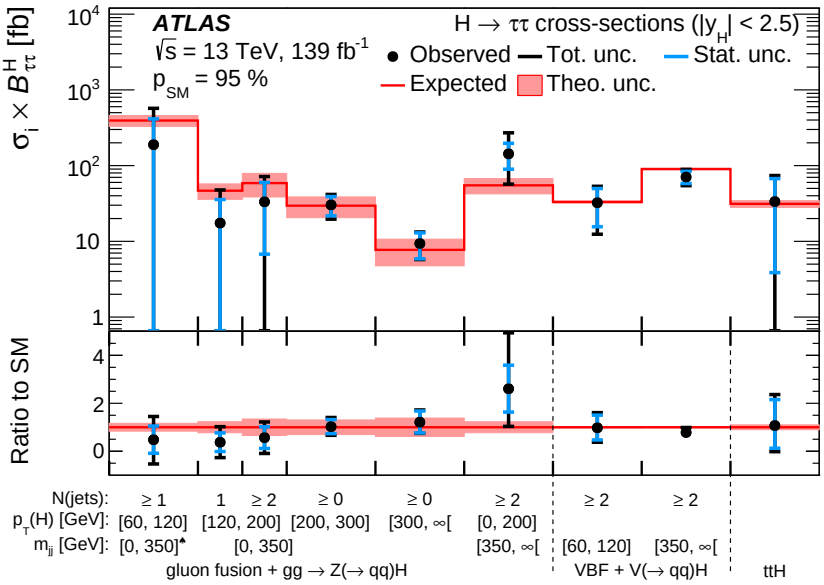


Figure 6.1: Measured values for $\sigma \times \mathcal{B}(H \to \tau\tau)$ in the nine fiducial volumes defined in the STXS measurement in the first ATLAS analysis targeting $H \to \tau\tau$ using the full Run 2 dataset. The total (statistical) uncertainty in the measurement is indicated by the black (blue) error bars. The continuous lines indicate the SM predictions. [222].

Instead, a phase space enriched in VBF events is defined to ensure optimal sensitivity to BSM effects in the VBF production. This less model-dependent approach, while still relying on SM simulations, measures the cross-section as a function of different variables. These measurements are not only useful for probing SM Higgs kinematics, but also offer sensitivity to potential new physics scenarios, such as anomalous CP properties in the Higgs boson couplings.

These new results are derived from an analysis of the full Run 2 dataset, which includes 140 fb^{-1} of integrated luminosity collected at $\sqrt{s} = 13$ TeV.

The $H \to \tau\tau$ analysis presents a series of challenges from the experimental perspective. The final state of the signal process can be easily mimicked by background processes, such as $Z \to \tau\tau$, with a considerably larger cross-section. The τ-lepton is an unstable particle that usually decays before reaching the detector, leading to a complex final state with neutrinos that escape detection and complicate the Higgs boson reconstruction, leading to poorer resolution in the reconstructed mass and transverse momentum. Furthermore, several objects can be misidentified as τ-leptons. Following sections will describe how these challenges are addressed in the analysis to ensure a robust measurement of the Higgs boson production cross-section.

The two measurements of the analysis follow the same strategy. The events are classified according to the τ-lepton decay modes into three channels: $\tau_e\tau_\mu$, $\tau_{\text{lep}}\tau_{\text{had}}$, and $\tau_{\text{had}}\tau_{\text{had}}$. The $\tau_e\tau_e$ and $\tau_\mu\tau_\mu$ channels are excluded due to their lower branching ratios and sensitivity, due to the overwhelmingly large backgrounds from $Z \to \ell\ell$ decays.

The $Z \to \tau\tau$ process represents the largest background in all channels, sharing a similar final state to the signal but differing in the τ-lepton spin correlations and invariant mass $m_{\tau\tau}$. Accurate estimation of the Z+jets background is essential, and data-driven methods are utilized to reduce dependence on Monte Carlo simulations.

Additional backgrounds arise from events with misidentified hadronic τ-lepton decays, electrons, or muons. These are also estimated using data-driven techniques. In the $\tau_e\tau_\mu$ channel, the focus is on misidentified jets as electrons or muons. In the $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ channels, the misidentification of jets as hadronic τ-leptons is addressed.

Minor background contributions, such as those from $t\bar{t}$ and diboson production, are taken from simulations. For the $t\bar{t}$ background, the normalization is corrected using data.

For the STXS measurement, the signal regions are defined to target different bins from the Stage 1.2 of the STXS framework. The six ggF and the VH bins that were measured in the previous analysis are included in the new measurement. The $t\bar{t}H$ production is measured in three p_{T}^H bins, and the VBF production is measured in eight kinematic bins defined by the p_{T}^H and m_{jj}. Multivariate analysis techniques are used in the VH, $t\bar{t}H$, and VBF signal regions to enhance the sensitivity to the $H \to \tau\tau$ signal. Furthermore, inclusive measurements of the total cross-section times branching ratio for $H \to \tau\tau$ and the production mode cross-sections are performed.

The fiducial differential cross-section measurement is performed in a phase space enriched in VBF events. The fiducial phase space is defined to match the reconstructed-level selection criteria, avoiding large extrapolations and ensuring that the measurement is performed within the acceptance of the analysis. The reconstructed-level objects are unfolded to the particle level in order to compare the measured cross-sections to the SM predictions. Unfolded differential cross-sections are measured as a function of the p_{T}^H, $\Delta\phi_{jj}^{\text{signed}}$ and $p_{\text{T}}(j_0)$. A double-differential cross-section measurement is also presented, measuring the cross-section as a function of p_{T}^H and $\Delta\phi_{jj}^{\text{signed}}$. The unfolded cross-section results are interpreted in the context of the Standard Model Effective Field Theory (SMEFT) framework [225], providing constraints on the Wilson coefficients of the SMEFT operators.

The results presented in this chapter, based on the full Run 2 dataset, corresponding to an integrated luminosity of 140 fb^{-1}, are the legacy results of the ATLAS collaboration for the $H \to \tau\tau$ decay channel with the full Run 2 dataset using pp collisions at $\sqrt{s} = 13$ TeV. They are published in Ref. [226].

6.2 Object definition and trigger selection

To select signal events, the analysis relies on the reconstruction and identification of objects in the final state: electrons, muons, visible hadronic τ-lepton decay products, jets, and missing transverse momentum. The number of electrons, muons,

and hadronic τ-leptons is used to classify events into three distinct channels: $\tau_e\tau_\mu$, $\tau_{\text{lep}}\tau_{\text{had}}$, and $\tau_{\text{had}}\tau_{\text{had}}$. Here, τ_{lep} refers to a τ-lepton that decays leptonically into a final state with either an electron or a muon, while τ_{had} refers to a τ-lepton that decays hadronically.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, both τ-leptons decay hadronically into hadrons and a neutrino, while the $\tau_{\text{lep}}\tau_{\text{had}}$ channel features one τ decaying leptonically and one decaying hadronically. The $\tau_e\tau_\mu$ channel, on the other hand, includes events with one electron and one muon, both from the leptonic decays of τ-leptons. Throughout this discussion, electrons and muons are collectively referred to as light leptons (ℓ).

6.2.1 Trigger selection

Various triggers are applied to select events, and the trigger efficiency in MC simulations is adjusted to match real data using scale factors.

For the $\tau_e\tau_\mu$ channel, both single-lepton and dilepton triggers are employed. The trigger selection requires matching between the trigger objects and the offline lepton objects used in the analysis. A logical OR between single-electron, single-muon, and electron-muon triggers is performed. For electrons and muons, the lowest unprescaled [1] triggers in the 2015 dataset have p_{T} thresholds of 24 GeV and 20 GeV, respectively, which increase to 26 GeV for the 2016–2018 period. These thresholds are adjusted to 25 GeV for electrons and 21 GeV for muons in 2015, rising to 27 and 27.3 GeV, respectively, in 2016–2018. Additionally, an electron-muon (dilepton) trigger with online (offline) thresholds of 17 (18) GeV for electrons and 14 (14.7) GeV for muons is used across the full Run 2 dataset (2015–2018).

Trigger	Data period	Online p_{T} threshold	Offline p_{T} threshold
Single electron	2015	p_{T} (e) > 24 GeV	p_{T} (e) > 25 GeV
	2016-2018	p_{T} (e) > 26 GeV	p_{T} (e) > 27 GeV
Single muon	2015-2018	p_{T} (μ) > 20 GeV	p_{T} (μ) > 21 GeV
		p_{T} (μ) > 26 GeV	p_{T} (μ) > 27.3 GeV
Electron-muon	2015	p_{T} (e) > 17, p_{T} (μ) > 14 GeV	p_{T} (e) > 18, p_{T} (μ) > 14.7 GeV
	2016-2018	p_{T} (e) > 17, p_{T} (μ) > 14 GeV	p_{T} (e) > 18, p_{T} (μ) > 14.7 GeV
Ditau	2015-2018	p_{T} (leading $\tau_{\text{had-vis}}$) > 35 GeV	p_{T} (leading $\tau_{\text{had-vis}}$) > 40 GeV
		p_{T} (subleading $\tau_{\text{had-vis}}$) > 25 GeV	p_{T} (subleading $\tau_{\text{had-vis}}$) > 30 GeV

Table 6.1: The p_{T} thresholds applied to the online and offline electrons, muons, and $\tau_{\text{had-vis}}$ objects.

For the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, the same unprescaled single-lepton triggers are used, with identical p_{T} thresholds and variations across the data-taking periods, as in the $\tau_e\tau_\mu$ channel.

[1] An unprescaled trigger is one that is not subject to a prescale factor, meaning it records all events that pass the trigger requirements without applying a reduction.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, di-tau triggers are used. The hadronically decaying τ-leptons reconstructed offline should match the online $\tau_{\text{had-vis}}$ objects. The di-tau trigger requires the leading τ_{had} candidate to exceed an online (offline) p_{T} threshold of 35 (40) GeV and the subleading $\tau_{\text{had-vis}}$ candidate to exceed 25 (30) GeV. To account for the increased luminosity during 2016–2018, an additional Level-1 jet trigger with a p_{T} threshold of 25 GeV is required. Additionally, the leading jet in this channel must have $p_{\text{T}} > 70$ GeV to ensure the trigger operates within its efficiency plateau. Finally, the leading jet must have $|\eta| < 3.2$.

The trigger p_{T} thresholds applied to the online and offline objects are summarized in Table 6.1.

6.2.2 Object definition

The reconstruction and identification of electrons, muons, τ-leptons, jets, and missing transverse momentum is done following the algorithms described in Chapter 5. The additional requirements on the objects in the $H \to \tau\tau$ analysis are described in this section.

Electrons are required to pass the *Loose* identification criteria, have $p_{\text{T}} > 15$ GeV, and be within the fiducial volume of the ID and electromagnetic calorimeter, excluding the transition region between the barrel and endcap calorimeters ($1.37 < |\eta| < 1.52$). For the $\tau_e\tau_\mu$ and $\tau_e\tau_{\text{had}}$ channels, electrons must pass the *Medium* identification and *Loose* isolation criteria. The efficiency for *Medium* identification ranges from 80% to 90%, and the isolation efficiency ranges from 90% for $p_{\text{T}} = 15$ GeV candidates to above 98% for $p_{\text{T}} = 30$ GeV candidates. In both channels, the transverse momentum threshold is raised by 1 GeV above the nominal trigger threshold to ensure operation within the trigger efficiency plateau, as indicated in section 6.2.1.

Muons must pass the *Loose* identification criteria, corresponding to an efficiency of above 97%, have $p_{\text{T}} > 10$ GeV, and be within $|\eta| < 2.5$. In the $\tau_e\tau_\mu$ and $\tau_\mu\tau_{\text{had}}$ channels, muons must also satisfy a *Tight* isolation criterion, with an efficiency increasing from 85% for $p_{\text{T}} = 10$ GeV to 99% for $p_{\text{T}} = 50$ GeV. For the $\tau_\mu\tau_{\text{had}}$ channel, the transverse momentum requirement is increased to ensure high efficiency in the single-muon trigger. Similar tightening is applied in the $\tau_e\tau_\mu$ channel.

In the $H \to \tau\tau$ analysis, only jets with $p_{\text{T}} > 20$ GeV are considered. The JVT is applied to jets with $p_{\text{T}} < 60$ GeV and $|\eta| < 2.5$ to suppress those not associated with the primary vertex, and the fJVT requirement is used for jets with $p_{\text{T}} < 60$ GeV in the forward region ($|\eta| > 2.5$). b-tagged jets are identified using the DL1r algorithm, with a fixed 85% efficiency working point in the $\tau_e\tau_\mu$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels, and a 70% efficiency working point in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, corresponding to rejection factors for b-tagged jets from c-quarks and light partons of 2.6 and 29, respectively, for the 85% working point.

The $\tau_{\text{had-vis}}$ candidates are required to have $p_{\text{T}} > 20$ GeV and $|\eta| < 2.47$, excluding the barrel-endcap transition region. The charge of $\tau_{\text{had-vis}}$ candidates, defined as the sum of charges of associated tracks, must be ± 1. The $\tau_{\text{had-vis}}$ candidates are classified as 1-prong or 3-prong, based on the number of associated tracks. The RNN described in section 5.5 is used to distinguish $\tau_{\text{had-vis}}$ from quark- or gluon-initiated jets, and the eBDT described also in section 5.5 is used to suppress electrons misidentified as $\tau_{\text{had-vis}}$. The $\tau_{\text{had-vis}}$ candidates are required to pass the *Medium* identification criteria, with efficiencies of 75% and 60% for 1-prong and 3-prong decays, respectively.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, $\tau_{\text{had-vis}}$ candidates are required to match the $\tau_{\text{had-vis}}$ trigger objects, with the transverse momentum threshold tightened to ensure operation within the trigger plateau efficiency. In the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, the $\tau_{\text{had-vis}}$ candidate with the highest transverse momentum is selected, while others are treated as jets. A *Medium eBDT* requirement is applied to 1-prong $\tau_{\text{had-vis}}$ candidates, with an efficiency of 85%. Additionally, the transverse momentum threshold is raised to $p_{\text{T}} > 30$ GeV to enhance background rejection.

The missing transverse momentum vector, $E_{\text{T}}^{\text{miss}}$, is reconstructed as indicated in section 5.6.

6.3 Higgs mass and transverse momentum reconstruction

The event categorization, the discrimination between signal and background events, and the signal extraction in the $H \to \tau\tau$ analysis rely on the accurate reconstruction of the Higgs boson mass and transverse momentum. This section describes the methods used to estimate these quantities.

6.3.1 Higgs mass reconstruction

The reconstruction of the invariant mass of the di-τ system plays an important role in analyses targeting the $H \to \tau\tau$ decay, as it helps discriminate between the signal events and irreducible background contributions, such as those arising from $Z \to \tau\tau$ processes. Since the decay of the Higgs boson into τ-leptons produces unstable particles whose decays involve undetected neutrinos, reconstructing the full mass of the system requires dealing with invisible particles. These undetected neutrinos complicate the direct measurement, and thus specialized techniques must be employed to estimate the mass of the di-τ system. This section introduces and explains the three approaches used in this analysis for reconstructing the invariant mass: the visible mass method and the collinear approximation, and a more advanced approach called the Missing Mass Calculator (MMC).

The visible mass method offers a straightforward technique for estimating the di-τ system's mass, but it makes a significant simplification by only accounting

for the visible decay products of the τ-leptons, ignoring the contribution from the missing energy carried by neutrinos. As a result of this omission, the reconstructed mass is shifted toward lower values compared to the true mass. The visible mass, denoted as $m_{\tau\tau}^{\text{vis}}$, is given by the energy and momenta of the two reconstructed visible τ-leptons:

$$m_{\tau\tau}^{\text{vis}} = \sqrt{2p_{\tau_1,\text{vis}}p_{\tau_2,\text{vis}}\left(1 - \cos\Delta\phi_{\tau_1,\tau_2}\right)} \tag{6.1}$$

Here, $p_{\tau_1,\text{vis}}$ and $p_{\tau_2,\text{vis}}$ are the magnitudes of the momenta of the visible products of the τ-lepton decays, and $\Delta\phi_{\tau_1,\tau_2}$ is the azimuthal angle between them. This method is computationally simple and often used as a baseline but does not provide a complete picture since the invisible decay products are neglected.

The collinear approximation offers a more refined approach by incorporating the missing transverse momentum into the mass calculation. This method is based on the assumption that the invisible decay products, namely neutrinos, are emitted in the same direction as the visible decay products. In addition, it assumes that these neutrinos are the sole contributors to the event's $E_{\text{T}}^{\text{miss}}$.

Using this assumption, the fraction of each τ-lepton's momentum carried by its visible decay products can be estimated. These fractions, represented by x_1 and x_2, can be expressed as:

$$x_1 = \frac{p_{x_2}^{\text{vis}}p_{y_1}^{\text{vis}} - p_{y_2}^{\text{vis}}p_{x_1}^{\text{vis}}}{p_{x_2}^{\text{vis}}p_{y_1}^{\text{vis}} + E_x^{\text{miss}}p_{x_1}^{\text{vis}} - p_{y_2}^{\text{vis}}p_{x_1}^{\text{vis}} - E_y^{\text{miss}}p_{y_1}^{\text{vis}}}, \tag{6.2}$$

$$x_2 = \frac{p_{x_2}^{\text{vis}}p_{y_1}^{\text{vis}} - p_{y_2}^{\text{vis}}p_{x_1}^{\text{vis}}}{p_{x_2}^{\text{vis}}p_{y_1}^{\text{vis}} - E_x^{\text{miss}}p_{x_2}^{\text{vis}} - p_{y_2}^{\text{vis}}p_{x_1}^{\text{vis}} + E_y^{\text{miss}}p_{y_2}^{\text{vis}}}. \tag{6.3}$$

The x and y subscripts refer to the components of the momenta and missing transverse momentum in the transverse plane. Once the momentum fractions x_1 and x_2 are determined, the collinear mass of the di-τ system, $m_{\tau\tau}^{\text{coll}}$, is given by the following relationship, which takes into account the invariant mass of the visible system and the fractions of the momenta:

$$m_{\tau\tau}^{\text{coll}} = \frac{m_{\tau\tau}^{\text{vis}}}{\sqrt{x_1x_2}}. \tag{6.4}$$

This approach provides a better resolution for events with a boosted topology, where the τ-leptons and their decay products are approximately collinear. However, it tends to overestimate the mass in events where the collinear assumption is invalid, such as when the decay products are produced in a back-to-back configuration. In these cases, the method can yield unphysical results or excessively high mass values due to the incorrect estimation of the momentum fractions.

The Missing Mass Calculator (MMC) is a more sophisticated algorithm that overcomes the limitations of both the visible mass method and the collinear approximation. Unlike the collinear approximation, the MMC does not rely on the

assumption that the neutrinos are emitted along the same direction as the visible decay products, making it applicable to a wider range of event topologies. The MMC attempts to reconstruct the most likely kinematic configuration of the full di-τ system, including the invisible neutrinos, by maximizing a likelihood function based on probability density functions derived from τ-decay kinematics.

In this method, the six to eight unknown variables required to fully describe the event depend on the τ-lepton decay mode. These variables include the components of the invisible momenta $\vec{p}^{\,\text{miss}}_{1(2)}$ of the neutrinos from each τ decay, as well as the invariant mass of the invisible system $m^{\text{miss}}_{1(2)}$ for each leptonically decaying τ-lepton. The system of equations used to estimate these variables is derived from the mass-shell constraints:

$$\begin{aligned}
E_x^{\text{miss}} &= p_1^{\text{miss}} \sin\theta_1^{\text{miss}} \cos\phi_1^{\text{miss}} + p_2^{\text{miss}} \sin\theta_2^{\text{miss}} \cos\phi_2^{\text{miss}} \\
E_y^{\text{miss}} &= p_1^{\text{miss}} \sin\theta_1^{\text{miss}} \sin\phi_1^{\text{miss}} + p_2^{\text{miss}} \sin\theta_2^{\text{miss}} \sin\phi_2^{\text{miss}} \\
m_{\tau_1}^2 &= \left(m_1^{\text{miss}}\right)^2 + \left(m_1^{\text{vis}}\right)^2 + 2E_1^{\text{vis}}E_1^{\text{miss}} - 2p_1^{\text{vis}}p_1^{\text{miss}} \cos\left(\theta_1^{\text{vis}} - \theta_1^{\text{miss}}\right) \\
m_{\tau_2}^2 &= \left(m_2^{\text{miss}}\right)^2 + \left(m_2^{\text{vis}}\right)^2 + 2E_2^{\text{vis}}E_2^{\text{miss}} - 2p_2^{\text{vis}}p_2^{\text{miss}} \cos\left(\theta_2^{\text{vis}} - \theta_2^{\text{miss}}\right),
\end{aligned} \tag{6.5}$$

where $E^{\text{vis}}_{1(2)}$ and $p^{\text{vis}}_{1(2)}$ are the energies and momenta of the visible products of the τ-leptons decay, $m^{\text{vis}}_{1(2)}$ are the masses of the visible products of the τ-leptons decay, $m^{\text{miss}}_{1(2)}$ are the invariant masses of the neutrino systems, and $\theta^{\text{vis}}_{1(2)}$ ($\phi^{\text{vis}}_{1(2)}$) and $\theta^{\text{miss}}_{1(2)}$ ($\phi^{\text{miss}}_{1(2)}$) are the polar (azimuthal) angles of the visible and invisible products of the τ-leptons decay.

For τ-leptons that decay hadronically, only one neutrino is produced, and thus $m^{\text{miss}}_{1(2)}$ is set to zero. Even with these mass-shell constraints, the system remains underdetermined. The MMC addresses this using probability density functions that depend on the τ-decay kinematics, which are estimated from Monte Carlo simulations. The algorithm scans over the unknown parameters of the neutrino momenta and chooses the most likely configuration based on the likelihood of each solution, effectively optimizing the mass reconstruction.

The MMC algorithm's performance is sensitive to the resolution of the missing transverse momentum $E_{\text{T}}^{\text{miss}}$, which introduces further challenges. To account for this, the algorithm performs additional scans over the components of the missing transverse momentum, E_x^{miss} and E_y^{miss}, with each scan point weighted according to a Gaussian probability based on the scalar sum of transverse energy in the calorimeter.

The MMC algorithm estimates the di-τ mass using three methods: the maximum weight (MAXW), the most likely mass (MLM), and the most likely neutrino momenta (MLNU3P). These methods rely on phase space sampling through a Markov Chain Monte Carlo (MCMC) approach, which generates phase space points weighted by their likelihood.

The MAXW estimator determines the di-τ mass as the value corresponding to the single phase space point with the highest likelihood. While this method

provides a coherent estimate of both the neutrino momenta and the mass, it can be less reliable in cases where multiple phase space points have similar probabilities due to challenging event kinematics.

The MLM method improves on this by utilizing all sampled phase space points. It records the di-τ masses from each point in a histogram and defines the mass as the maximum of this histogram. The histogram is automatically weighted by the likelihoods of the sampled points through the MCMC process, resulting in a more robust mass estimate. However, the MLM method does not associate specific neutrino momenta with the mass.

The MLNU3P method adopts a similar approach to MLM but focuses on the neutrino momenta instead of the di-τ mass. Histograms of the neutrino momentum components are constructed, and the most likely values are taken from the maxima of these histograms. The di-τ mass is then reconstructed using these momenta, providing both mass and neutrino momenta estimates. However, treating the momentum components independently can introduce inconsistencies.

The di-τ mass $m_{\tau\tau}^{\text{MMC}}$ estimate used in the analysis is determined with the MLM method, which provides the most robust $m_{\tau\tau}^{\text{MMC}}$ estimate [227].

While the MMC typically provides the most accurate mass reconstruction in a broad range of event topologies, it can occasionally fail in about 1% of events, particularly those with large fluctuations in the $E_{\text{T}}^{\text{miss}}$ measurement. In such cases, the algorithm may not converge to a solution.

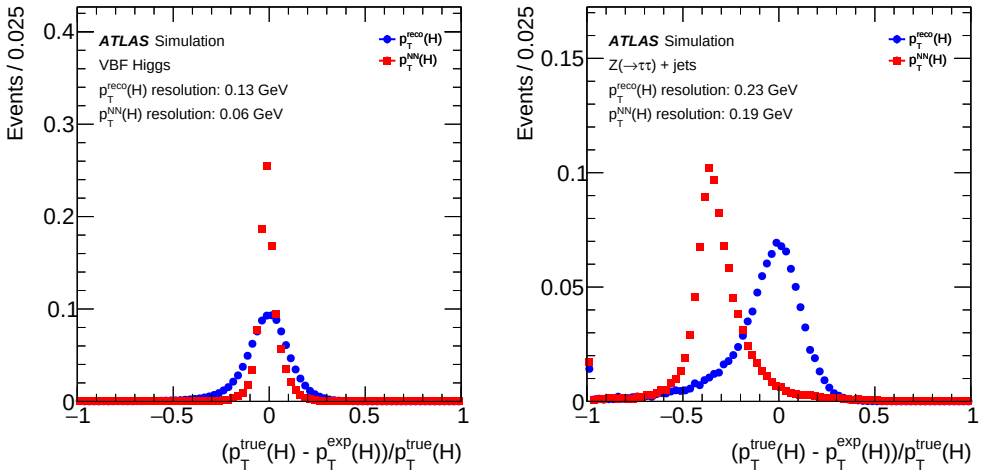
The MMC is the primary method used in the $H \to \tau\tau$ analysis to build the discriminants used to extract the signal strength in the statistical analysis. The signal extraction is described in section 6.8. In the cases where the MMC does not converge to a valid solution, the collinear approximation is used as a fallback method to estimate the di-τ mass. The MMC method fails to converge in approximately 1.3% of signal events in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, 0.7% in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, and 0.2% in the $\tau_e\tau_\mu$ channel. These events are recovered using the collinear approximation.

6.3.2 Transverse momentum reconstruction

In the $H \to \tau\tau$ analysis, the transverse momentum of the Higgs boson, p_{T}^H, plays a crucial role in defining various categories within the STXS framework. This variable is particularly important as it can potentially exhibit sensitivity to BSM effects. Traditionally, the Higgs boson candidate has been reconstructed using the four-momenta of the τ-leptons and the $E_{\text{T}}^{\text{miss}}$.

In this analysis, a neural network approach has been introduced to enhance the reconstruction of p_{T}^H. The network utilizes four input variables: the angular separation $\Delta R_{\tau\tau}$, the azimuthal angle difference $\Delta\phi_{\tau\tau}$ between the two τ-leptons, the transverse momentum of the system formed by their four-momenta and $E_{\text{T}}^{\text{miss}}$, and the collinear mass $m_{\tau\tau}^{\text{coll}}$. Although the NN is trained using simulated events from ggF production, it can also be applied to other production modes, such as VBF

and $t\bar{t}H$. The NN-based reconstruction demonstrated a significant improvement in resolution, even for events from the VBF and $t\bar{t}H$ production modes, making retraining the NN unnecessary for these modes.



(a) (b)

Figure 6.2: Comparison of the p_{T}^{H} resolution relative to the truth-level simulation for the p_{T}^{H} reconstructed from the sum of the four-momenta of the τ-leptons and the $E_{\mathrm{T}}^{\mathrm{miss}}$ versus the NN-based reconstruction for (a) VBF-produced Higgs bosons and (b) $Z(\rightarrow \tau\tau)$ + jets processes. Since the neural network uses the invariant mass of the $\tau\tau$ system and is specifically optimized for $H \rightarrow \tau\tau$ decays, it results in a bad reconstruction of p_{T} for processes such as $Z(\rightarrow \tau\tau)$ + jets [226].

The NN significantly improves the resolution of p_{T}^{H} compared to the traditional method. This enhancement is evident when comparing the truth and reconstructed p_{T}^{H} distributions, where the NN approach shows a resolution improvement. The NN-reconstructed p_{T}^{H} and the p_{T}^{H} reconstructed from the sum of the τ-leptons' four-momenta and $E_{\mathrm{T}}^{\mathrm{miss}}$ are compared in Figure 6.2a for Higgs boson events produced via the VBF mechanism. The NN-based reconstruction shows a narrower distribution, with an improvement in resolution of approximately 50% compared to the traditional method. Since the NN uses the invariant mass of the $\tau\tau$ system, it is able to constrain the $E_{\mathrm{T}}^{\mathrm{miss}}$, leading to a more precise p_{T}^{H} estimation compared to the traditional method.

While the NN offers considerable advantages in terms of resolution, especially for signal events, it introduces some misreconstruction in background processes like $Z \rightarrow \tau\tau$ + jets. This effect is illustrated in Figure 6.2b, which shows the p_{T}^{H} resolution for $Z \rightarrow \tau\tau$ + jets events. The NN-based reconstruction results in a narrower distribution for the $Z \rightarrow \tau\tau$ + jets, but it also introduces a bias in the p_{T}^{H} estimation, since it is specifically optimized for Higgs boson events. Nonetheless, the overall improvement in p_{T}^{H} reconstruction, especially for Higgs production events, demonstrates the effectiveness of this novel technique in the $H \rightarrow \tau\tau$ analysis, where a precise p_{T}^{H} estimation is required to classify the events into different STXS categories that aim to measure differential Higgs boson production cross-sections as a function of p_{T}^{H}.

6.4 Multivariate analysis

Event categorization in the $H \to \tau\tau$ analysis strongly relies on machine learning techniques to enhance sensitivity, improve signal-to-background separation, and optimize signal purity in signal regions. Three distinct multivariate algorithms are employed: two BDTs designed to enhance sensitivity to the VBF and VH production modes, and a BDT multiclassifier used to define signal-pure regions for measuring the $t\bar{t}H$ production mode. The characteristics of these classifiers are described in this section, while their application in event categorization is discussed in Section 6.5.

6.4.1 The VBF tagger

The distinctive signature of VBF production in H events allows for the creation of high-purity signal regions. Due to the sizable branching ratio of the $H \to \tau\tau$ decay, this final state benefits from the VBF topology, allowing for precise measurements of the VBF production cross-section. However, two major backgrounds must be carefully controlled to fully exploit this sensitivity. The first is the $Z \to \tau\tau$ background with additional jets, particularly from electroweak processes, which can closely resemble the VBF final state. The second is contamination from the ggF Higgs production, which can mimic the VBF topology in the presence of additional jets. Given the larger theoretical uncertainties on the ggF cross-section compared to VBF, this contamination increases the systematic uncertainty in the VBF cross-section measurement.

Due to the different mechanisms behind jet production in VBF and the main backgrounds, the kinematic properties of the two leading jets differ between the signal and background events. These differences are exploited to define a high-purity signal region by training a BDT, which uses the kinematic features of the leading jets as input variables. The BDT is trained with VBF $H \to \tau\tau$ signal samples, while the ggF and $Z \to \tau\tau$ events are treated as backgrounds. In the training, all Monte Carlo weights are applied, and events with negative weights are excluded. The relative normalization between the ggF and $Z \to \tau\tau$ samples is kept equal during training to avoid biasing the BDT towards one of the backgrounds. The three analyses channels, $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$, and $\tau_e\tau_\mu$, are combined in the training.

Seven jet-related kinematic variables are used as inputs for the BDT training:

- $p_{\text{T}}(\text{Hjj})$: the transverse momentum of the system formed by the two τ candidates, missing transverse momentum, and two leading jets.
- $\eta^{j_0} \times \eta^{j_1}$: the product of the pseudorapidity of the leading and sub-leading jets.
- p_{T}^{jj}: the transverse momentum of the di-jet system.

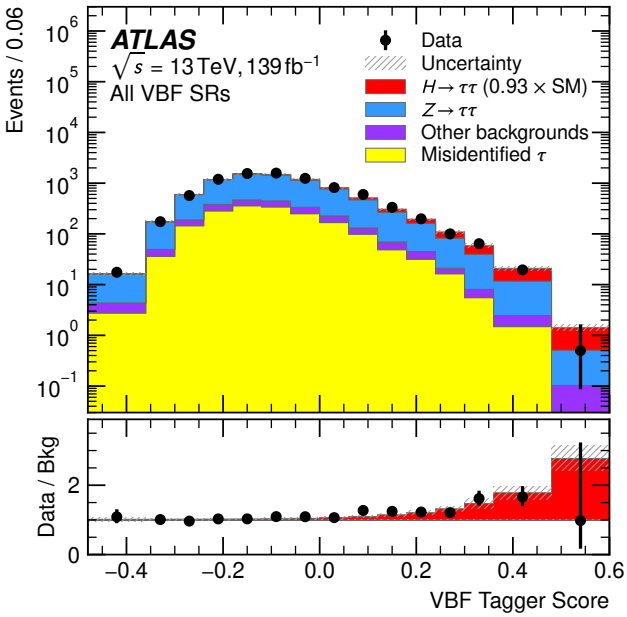


Figure 6.3: Distribution of the BDT discriminant used to separate $H \to \tau\tau$ events produced through VBF from ggF $H \to \tau\tau$ events and $Z \to \tau\tau$ events [222]. The events in this distribution are required to pass the VBF inclusive preselection criteria, explained in Section 6.5.

- $\Delta\eta_{jj}$: the pseudorapidity separation between the two jets.
- $\Delta\phi_{jj}$: the difference in azimuthal angle between the two jets.
- m_{jj}: the invariant mass of the di-jet system.
- $p_{\mathrm{T}}^{j_1}$: the transverse momentum of the sub-leading jet.

One concern with using machine learning classifiers is the potential for introducing biases that affect the signal mass reconstruction. This occurs when the classifier modifies the shape of the mass distribution used in the likelihood fit. Although sculpting the mass variable is not problematic if both data and simulations are affected similarly, it is generally preferred to preserve the original mass shape, even after applying the classifier. To mitigate this effect, careful selection of input variables remains critical. Focusing on the kinematic properties of the two leading jets, which are a signature of the VBF topology, helps reduce the sculpting effect. The variables selected for the BDT training are not highly correlated with the Higgs boson mass reconstructed with the MMC technique, ensuring that the mass distribution is not sculpted by the BDT.

The BDT training is performed using a four-fold cross-validation scheme to ensure robust performance across data subsets and optimal use of the available statistics. Developed for the full Run 2 analysis targeting $H \to \tau\tau$ events [222], the BDT has been applied in the analysis presented in this chapter. The BDT discriminant is used to classify events into a high-purity VBF signal region and a

region dominated by ggF and $Z \to \tau\tau$ events. The details of event categorization using the BDT are described in Section 6.5.1.2. The distribution of the BDT discriminant in events passing a VBF inclusive preselection is shown in Figure 6.3.

6.4.2 The VH tagger

To optimize the sensitivity to the VH production mode in the $H \to \tau\tau$ analysis, a binary BDT classifier is trained to distinguish VH events from other Higgs production mechanisms. The BDT is trained using a 2-fold cross-validation strategy, with input variables including the m_{jj}, the angular separation between the dijet and di-τ systems ($\Delta\phi_{\text{jj},\tau\tau}$), the transverse momenta of the di-τ system ($p_{\text{T}}^{\tau\tau}$) and the dijet system (p_{T}^{jj}), the ratio of these transverse momenta ($p_{\text{T}}^{\tau\tau}/p_{\text{T}}^{\text{jj}}$), the $E_{\text{T}}^{\text{miss}}$, the pseudorapidity difference between the two jets ($\Delta\eta_{\text{jj}}$), and their angular separation (ΔR_{jj}). Additional variables were tested, but their impact on the BDT performance was negligible.

The BDT is trained with VH events as signal and ggF, VBF, and $t\bar{t}H$ events as background. The output of the BDT is used to categorize events into a VH-enriched signal region and a region dominated by background events, which includes not only non-Higgs processes but also Higgs events from other production modes. As for the VBF tagger, this BDT was initially trained for the full Run 2 analysis targeting $H \to \tau\tau$ events [222] and is applied in the current analysis. The process for event categorization using this BDT is detailed in Section 6.5.1.3.

Figure 6.4 shows the distribution of the BDT discriminant for events that pass a VH inclusive preselection. The categorization described in Section 6.5.1.3 is based on this distribution.

6.4.3 The $t\bar{t}H$ multiclassifier

After a $t\bar{t}H$ inclusive preselection, that will be discussed in detail in Section 6.5, events are categorized into regions enriched in $t\bar{t}H$ signal and regions dominated by $Z \to \tau\tau$ and $t\bar{t}$ background events. A multiclass BDT specifically trained for this purpose is employed. The multiclass BDT outputs three distinct scores, one for each of the processes (or *classes*). The usage of a multiclass BDT allows the simultaneous discrimination between the $t\bar{t}H$, $Z \to \tau\tau$, and $t\bar{t}$ processes, allowing not only the separation of the $t\bar{t}H$ signal from the backgrounds in signal-enriched regions, but also the definition of high-purity control regions for the $Z \to \tau\tau$ and $t\bar{t}$ processes. These control regions are used to estimate the normalization of the backgrounds in the simultaneous fit of all the regions to the data.

The input variables for the BDT can be grouped into four categories. A first set of variables describes different jet properties, and includes the η of the five leading jets, the scalar sum of the p_{T} of all the jets in the event and the scalar sum of the p_{T} of all the b-tagged jets. It also includes the m_{jj} of the best W-boson candidate (built as the pair of non-b-tagged jets in the event with the closest

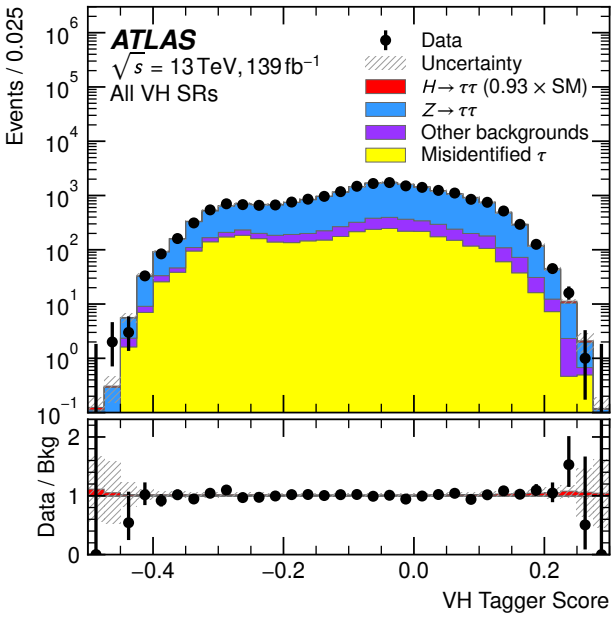


Figure 6.4: Distribution of the BDT discriminant used to separate $H \to \tau\tau$ events produced through VH from background events [222]. The events in this distribution are required to pass the VH inclusive preselection criteria, explained in Section 6.5.

invariant mass to the m_W value) and the invariant mass of the three jets coming from the best top-quark candidate (built as the system of the W candidate and a b-tagged jet in the event with the closest invariant mass to the m_t value). A second set of variables describes the τ-leptons, including the p_{T} of the di-τ system and the p_{T} of the leading and sub-leading τ-leptons. Two additional variables describe the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event, including the $E_{\mathrm{T}}^{\mathrm{miss}}$ itself and the smallest azimuthal angle between the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and a τ-lepton. Finally, a set of variables related to angular distances between objects in the event is added, including the minimum angular separation between jets, the minimum angular separation between a b-tagged jet and a τ-lepton, and the $\Delta\eta$ and ΔR between the two τ-leptons.

Two separate BDTs are trained to optimize performance in different regions of p_{T}^H. A low p_{T}^H model is trained with events where $p_{\mathrm{T}}^H < 200$ GeV, while a high p_{T}^H model handles events with $p_{\mathrm{T}}^H > 200$ GeV. This approach is necessary due to the different background compositions in each region, with most of the $t\bar{t}$ background concentrated at low p_{T}^H. The BDT trained for high p_{T}^H shows better rejection of $t\bar{t}$ events in that region compared to a model trained inclusively.

Each of the output scores from the multiclass BDT indicates the likelihood of the event belonging to one of the three classes: $t\bar{t}H$, $Z \to \tau\tau$, and $t\bar{t}$. Cuts on the distributions of these scores are used to define signal- and background-enriched regions. It is important to note that the sum of the three BDT outputs is normalized to one: $\mathrm{BDT}_{t\bar{t}H} + \mathrm{BDT}_{Z} + \mathrm{BDT}_{t\bar{t}} = 1$. The normalized BDT scores for the $t\bar{t}H$, $Z \to \tau\tau$, and $t\bar{t}$ classes for events with $p_{\mathrm{T}}^H < 200$ GeV and $p_{\mathrm{T}}^H \geq 200$ GeV are shown in Figure 6.5.

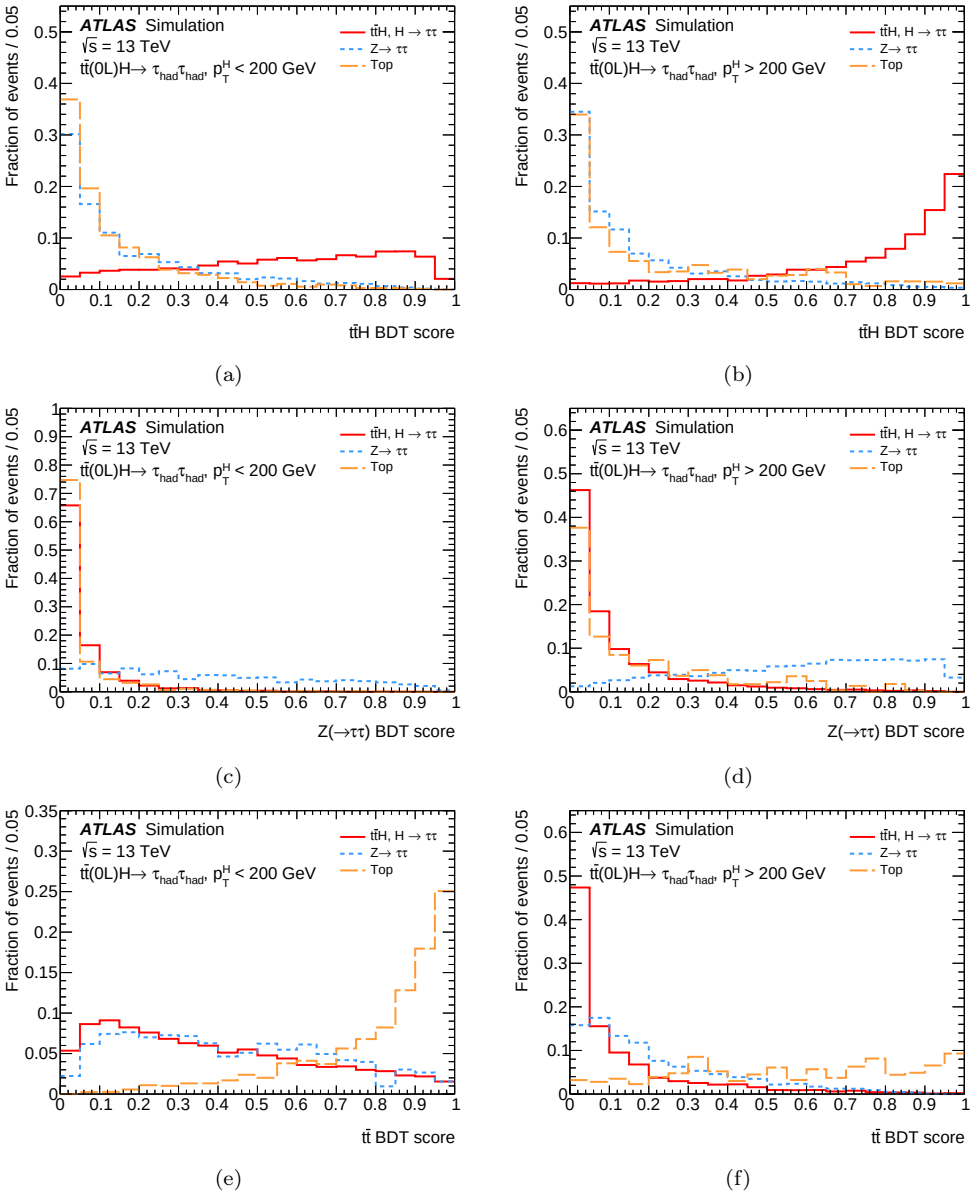


Figure 6.5: Distributions of the scores of the multiclass BDT training used for the event categorization in the $t\bar{t}H$ region. The BDT output scores for the $t\bar{t}H$ (a,b), $Z \to \tau\tau$ (c,d) and $t\bar{t}$ (e,f) classes are shown for $p_{\mathrm{T}}^{H} < 200$ GeV (a,c,e) and $p_{\mathrm{T}}^{H} \geq 200$ GeV (b,d,f).

6.5 Event selection and categorization

The event selection process for this analysis is conducted in two primary stages. Initially, events are assigned to one of three mutually exclusive channels, which are determined by the final states resulting from the τ-lepton decays. Once classified into a channel, events are further subdivided into specific categories, with the goal of isolating production mechanisms and measuring differential cross-sections as a function of different kinematic properties. The details of the selection criteria for both channel assignment and categorization are discussed in the following subsections.

Each channel is subject to specific selection criteria aimed at improving the sensitivity to the Standard Model Higgs boson signal. The selection criteria for the three channels are summarized in Table 6.2.

	$\tau_e\tau_\mu$	$\tau_{\text{lep}}\tau_{\text{had}}$		$\tau_{\text{had}}\tau_{\text{had}}$
		$e\tau_{\text{had}}$	$\mu\tau_{\text{had}}$	
Preselection				
Object counting	# of $e = 1$, # of $\mu = 1$, # of $\tau_{\text{had-vis}} = 0$	# of $e/\mu = 1$, # of $\tau_{\text{had-vis}} = 1$		# of $e/\mu = 0$, # of $\tau_{\text{had-vis}} = 2$
p_{T} cut	e/μ : p_{T} cut 10 GeV to 27.3 GeV	e/μ : p_{T} cut 21 GeV to 27.3 GeV, $\tau_{\text{had-vis}}$: $p_{\text{T}} > 30$ GeV		$\tau_{\text{had-vis}}$: $p_{\text{T}} > 40$,30 GeV
ID, Isolation, and eveto	e/μ: Medium e: FCLoose, μ: FCTightTrackOnly	e/μ: Medium, $\tau_{\text{had-vis}}$: RNN Medium e: FCLoose, μ: FCTightTrackOnly 1-prong $\tau_{\text{had-vis}}$: eleBDT e-veto		$\tau_{\text{had-vis}}$: RNN Medium
Charge product	Opposite charge	Opposite charge		Opposite charge
Kinematics	$m_{\tau\tau}^{\text{coll}} 66$ GeV $30 < m_{e\mu} < 100$ GeV	$m_{\text{T}} < 70$ GeV		
b-veto	# of b-jets = 0 WP: DL1r_FixedCutBEff_85	# of b-jets = 0 WP: DL1r_FixedCutBEff_85		# of b-jets = 0 WP: DL1r_FixedCutBEff_70 not applied in tt(0L)$H \to \tau_{\text{had}}\tau_{\text{had}}$
$E_{\text{T}}^{\text{miss}}$	$E_{\text{T}}^{\text{miss}} > 20$ GeV	$E_{\text{T}}^{\text{miss}} > 20$ GeV		$E_{\text{T}}^{\text{miss}} > 20$ GeV
Leading jet	$p_{\text{T}} > 40$ GeV	$p_{\text{T}} > 40$ GeV		$p_{\text{T}} > 70$ GeV, $\lvert\eta\rvert < 3.2$
Angular	$\Delta R_{e\mu} < 2.0$, $\lvert\Delta\eta_{e\mu}\rvert < 1.5$	$\Delta R_{\ell\tau_{\text{had-vis}}} < 2.5$, $\lvert\Delta\eta_{\ell\tau_{\text{had-vis}}}\rvert < 1.5$		$0.6 < \Delta R_{\tau_{\text{had-vis}}\tau_{\text{had-vis}}} < 2.5$ $\lvert\Delta\eta_{\tau_{\text{had-vis}}\tau_{\text{had-vis}}}\rvert < 1.5$
Coll. app. x_1/x_2	$0.1 < x_1 < 1.0$, $0.1 < x_2 < 1.0$	$0.1 < x_1 < 1.4$, $0.1 < x_2 < 1.2$		$0.1 < x_1 < 1.4$, $0.1 < x_2 < 1.4$

Table 6.2: Summary of the event selection for all sub-channels in the $H \to \tau\tau$ analysis.

In the $\tau_e\tau_\mu$ channel, events are required to contain exactly one reconstructed electron and one reconstructed muon, each meeting the criteria defined in Section 6.2.2. The product of the two lepton charges must be negative. Additionally, the collinear mass is required to be above 66 GeV, and the invariant mass of the $e\mu$ system must lie between 30 and 100 GeV. Further background reduction is achieved by rejecting events that contain b-tagged jets. Angular cuts on the separation between the two leptons ($\Delta R_{e\mu}$ and $|\Delta\eta_{e\mu}|$) are imposed to optimize the selection. Additional criteria involving $E_{\text{T}}^{\text{miss}}$ and the momentum fractions of the visible τ-lepton decays (x_0 and x_1, for the leading and subleading τ decay products, respectively) are applied to refine the reconstruction of the system's invariant mass. Moreover, events must have a leading jet with p_{T} above 40 GeV to reduce background contributions, since the final state from the Higgs boson signal typically involves at least one energetic jet.

For the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, events are required to have one light lepton (either an electron or a muon) and one $\tau_{\text{had-vis}}$, both of which must fulfill specific selection criteria. The charges of the lepton and the $\tau_{\text{had-vis}}$ must have opposite signs.

Additionally, the transverse mass of the lepton and $\tau_{\text{had-vis}}$ system is required to be larger than 70 GeV to reject W+jets backgrounds. A b-tagged jet veto is also imposed. Like the $\tau_e\tau_\mu$ channel, angular separation criteria between the lepton and the $\tau_{\text{had-vis}}$ ($\Delta R_{\ell\tau_{\text{had-vis}}}$ and $|\Delta\eta_{\ell\tau_{\text{had-vis}}}|$) are imposed. To reduce background contamination from $Z \to ee$ decays misidentified as τ decays, an electron veto based on the eBDT described in Section 5.5 is applied to the $\tau_{\text{had-vis}}$ candidates in events where a one-prong τ-lepton decay is reconstructed. This electron veto aims to prevent these events, which can mimic true $H \to \tau\tau$ decays. Additional requirements on $E_{\text{T}}^{\text{miss}}$ and the x_0 and x_1 variables are imposed to further enhance the mass reconstruction performance. The same $p_{\text{T}} > 40$ GeV requirement for the leading jet is applied in this channel as well.

In the $\tau_{\text{had}}\tau_{\text{had}}$ channel, events must contain exactly two hadronically decaying τ-leptons. The transverse momentum thresholds for the two $\tau_{\text{had-vis}}$ candidates are dictated by the lowest unprescaled ditau trigger used during data collection. Given the high event rate for low-p_{T} τ-lepton candidates, additional criteria were introduced to reduce the trigger rate, as discussed in Section 6.2.1. These include constraints on the angular separation between the two $\tau_{\text{had-vis}}$ objects ($\Delta R_{\tau_{\text{had-vis}}\tau_{\text{had-vis}}} > 0.6$) and a requirement for at least one central jet ($|\eta| < 3.2$) with $p_{\text{T}} > 70$ GeV. As in the other channels, the charges of the two $\tau_{\text{had-vis}}$ candidates must be opposite. A veto on events containing b-tagged jets is applied, except for the $ttH \to \tau_{\text{had}}\tau_{\text{had}}$ category, where top-associated Higgs production is considered. To improve the reconstruction of the system's invariant mass, additional criteria on $E_{\text{T}}^{\text{miss}}$ and the visible tau momentum fractions (x_0 and x_1) are implemented.

6.5.1 Event selection for the STXS measurement

Once the events are classified into one of the three channels, they are further categorized. Firstly, the topology of the events is used to classify them into different categories that target specific Higgs production mechanisms: VBF, VH, $t\bar{t}H$, and ggF (boosted topologies). Then, the events are further divided into categories based on kinematic properties, such as the transverse momentum of the Higgs boson, p_{T}^{H}, and the invariant mass of the dijet system, m_{jj} or the number of jets. These divisions are designed to target specific regions of the phase space defined by the STXS framework, and where the differential cross-sections are measured. Further division of the events is performed in some categories using the multivariate classifiers described in Section 6.4.1, 6.4.2, and 6.4.3.

The definition of the categories targets phase space regions defined by the STXS framework, which performs a simplified fiducial volume definition and a categorization of the final states based on the kinematic properties of the Higgs boson production and the associated objects. The STXS framework, already introduced in Section 1.5, and the reduced STXS categorization used in this measurement are described in Section 6.5.1.1. The categorization of the reconstructed events is described in detail in Sections 6.5.1.2, 6.5.1.3, 6.5.1.4 and 6.5.1.5.

6.5.1.1 STXS Stage 1.2 in the $H \to \tau\tau$ analysis

The STXS framework was developed collaboratively by ATLAS, CMS, and the theoretical community to enhance the precision of Higgs boson cross-section measurements while minimizing theoretical assumptions and dependencies [63]. Conventional methods, such as signal strength and coupling strength measurements, use kinematic properties predicted by the SM and rely on extrapolating from the observable phase space to the total phase space. This approach introduces model-dependent theoretical uncertainties, as both coupling strengths and global signal strengths are normalized to SM predictions. The STXS framework targets cross-section measurements in well-defined phase space regions, aligned with different Higgs boson production modes.

These STXS bins serve as the primary signal templates in the analysis. Defined with a focus on reducing theoretical uncertainties, the bins help isolate effects possibly linked to new physics while enhancing experimental sensitivity. Each bin represents exclusive phase space regions in specific production modes and is optimized to reduce theoretical model dependence that would otherwise affect coupling strength measurements. Although individual bins are configured for distinct measurements, they can be grouped for analyses with limited statistical power to provide broader measurement insights.

The STXS methodology simultaneously fits signal expectations in all defined bins, providing cross-section measurements for each fiducial bin. The measurements can be interpreted then in Effective Field Theory (EFT) models, evaluating the compatibility of the obtained results with BSM scenarios (see for example Ref. [228]), or providing scaling factors for coupling strengths for the Yukawa couplings of the Higgs boson. These measurements can also contribute to longer-term analyses by including the full covariance matrix among STXS bins.

In the STXS framework, the fit model unfolds observed events across all decay channels into these defined bins, requiring well-specified definitions of final-state particles at truth or particle level. Simplified particle definitions provide a consistent approach across different theoretical predictions, including both analytic calculations and simulations.

To introduce the definition of the simplified fiducial volume in the STXS framework (and later, the fiducial volume definition for the unfolded fiducial cross-section measurements in Section 6.5.2.1), it is necessary to define the concepts of the objects at particle and reconstruction level. Particle-level objects are defined as stable particles in the MC simulation. These objects are not subject to the detector simulation, and the hadronization of the quarks is included. Particle-level objects can also be referred to as truth-level objects. The common definition of the simplified fiducial volume for all the analyses performing STXS measurements allows for a consistent combination of the results from different analyses, since they all target the same fiducial phase space.

The reconstruction-level objects are defined as the objects after the simulation

of the interaction of the particles with the detector. The reconstruction-level objects are the ones used in the event selection and the definition of the analysis categories. The unfolding between the reconstruction-level and particle-level objects in the STXS framework is performed implicitly at the statistical analysis level: different signal templates are used for different STXS bins, and the corresponding signal strengths for each STXS bin are extracted from the fit, that is performed on distributions built from the reconstruction-level objects.

For the Higgs boson selection, only on-shell states are considered, with a rapidity cut of $|y_H| < 2.5$, as current analyses are not sensitive to regions beyond this range. Leptons, including electrons and muons from vector boson decays, are taken as *dressed*, incorporating final-state photons in their momentum, while τ-leptons are defined by the sum of their decay products, with no restrictions on transverse momentum or pseudorapidity for fully leptonically decaying weak vector bosons.

Production Mode	N_{jets}	p_{T}^{H} (GeV)	m_{jj} (GeV)
$gg \to H$	1	$120 - 200$	< 200
	≥ 1	$60 - 120$	–
	≥ 2	$120 - 200$	< 350
	≥ 2	< 200	≥ 350
	–	$200 - 300$	–
	–	≥ 300	–
$qq' \to Hqq'$	≥ 2	–	$60 - 120$
	≥ 2	< 200	$350 - 700$
	≥ 2	< 200	$700 - 1000$
	≥ 2	< 200	$1000 - 1500$
	≥ 2	< 200	≥ 1500
	≥ 2	≥ 200	$350 - 700$
	≥ 2	≥ 200	$700 - 1000$
	≥ 2	≥ 200	≥ 1000
ttH	–	< 200	–
	–	$200 - 300$	–
	–	≥ 300	–

Table 6.3: STXS bins in the $H \to \tau\tau$ analysis.

The STXS jet definition employs anti-k_t clustering with a radius $R = 0.4$, considering stable particles with lifetimes above 10 ps and excluding Higgs boson decay products, as these are treated separately. Stable leptons and neutrinos from vector boson decays are excluded, but hadronic decay products from these weak bosons

are incorporated. Truth jets have a minimum transverse momentum threshold of 30 GeV, helping to control phase space extrapolation effects in measurements.

As explained in Section 1.5, the STXS framework partitions the Higgs boson production phase space into several bins, targeting specific production mechanisms: gluon-gluon fusion ($gg \to H$), vector boson fusion and quark-initiated VH production ($qq' \to qq'H$), VH production with leptonic decays ($VH \to (\ell\ell, \ell\nu)H$), and production with a top quark pair ($t\bar{t}H$). The categories are further subdivided based on variables like p_{T}^{H}, the number of jets or m_{jj}. The full split of the STXS bins in the so-called Stage 1.2 can be seen in Figure 1.6.

Analysis targeting just one decay channel, such as the $H \to \tau\tau$ analysis, can only measure a subset of the STXS bins. The STXS framework is designed to be flexible and allow for the measurement of a subset of the bins, merging bins with low sensitivity to maintain reasonable statistical uncertainties in the measured bins. The bin definitions in a concrete STXS stage can evolve over time as more data is collected and the experimental and theoretical uncertainties are reduced, providing finer granularity in the phase space definition and enhancing the sensitivity to new physics.

The $H \to \tau\tau$ STXS measurement presented in this chapter targets 18 STXS bins, defined using the Stage 1.2 framework. In the $gg \to H$ volume, the bins are defined based on the number of jets and the p_{T}^{H} of the Higgs boson. The $H \to \tau\tau$ analysis targets six bins in this volume. Seven bins are defined for the $qq' \to qq'H$ volume, based on p_{T}^{H} and m_{jj}. Finally, three bins are defined for the $t\bar{t}H$ volume.

The definition of the 18 STXS bins as a function of the properties of the final state particles is summarized in Table 6.3.

6.5.1.2 VBF categorization

The VBF category is designed to select events in which the Higgs boson is produced in association with two jets from the hard scattering process. These events can be distinguished from other Higgs production mechanisms and the $Z \to \tau\tau$ + jets background, as they are characterized by two high-p_{T} jets with a large separation in rapidity. Consequently, the invariant mass of the dijet system (m_{jj}) is typically large.

Following the fiducial volume definition of the STXS framework, events in the electroweak qqH bins with large dijet mass must satisfy two conditions: the dijet mass must exceed 350 GeV, and the subleading jet must have $p_{\mathrm{T}} > 30$ GeV.

Additional selection criteria are applied to reduce the $Z \to \tau\tau$ + jets background. The pseudorapidities of the two leading jets must fulfill the condition $\eta_{j0} \cdot \eta_{j1} < 0$, ensuring that the jets are in opposite hemispheres of the detector. Moreover, the absolute pseudorapidity difference ($|\Delta\eta_{jj}|$) between the jets must be greater than 3. The visible τ decay products are also required to lie within the rapidity of the VBF jets.

To achieve sensitivity to the different STXS bins targeted in the measurement, events are further categorized based on the reconstructed p_{T}^{H} using the NN described in Section 6.3.2, and the dijet invariant mass, m_{jj}. The events are divided into eight categories according to these variables: two groups based on p_{T}^{H} ($p_{\mathrm{T}}^{H} < 200$ GeV and $p_{\mathrm{T}}^{H} \geq 200$ GeV), which are further subdivided into four bins according to the dijet mass, $m_{\mathrm{jj}} \in [350, 700, 1000, 1500, \infty)$ GeV.

To improve separation between VBF Higgs events and those from the ggF process or the $Z \to \tau\tau$ + jets background, each event category is split into two subcategories based on the BDT score from the discriminant described in Section 6.4.1. These subcategories are referred to as VBF_0 (low BDT score) and VBF_1 (high BDT score) for each STXS bin. The thresholds for separating the two subcategories are optimized individually for each STXS bin, aiming to maximize the combined significance. The figure of merit used for this optimization is $\sigma_{\mathrm{tot}} = \sqrt{\sigma_0^2 + \sigma_1^2}$, where the significance for each subcategory ($i = 0$ for VBF_0 and $i = 1$ for VBF_1) is defined as $\sigma_i = \frac{N_s}{\sqrt{N_s + N_b}}$, with N_s and N_b representing the number of VBF-produced and background events in the output categories, respectively.

Figure 6.6 shows the distribution of the BDT score for both signal and background events across all the categories, considering all three analysis channels together. The optimal cut values for the BDT score, which vary depending on the STXS bin, are presented in Table 6.4.

STXS bin	1	2	3	4
p_{T}^{H} [GeV]	0-200			
m_{jj} [GeV]	350-700	700-1000	1000-1500	>1500
Cut point	-0.04	0.03	0.13	0.16

STXS bin	5	6	7	8
p_{T}^{H} [GeV]	>200			
m_{jj} [GeV]	350-700	700-1000	1000-1500	>1500
Cut point	0.01	0.07	0.17	0.25

Table 6.4: Optimized cut values for the VBF tagger BDT score in each STXS bin.

6.5.1.3 VH categorization

The sensitivity to the VH production mode, where the Higgs boson is produced in association with a vector boson (either Z or W), is driven by analyses targeting events in which the vector boson decays hadronically. This phase space is targeted in a separate analysis, described in detail in Ref. [229].

The selection of the analysis described in this chapter targets VH events where the vector boson decays hadronically. The selection criteria at the reconstruction

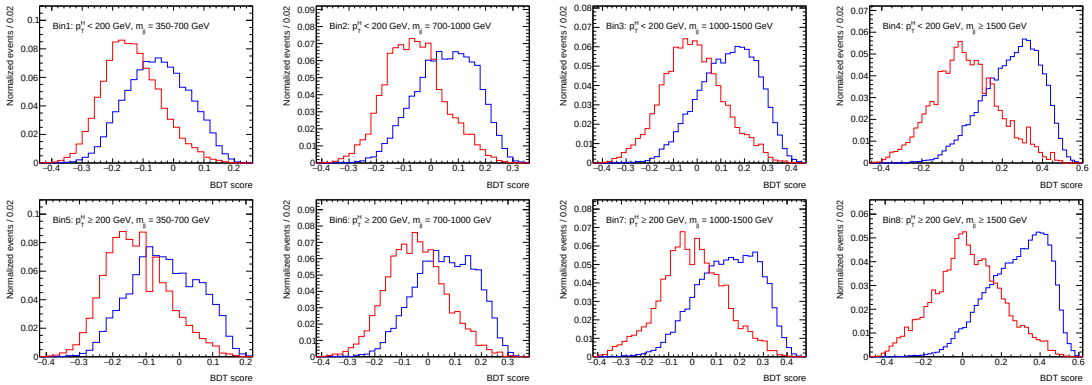


Figure 6.6: VBF tagger BDT score distributions for each STXS bin, with signal (VBF production) shown in red and background ($Z \to \tau\tau$ and ggF production) in blue.

level follow the definition of the STXS fiducial volume, requiring 60 GeV $< m_{\mathrm{jj}} <$ 120 GeV and a subleading jet with $p_{\mathrm{T}} > 30$ GeV.

For events in the $\tau_e\tau_\mu$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels, an additional requirement is placed on the leading jet's p_{T}, which must exceed 40 GeV. The impact of this cut on qqH events in the VH category is found to be negligible, as the leading jet in these events typically has a high transverse momentum. The cut is primarily intended to suppress background contributions.

To improve the separation between VH Higgs events and those from ggF or $Z \to \tau\tau$ + jets processes, the events are further divided into two subcategories, VH_0 and VH_1, based on the output of the VH MVA discriminant described in Section 6.4.2. The optimal cut value (events where the MVA score is above 0.1 are classified as VH_1) to separate these categories is not changed with respect to the previous analysis [222]. The cut value was optimized trying different values and performing statistical fits using Asimov datasets without including systematic uncertainties to minimize the expected statistical uncertainty in the determination of the VH signal (therefore maximizing the significance to the signal).

6.5.1.4 $t\bar{t}H$ categorization

The selection of events in the $t\bar{t}H$ channel closely follows the criteria used in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel, with the primary difference being the inversion of the b-jet veto. The final state in the $t\bar{t}H$ category typically includes six jets, two of which are expected to originate from b-quark hadronization. To allow for flexibility in the selection, ≥ 5 jets (with ≥ 2 b-tagged jets) or ≥ 6 jets (with ≥ 1 b-tagged jets) are required

Therefore, the event selection for the $t\bar{t}H$ category requires either six jets with $p_{\mathrm{T}} > 20$ GeV, including at least one b-tagged jet, or five jets, including at least two b-tagged jets.

Events that meet these selection criteria are further categorized using the $t\bar{t}H$ multiclassifier described in Section 6.4.3. As explained before, the multiclassifier is trained to distinguish between three types of events: $t\bar{t}H$, $Z \to \tau\tau$ and $t\bar{t}$. To define

	p_{T}^{H} bins in GeV		
	<200	[200, 300]	>300
Signal region	$\mathrm{BDT}_{t\bar{t}H}^{\mathrm{low}} > 0.65$	$\mathrm{BDT}_{t\bar{t}H}^{\mathrm{high}} > 0.65$	
$Z \to \tau\tau$ control region	$\mathrm{BDT}_{t\bar{t}H}^{\mathrm{low}} < 0.65$ $\mathrm{BDT}_{Z}^{\mathrm{low}} > 0.2$	$\mathrm{BDT}_{t\bar{t}H}^{\mathrm{high}} < 0.65$ $\mathrm{BDT}_{Z}^{\mathrm{high}} > 0.2$	
$t\bar{t}$ control region	$\mathrm{BDT}_{t\bar{t}H}^{\mathrm{low}} < 0.65$ $\mathrm{BDT}_{Z}^{\mathrm{low}} < 0.2$	$\mathrm{BDT}_{t\bar{t}H}^{\mathrm{high}} < 0.65$ $\mathrm{BDT}_{Z}^{\mathrm{high}} < 0.2$	

Table 6.5: Definition of the signal and control regions for the categories as a function of p_{T}^{H}. $\mathrm{BDT}_{t\bar{t}H}^{\mathrm{low}}$ and $\mathrm{BDT}_{Z}^{\mathrm{low}}$ denote the $t\bar{t}H$ and $Z \to \tau\tau$ scores from the low p_{T}^{H} training, while $\mathrm{BDT}_{t\bar{t}H}^{\mathrm{high}}$ and $\mathrm{BDT}_{Z}^{\mathrm{high}}$ denote the $t\bar{t}H$ and $Z \to \tau\tau$ scores from the high p_{T}^{H} training.

the signal-enriched regions, cuts are applied on the $t\bar{t}H$ score, with the optimal values determined by maximizing the expected statistical significance of the signal (the Poisson-Binomial model defined in Ref. [230] is used as a figure of merit). The $Z \to \tau\tau$ control region is defined by inverting the $t\bar{t}H$ score cut and requiring a high $Z \to \tau\tau$ score, while the top control region is defined by inverting both the $t\bar{t}H$ and $Z \to \tau\tau$ score cuts. Different multiclassifiers are used for events above and below 200 GeV in p_{T}^{H}. The summary of the selection criteria for the $t\bar{t}H$ category is presented in Table 6.5. The events in the signal regions are further classified into three different categories according to p_{T}^{H}, with boundaries at 200 and 300 GeV. To perform this classification, the p_{T}^{H} estimated with the NN described in section 6.3.2 is used. This division is motivated by the scope of the analysis, which aims to measure the differential cross-section of the $t\bar{t}H$ Higgs boson production in three different p_{T}^{H} regions.

The signal regions in each of the three p_{T}^{H} bins are further divided into two categories: a sideband region and a window region. The window region contains events with a reconstructed Higgs boson mass within $100 < m_{\tau\tau} < 150$ GeV, while the sideband region contains the remaining events. The sideband region has a lower purity in signal events, but can be used to constrain the background processes in the analysis.

6.5.1.5 Boost categorization

Events that fail the VBF and VH selection criteria are considered for the Boost phase space, which is sensitive to the ggF production mode. The reconstructed p_{T}^{H} is calculated from the visible τ decay products and the $E_{\mathrm{T}}^{\mathrm{miss}}$. Events must satisfy $p_{\mathrm{T}}^{H} > 100$ GeV to enter the boost category.

As per the STXS fiducial definition for the ggF production mode, events in the boost category are further subdivided based on p_{T}^{H} and the number of jets (N_{jets}) with $p_{\mathrm{T}} > 30$ GeV. Specifically, events with $p_{\mathrm{T}} < 200$ GeV are split into 1-jet and $\geq$2-jet categories. For events with $p_{\mathrm{T}}^{H} \geq 200$ GeV, only events with at least one jet are considered, without further dividing by jet multiplicity. The boost categories are the same as in the analysis described in Ref. [222]. The p_{T}^{H} reconstruction is not performed using the NN trained for the analysis round described in this thesis, but estimated summing the four-momentum of the di-τ system and the $E_{\mathrm{T}}^{\mathrm{miss}}$.

The selection criteria for the Boost phase space are summarized in Table 6.6. The selection criteria used to define the inclusive VBF, VH, $t\bar{t}H$, and Boost categories are summarized in Table 6.7. The event yields and signal purity in each signal region (adding together the three analysis channels) are shown in Figure 6.7. The diagonal structure of the signal purity matrix indicates that the signal regions are generally pure in the signal processes that they target. It is possible to observe that the VBF_0 regions are highly contaminated with ggF events with two jets, especially at $p_{\mathrm{T}}^{H} < 200$ GeV. The VBF_1 regions have high purity in VBF events.

$N_{\mathrm{jets}}(p_{\mathrm{T}} > 30$ GeV)	p_{T}^{H} range in GeV			
	100, 120	120, 200	200, 300	300, ∞
Exactly 1	boost_0_1J	boost_1_1J	boost_2_ge1J	boost_3_ge1J
At least 2	boost_0_ge2J	boost_1_ge2J		

Table 6.6: Definition of the six categories in the Boost phase space.

6.5.2 Event selection for the fiducial differential cross-section measurement

To explore the characteristics of VBF Higgs boson production, an alternative method involves measuring the cross-section in a specific fiducial phase space. This phase space is defined by detector acceptance and event selection criteria that primarily focus on VBF production. Unlike process-specific measurements, this approach is inclusive of all production modes, reducing reliance on assumptions about the SM. The cross-section measurements are unfolded to correct for detector inefficiencies, using MC simulations to model VBF, ggF, VH, and $t\bar{t}H$ Higgs production.

The definition of the fiducial phase space, based on particle-level objects, is crucial for the unfolded differential cross-section measurement. This definition is fundamental in order to match the particle-level objects with the reconstructed objects to evaluate the acceptance in the measurement and to perform the unfolding procedure, used to correct the effects of the detector response on the reconstructed objects and to obtain the differential cross-sections at particle level. The definition

VBF inclusive	sub-leading jet $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ $m_{\mathrm{jj}} > 350\,\mathrm{GeV}$, $\lvert\Delta\eta_{jj}\rvert > 3$ $\eta(j_0) \times \eta(j_1) < 0$ lepton centrality: visible decay products of the τ leptons between VBF jets
VH **inclusive**	$60\,\mathrm{GeV} < m_{\mathrm{jj}} < 120\,\mathrm{GeV}$ sub-leading jet $p_{\mathrm{T}} > 30\,\mathrm{GeV}$
$t\bar{t}H$ **inclusive**	# of jets ≥ 6 and # of b-jets ≥ 1 or # of jets ≥ 5 and # of b-jets ≥ 2
Boost inclusive	Not VBF inclusive Not VH inclusive $p_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$

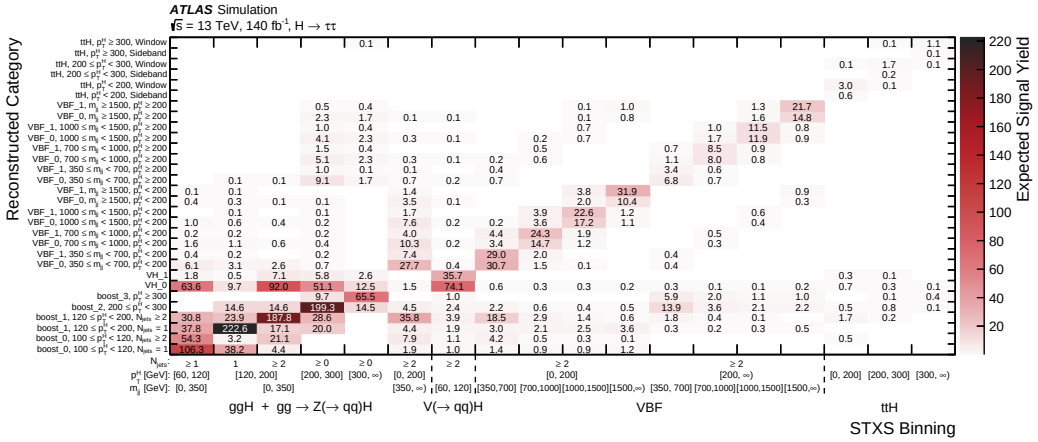
Table 6.7: Summary of the event selection for all sub-channels.

of the fiducial phase space is based on particle-level objects, and the definition of these objets is presented in Section 6.5.2.1.

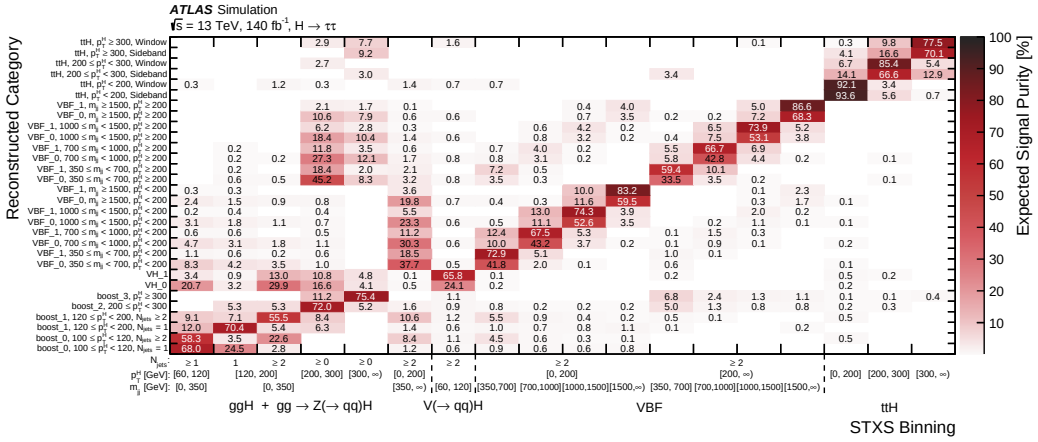
The fiducial differential cross-section in bin i is calculated by accounting for detector effects and background subtraction, as expressed below, and is expressed as:

$$\sigma_i^{\mathrm{fid}} = \frac{1}{\mathcal{L}\epsilon_i} \sum_j \mathcal{M}_{ij}^{-1} f_j \left(N_j^{\mathrm{data}} - N_j^{\mathrm{bkg}} \right), \tag{6.6}$$

where j represents the reconstruction-level bin, $\mathcal{L}$ is the total integrated luminosity, and $\mathcal{M}$ is the migration matrix that connects particle-level bin i with reconstruction-level bin j, accounting for migration between bins due to detector effects. The selection efficiency ϵ_i is defined as the fraction of events passing both fiducial and reconstruction-level selection criteria relative to those passing only the fiducial selection. The correction factor f_j accounts for events outside the acceptance and is given by the ratio of events passing both selections to those passing only the reconstruction-level selection. The observed signal is derived by subtracting the expected background events (N_j^{bkg}) from the total observed events in data (N_j^{data}). The migration matrix, efficiency, and out-of-acceptance corrections are determined using SM Higgs boson MC simulations, including the four main production modes.



(a)



(b)

Figure 6.7: (a) Expected SM $H \to \tau\tau$ signal yield and (b) expected signal purity in each signal region of the analysis (y-axis). Yields are summed over the three $\tau\tau$ channels. Only entries with a value above 0.1 are shown in the plots.

Although the analysis primarily focuses on VBF production, contributions from other production modes are also included as signal processes.

The following sections describe the fiducial phase space definition and the differential variables considered. The results of the fiducial cross-section measurement are presented later this chapter. A detailed description of the profile likelihood unfolding procedure used to extract the differential cross-sections is provided in Appendix A.

6.5.2.1 Fiducial volume definition

The fiducial cross-sections are computed within a well-defined phase space at particle level. This space closely mirrors the inclusive VBF event selection used at reconstruction level, outlined in the next section. In this phase space, approximately 83% of the Higgs signal arises from VBF production, with 17% originating from ggF, as shown in Table 6.8. The combined signal is modeled by summing VBF,

ggF, VH, and $t\bar{t}H$ processes, with their relative contributions fixed according to SM expectations.

	$\tau_e\tau_\mu$	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{had}}\tau_{\text{had}}$
Per Channel			
Object counting	$N_e = 1$, $N_\mu = 1$, $N_{\tau_{\text{truth}}} = 0$	$N_{e/\mu} = 1$, $N_{\tau_{\text{truth}}} = 1$	$N_{e/\mu} = 0$, $N_{\tau_{\text{truth}}} = 2$
p_{T} cut	e/μ : p_{T} cut 10 GeV to 27.3 GeV	e/μ : p_{T} cut 27.0 GeV to 27.3 GeV, τ_{truth}: $p_{\text{T}} > 30$ GeV	τ_{truth}: $p_{\text{T}} > 40$,30 GeV
Kinematics	$m_{\tau\tau}^{\text{coll}} > m_Z - 25$ GeV $30 < m_{e\mu} < 100$ GeV	$m_{\text{T}} < 70$ GeV	
Angular	$\Delta R_{e\mu} < 2.0$, $\lvert\Delta\eta_{e\mu}\rvert < 1.5$	$\Delta R_{\ell\tau_{\text{truth}}} < 2.5$, $\lvert\Delta\eta_{\ell\tau_{\text{truth}}}\rvert < 1.5$	$0.6 < \Delta R_{\tau_{\text{truth}}\tau_{\text{truth}}} < 2.5$ $\lvert\Delta\eta_{\tau_{\text{truth}}\tau_{\text{truth}}}\rvert < 1.5$
x_1 and x_2	$0.1 < x_1 < 1.0$, $0.1 < x_2 < 1.0$	$0.1 < x_1 < 1.4$, $0.1 < x_2 < 1.2$	$0.1 < x_1 < 1.4$, $0.1 < x_2 < 1.4$

Common selection	leading jet $p_{\text{T}} > 40$ GeV, sub-leading jet $p_{\text{T}} > 30$ GeV $E_{\text{T}}^{\text{miss}} > 20$ GeV Opposite charge of τ-decay products $m_{\text{jj}} > 600$ GeV, $\lvert\Delta\eta_{jj}\rvert > 3.4$, $p_{\text{T}}(jj) > 30$ GeV $\eta(j_0) \times \eta(j_1) < 0$ lepton centrality: visible decay products of the τ-leptons between VBF jets $p_{\text{T}}(\text{Hjj}) < 50$ GeV

Table 6.8: Summary of the fiducial particle-level event selection for all channels. The requirements on light-lepton p_{T} and τ_{truth} p_{T} follow those reported in Table 6.2.

At particle level, object definitions rely on stable particles. Electrons and muons are required to originate directly from the hard interaction rather than from hadrons or their decays. Final-state radiation photons within a cone of $\Delta R = 0.1$ around the lepton direction are added to correct the lepton momentum. Hadronic τ-leptons (τ_{truth}) are reconstructed by summing the four-momenta of all stable decay products, excluding neutrinos. All leptons are required to have pseudorapidity $|\eta| < 2.5$, with electrons further restricted to avoid the LAr transition region ($1.37 < |\eta| < 1.52$). Jets are defined by the anti-k_t algorithm with a radius parameter $R = 0.4$, excluding prompt leptons and Higgs decay products. Jets must satisfy $|\eta| < 4.5$ and be separated from τ_{truth} and light leptons by $\Delta R > 0.4$ and $\Delta R > 0.2$, respectively. Missing transverse momentum is computed as the sum of the transverse momenta of all neutrinos not originating from hadrons.

6.5.2.2 Event selection at reconstruction level

The event selection for the fiducial cross-section measurement is based on the inclusive VBF selection criteria described for the STXS measurement. The selection criteria should be as close as possible to the fiducial phase space definition, to

minimize the extrapolations and perform the measurement within the acceptance. Events are required to have at least two jets with $p_{\mathrm{T}} > 30$ GeV. The two jets should be in opposite hemispheres of the detector, and the visible decay products of the τ-leptons should be between the two jets.

In order to reduce as much as possible the ggF contamination in the VBF categories, the cuts on m_{jj} and $\Delta\eta_{jj}$ are tightened with respect to the STXS measurement. The m_{jj} cut is set to 600 GeV, and the $\Delta\eta_{jj}$ cut is increased to 3.4. An additional requirement on the di-jet transverse momentum is added, with $p_{\mathrm{T}}(jj) > 30$ GeV. The transverse momentum of the di-jet system and the Higgs boson is required to be below 50 GeV.

As in the STXS categorization, the VBF events are further divided in two categories, VBF_0 and VBF_1, based on the output of the BDT score. The VBF_0 category is enriched in ggF and $Z \to \tau\tau$ events, while the VBF_1 category has a higher purity in VBF events. The separation between VBF_0 and VBF_1 does not require modeling at the particle level because the unfolded cross-sections are not provided separately for these categories. Instead, the unfolded cross-sections are obtained for the inclusive VBF category, as the inclusive region at reconstruction level is matched to the fiducial phase space at particle level.

6.5.2.3 Differential variables and binning

Differential cross-sections are measured as functions of variables that are sensitive to the kinematics of VBF Higgs production. These measurements aim to both validate the SM and explore possible deviations from it. Among the key jet-related variables is the transverse momentum of the leading jet, $p_{\mathrm{T}}(j_0)$, and the signed azimuthal angle difference between the two jets, $\Delta\phi_{jj}^{\mathrm{signed}}$, ordered by the signed jet rapidity. This angle, $\Delta\phi_{jj}^{\mathrm{signed}}$, is a powerful observable for probing the charge-parity (CP) properties of the Higgs boson. Additionally, the p_{T}^{H} is considered, with the reconstruction performed using the neural network (NN) approach described in Section 6.3.2. In order to improve sensitivity to the CP nature of the Higgs, $\Delta\phi_{jj}^{\mathrm{signed}}$ is also measured as a function of p_{T}^{H}. Binning for these distributions is determined based on the available data statistics, as outlined in Table 6.9.

	$p_{\mathrm{T}}(j_0)$ [GeV]	p_{T}^{H} [GeV]	$\Delta\phi_{jj}^{\mathrm{signed}}$	$\Delta\phi_{jj}^{\mathrm{signed}}$ vs p_{T}^{H} [GeV]
Bin 1	$[40, 95]$	$[0, 110]$	$[-\pi, -\pi/2]$	$\Delta\phi_{jj}^{\mathrm{signed}} < 0$ & $p_{\mathrm{T}}^{H} < 200$
Bin 2	$[95, 130]$	$[110, 150]$	$[-\pi/2, 0]$	$\Delta\phi_{jj}^{\mathrm{signed}} > 0$ & $p_{\mathrm{T}}^{H} < 200$
Bin 3	$[130, 180]$	$[150, 200]$	$[0, \pi/2]$	$\Delta\phi_{jj}^{\mathrm{signed}} < 0$ & $p_{\mathrm{T}}^{H} > 200$
Bin 4	$[180, 500]$	$[200, 550]$	$[\pi/2, \pi]$	$\Delta\phi_{jj}^{\mathrm{signed}} > 0$ & $p_{\mathrm{T}}^{H} > 200$

Table 6.9: Binning of the differential variables.

For all the measured variables, the migration matrix shows a high degree of diagonality, with over 75% of events correctly matched for $p_{\mathrm{T}}(j_0)$, more than 85%

for p_{T}^{H}, greater than 95% for $\Delta\phi_{jj}^{\mathrm{signed}}$, and over 90% for the joint $\Delta\phi_{jj}^{\mathrm{signed}}$ versus p_{T}^{H} distribution. This reflects the relatively coarse binning used in comparison to the experimental resolution. The selection efficiency, ϵ_i, incorporates both object reconstruction efficiencies and event-level selection cuts, with values of 25%, 15%, and 43% for the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$, and $\tau_e\tau_\mu$ channels, respectively. The out-of-acceptance correction factors f_j vary between 74% and 83% depending on the channel.

6.6 Background estimation

The dominant background of the analysis arises from $Z(\to \tau\tau) + \text{jets}$ processes, which contribute approximately 80% of the total background, reaching up to 90% in the most boosted regions of phase space. These processes are modeled using MC simulations, correcting the normalization of the simulation using data in background-enriched control regions. These $Z(\to \tau\tau) + \text{jets}$ control regions are derived selecting $Z(\to \mu\mu) + \text{jets}$ and $Z(\to ee) + \text{jets}$, and applying an embedding technique to modify the selected events to emulate the $Z \to \tau\tau$ decay. The method is described in more detail in Section 6.6.1.1.

The second-largest background contribution, accounting for approximately 10% of the total background, arises from multijet events where jets are misidentified as electrons, muons, or $\tau_{\text{had-vis}}$. These misidentifications are estimated using data-driven methods in both signal and control regions of the analysis. In the $\tau_e\tau_\mu$ channel, the estimate includes cases where jets are misidentified as electrons, muons, or both. For the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel, it covers cases where a jet is misidentified as a $\tau_{\text{had-vis}}$ or where two jets are misidentified as a $\tau_{\text{had-vis}}$ and a light lepton. However, processes involving $W(\to \tau_{\mathrm{had}}\nu) + \text{jets}$ in which a jet is misidentified as a light lepton are modeled separately using MC simulations, contributing only 0.3% of the total background in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel. In the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel, the estimate accounts for cases where one or two jets are misidentified as $\tau_{\text{had-vis}}$. The methods used to estimate these background contributions are discussed in Section 6.6.2.

Top-quark-related processes are a significant source of background in the $\tau_e\tau_\mu$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels, as well as in the $t\bar{t}H$ category of the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. By applying a selection criterion based on the number of b-jets, control regions enriched in top-quark events are defined within the phase space of the analysis.

For the $t\bar{t}H$ category, the background estimation follows a similar approach to the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel, using the same MC samples. However, the additional requirement on the total number of jets and b-jets in this category necessitates separate consideration. The BDT multiclassifier discriminant used to define the signal region is also employed to establish control regions near the signal region in phase space, as described in Section 6.5.1.4.

Diboson processes, estimated using MC simulations, contribute approximately 2.5% to the total background. Additionally, $H \to WW$ events are included in the

analysis, although their contribution is minimal, representing 0.1% of the total background, with the majority (97%) of this contribution contained in the $\tau_e\tau_\mu$ channel.

6.6.1 $Z \to \tau\tau$ background

6.6.1.1 Object-level embedding

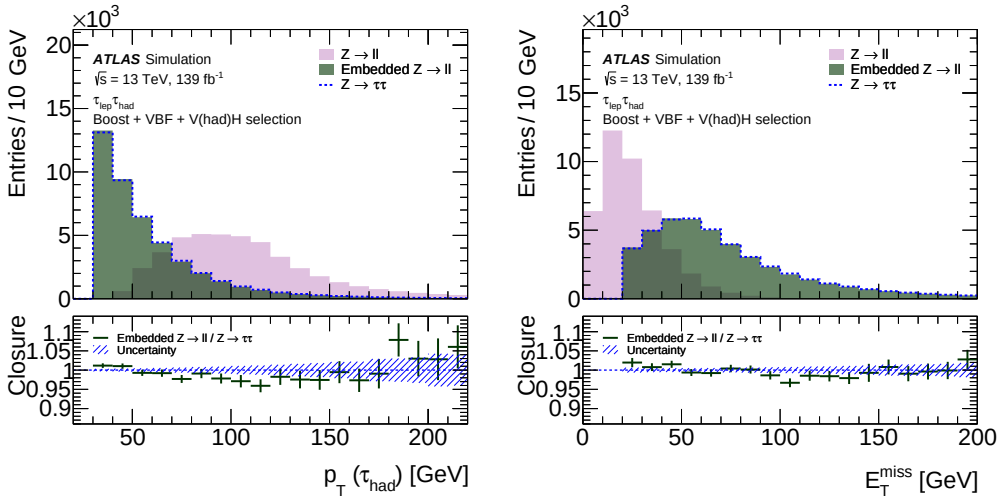
In the $H \to \tau\tau$ analysis, the $Z \to \tau\tau$ background represents a major source of background and requires careful estimation to ensure accurate results. A data-driven technique known as the embedding procedure is used to estimate this background. This method starts by selecting $Z \to \ell\ell$ events (where ℓ refers to either electrons or muons) and modifying their kinematics to simulate the characteristics of $Z \to \tau\tau$ events.

In the embedding procedure, $Z(\to \ell\ell)$ + jets events are selected using single-lepton triggers and are required to contain two oppositely charged light leptons (either electrons or muons). The leptons must pass strict identification and isolation criteria, and the invariant mass of the dilepton system must exceed 80 GeV. Once selected, these events undergo several modifications to transform the kinematics of the leptons into those of τ-leptons. The transverse momenta of the electrons or muons are rescaled using parameterizations derived from simulation, which account for the energy loss due to the presence of neutrinos in the τ decay. Separate parametrizations are used for leptonic and hadronic τ decays.

This rescaling is performed by applying per-lepton weights, which assume collinearity between the τ-lepton and its visible decay products. These weights correct the momentum of the leptons and also account for the reconstruction, identification, and isolation efficiencies of the τ decay products. After rescaling, all event kinematics, including the $E_{\mathrm{T}}^{\mathrm{miss}}$, are recalculated to reflect the new final state. Additionally, the event weights are corrected for trigger efficiencies derived for both single- and dilepton triggers.

Moreover, the embedded $Z \to \ell\ell$ events are split into three distinct subsamples, corresponding to the three primary τ decay modes: leptonic decays (τ_e and τ_μ) and hadronic decays (τ_h). Each subsample is treated separately, ensuring that the τ decay mode is properly accounted for in the analysis. This division prevents overlap in the event selection and ensures that each sample is used appropriately, without double-counting events across different final states.

One of the main innovations of the embedding technique, as applied in this analysis, is the object-level approach, which avoids the complications of embedding at the detector level, a method previously used in Run 1 [220]. In older versions of embedding, the selected leptons from $Z \to \ell\ell$ events were removed at the detector level and replaced with simulated τ-leptons, involving intricate manipulation of energy deposits in the detector cells. This procedure, while powerful, was computationally expensive and complex. In contrast, the current method simplifies



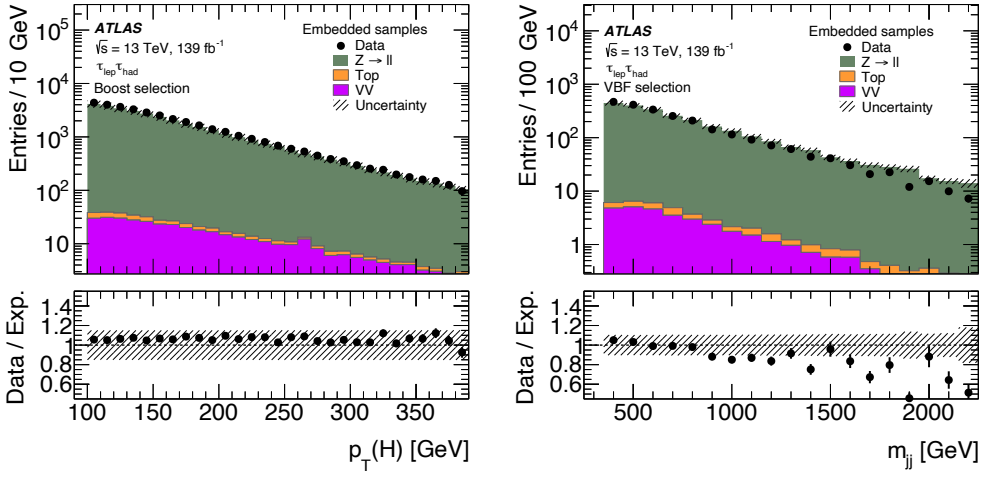
(a) (b)

Figure 6.8: Comparison of kinematic variables for $Z(\to \ell\ell)$ + jets simulated events in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, both before (light purple histogram) and after (dark green histogram) applying the embedding technique, across the combination of the boost, VBF, and VH phase spaces. The distribution from $Z(\to \tau\tau)$ + jets simulated events (dashed blue line) is also included for reference. (a) The p_{T} distribution of the τ_{had} in the event. (b) Distribution of $E_{\text{T}}^{\text{miss}}$. The lower panels illustrate the ratio of the embedded $Z(\to \ell\ell)$ + jets events to $Z(\to \tau\tau)$ + jets events, with error bars representing statistical uncertainties in the ratio and the dashed blue band indicating the statistical uncertainty in the $Z(\to \tau\tau)$ + jets simulation.

the embedding by working with fully reconstructed events, modifying only the p_{T} of the leptons and re-weighting the events to match the properties of $Z \to \tau\tau$ decays.

To ensure the validity of the object-based embedding, the $Z \to \ell\ell$ events processed through this technique are compared with fully simulated $Z \to \tau\tau$ events. Validation is carried out by examining key kinematic distributions such as the p_{T} of the objects or the $E_{\text{T}}^{\text{miss}}$. Two examples of these comparisons are shown in Figure 6.8, which display the p_{T} of the hadronically decaying τ-lepton and the $E_{\text{T}}^{\text{miss}}$. These comparisons show good agreement between the embedded-MC and the simulated $Z \to \tau\tau$ events. No data is shown in these figures.

In order to constrain the normalization of the $Z \to \tau\tau$ background, dedicated control regions are established using embedded $Z \to \ell\ell$ events. These control regions are designed to mimic the kinematic properties of the signal regions but are defined to be orthogonal to them. Specifically, for each signal category, a corresponding control region is constructed by applying the same selection criteria but replacing the τ-leptons with light leptons. The normalization of the $Z \to \tau\tau$ background is then determined by fitting these control regions simultaneously with the signal regions. This approach allows the data to inform the normalization of the $Z \to \tau\tau$ background in the signal region, while still relying on the accurate modeling provided by the embedding technique. The agreement between the embedded data and the embedded simulation in the control regions is shown in



(a) (b)

Figure 6.9: Comparison of Monte Carlo simulation predictions with data in $Z(\to \ell\ell)$+jets-enriched control regions, where the embedding technique is applied to both data and simulation to replicate the $\tau_{\text{lep}}\tau_{\text{had}}$ event selection: (a) p_{T}^{H} in the boosted categories, (b) m_{jj} in the VBF categories. The lower panels display the level of agreement between the embedded data and the embedded simulation. The uncertainty shown is the quadrature sum of the statistical uncertainty from simulated events and the systematic uncertainties of the simulation.

Figure 6.9, which displays the p_{T}^{H} in the boosted categories and the m_{jj} in the VBF categories.

Dedicated uncertainties are assigned to the embedding process, accounting for potential mismodeling of the τ decay kinematics and any residual differences between the embedded $Z \to \ell\ell$ events and the fully simulated $Z \to \tau\tau$ events. These uncertainties are generally small (below 1% in most of the regions) and are propagated throughout the analysis to ensure a robust estimate of the $Z \to \tau\tau$ background.

6.6.1.2 Treatment of electroweak $Z \to \tau\tau$ + jets production

To account for the electroweak contributions of Z+jets processes in the analysis, the SHERPA generator is utilized at leading order. Previous studies have demonstrated that this generator underestimates the electroweak component, particularly in comparison to the HERWIG generator, which provides better agreement with the data. A scale factor, determined by the ratio of the HERWIG to SHERPA results, is applied on the analysis to adjust the electroweak Z + jets contributions accordingly. This is particularly important at high p_{T}^{H} and m_{jj}, regions of phase space explored in the analysis, where the fraction of Z + jets events produced via electroweak processes is significantly larger.

To derive the scale factor, a comparison is made between embedded $Z \to ee$ samples generated with HERWIG, POWHEG, and SHERPA, along with $Z \to \tau\tau$ samples from SHERPA. This analysis in the embedded sample phase space allows

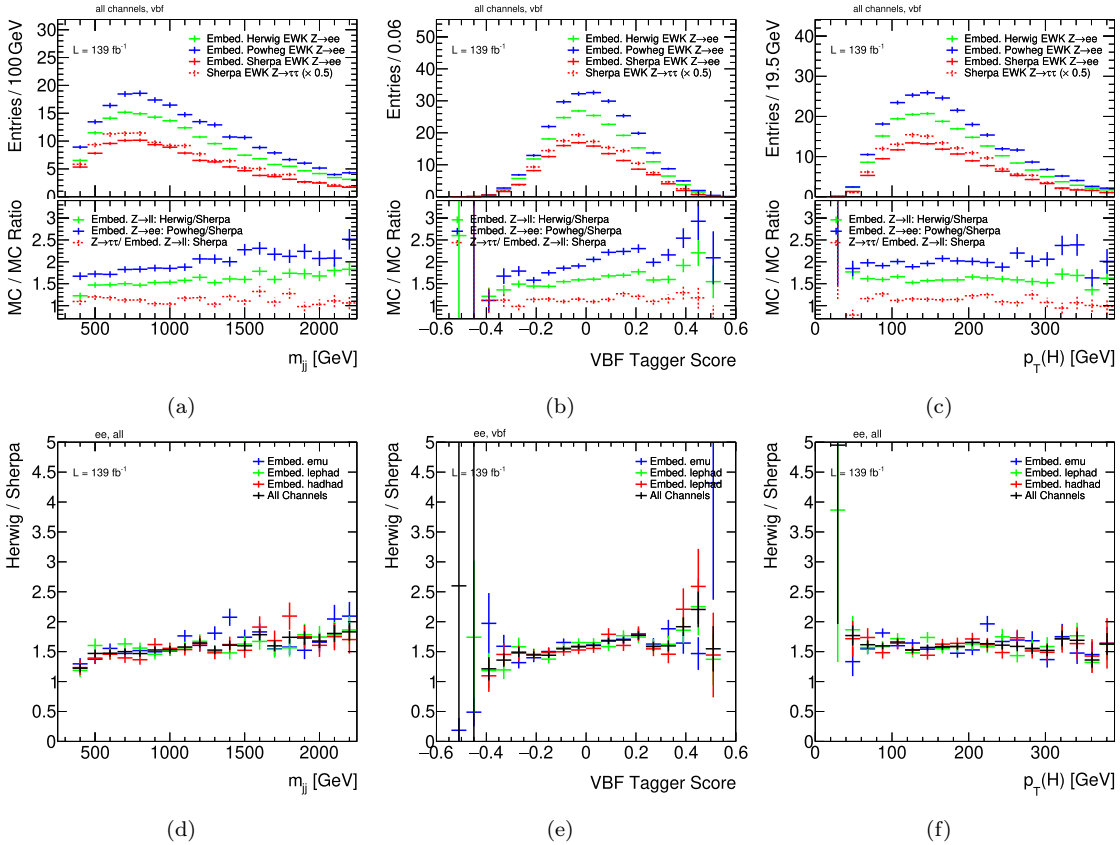


Figure 6.10: Embedded samples for various event generators and the ratio of these generators to SHERPA as a function of m_{jj}, VBF Tagger Score, and p_{T}^{H}. The top three panels show the distributions of $Z \to ee$ samples generated by HERWIG, POWHEG, and SHERPA, as well as SHERPA-generated $Z \to \tau\tau$ samples. Ratios in the top panels are each divided by the SHERPA sample. The bottom three panels show the HERWIG to SHERPA ratio broken down by decay mode. As the ratio plots are relatively constant, a single scale factor may be applied to account for the difference in generator prediction.

for an accurate calculation of the scaling factor. The results show that the ratio of HERWIG to SHERPA remains relatively stable across key kinematic variables like m_{jj}, the VBF tagger score, and p_{T}^{H}, and is consistent across decay modes. This consistency suggests that a single scale factor can be applied. This is visualized in Figure 6.10, which shows the distributions of the $Z \to ee$ samples and the HERWIG to SHERPA ratio for the $Z \to ee$ and $Z \to \tau\tau$ samples.

A scaling factor of approximately 1.7 is chosen, which reflects the overall trend observed in the phase space. This value, being slightly conservative, compensates for the larger electroweak fraction in the high m_{jj} region.

6.6.2 Misidentified background

The background due to misidentified objects, where jets are incorrectly identified as electrons, muons, or $\tau_{\text{had-vis}}$, is an important component of the analysis. This background is typically referred to as the misidentified background and contributes

approximately 5% to 25% of the total background, with the largest contributions seen in the less-boosted event categories (where the objects have a lower p_{T} and are more likely to be misidentified). The estimation of this background is data-driven and employs different techniques in the three analysis channels.

6.6.2.1 Misidentified background in the $\tau_e\tau_\mu$ channel

In the $\tau_e\tau_\mu$ channel, the estimation of misidentified leptons is performed using the matrix method. The method needs to define control regions enriched in misidentified leptons through the application of a loose selection. This is achieved by removing the lepton isolation criteria from the nominal signal region selection and loosening the identification requirements, particularly for electrons. This results in an event selection enriched with misidentified leptons, which predominantly arise from multijet events, which usually exhibit more activity in the jet cone compared to real leptons.

Each event contains two leptons, and the combination of the loose and tight selections for each lepton defines four categories, each with event yields N_{LL}, N_{LT}, N_{TL}, and N_{TT}. The last one, which corresponds to the events where both leptons pass the tight selection, is the signal region. The matrix method decomposes the observed yields in these regions as linear combinations of events with two real leptons (N_{rr}), one real and one misidentified lepton (N_{rf}, N_{fr}), and two misidentified leptons (N_{ff}).

The observed yields in the signal and control regions are used to solve a system of equations involving the efficiencies for real (ϵ_r) and misidentified leptons (ϵ_f) to pass the tight selection if they also pass the loose selection. The efficiencies are parameterized as a function of the p_{T} and η of the leptons, with ϵ_r estimated from MC simulations of Z boson decays and ϵ_f measured in data from same-charge dilepton events. These efficiencies are used to define also their counterparts, $\tilde{\epsilon}_r = 1 - \epsilon_r$ and $\tilde{\epsilon}_f = 1 - \epsilon_f$, which are the efficiencies for passing the loose selection while failing the tight selection.

Contributions from processes with real leptons are subtracted using MC predictions, accounting for approximately 35% of the selected events. Taking all the components into account, the system of equations that describes the observed yields in the signal and control regions is given by:

$$\begin{pmatrix} N_{\mathrm{TT}} \\ N_{\mathrm{LT}} \\ N_{\mathrm{TL}} \\ N_{\mathrm{LL}} \end{pmatrix} = M \begin{pmatrix} N_{rr} \\ N_{rf} \\ N_{fr} \\ N_{ff} \end{pmatrix} = \begin{pmatrix} \epsilon_r\epsilon_r & \epsilon_r\epsilon_f & \epsilon_f\epsilon_r & \epsilon_f\epsilon_f \\ \tilde{\epsilon}_r\epsilon_r & \tilde{\epsilon}_r\epsilon_f & \tilde{\epsilon}_f\epsilon_r & \tilde{\epsilon}_f\epsilon_f \\ \epsilon_r\tilde{\epsilon}_r & \epsilon_r\tilde{\epsilon}_f & \epsilon_f\tilde{\epsilon}_r & \epsilon_f\tilde{\epsilon}_f \\ \tilde{\epsilon}_r\tilde{\epsilon}_r & \tilde{\epsilon}_r\tilde{\epsilon}_f & \tilde{\epsilon}_f\tilde{\epsilon}_r & \tilde{\epsilon}_f\tilde{\epsilon}_f \end{pmatrix} \begin{pmatrix} N_{rr} \\ N_{rf} \\ N_{fr} \\ N_{ff} \end{pmatrix}. \tag{6.7}$$

The inversion of the matrix M allows for the determination of the number of events with fake leptons in the signal region, $N^{\mathrm{SR},e\mu}_{\mathrm{misidentified}} = N_{rf} + N_{fr} + N_{ff}$.

The uncertainties in estimating misidentified contributions in the $\ell\ell$ channel arise from several sources. One source is the statistical uncertainty in real and misidentified efficiencies, with separate uncertainties calculated for real electrons, real muons, misidentified electrons, and misidentified muons.

Another uncertainty comes from the subtraction of prompt MC events in the control regions used to isolate misidentified leptons. In this case, a 30% bin-by-bin uncertainty is assigned to the prompt yield subtraction. For the misidentified electron estimation, an additional charge-flip correction is required, introducing another source of uncertainty. This is handled by applying scale factor variations to the charge-flip events, recalculating the misidentified electron efficiencies, and taking the difference as a symmetrical uncertainty.

Further uncertainties are introduced by variations in the real efficiencies, which are evaluated by comparing efficiencies derived from different MC sources, such as $t\bar{t}$ or Z + jets processes. The largest bin-by-bin deviation is used as the uncertainty. Additionally, the flavor composition differences between the misidentified lepton CRs and the signal regions are addressed by comparing misidentified efficiencies in regions with and without a b-jet requirement, with the differences taken as a further uncertainty. Finally, uncertainties related to jet multiplicity are considered by rederiving fake efficiencies for different jet bins (1, 2, 3, and $\geq$4 jets), with the largest observed variation applied as the uncertainty.

6.6.2.2 Misidentified background in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel

For the $\tau_\ell\tau_{\text{had}}$ channel, where jets are misidentified as $\tau_{\text{had-vis}}$, the fake-factor method is employed. Events with a $\tau_{\text{had-vis}}$ candidate that pass a very loose identification criteria but fail the identification medium working point are selected in a control region. The fake factor is defined as the ratio of events where the misidentified $\tau_{\text{had-vis}}$ passes the medium working point to those that fail. These factors are parameterized as a function of p_{T} and the number of associated tracks. To derive the fake factors, two control regions are defined: one enriched in multijet events and the other enriched in W+jets processes. The relative weighting of the fake factors from these regions depends on the fraction of quark- versus gluon-initiated jets.

Jets misidentified as hadronically decaying τ leptons constitute a significant background in this channel. While electrons misidentified as $\tau_{\text{had-vis}}$ are negligible due to the use of the eBDT and are modeled directly with MC samples, jets originating from multijet and W+jets events remain important sources of background. The fake-factor method is applied to estimate their contribution to the signal region. This method uses an anti-tau region, which is defined by inverting the identification criteria of the $\tau_{\text{had-vis}}$ candidate, with a minimal RNN identification score of 0.01 to suppress gluon-initiated jets, which dominate without such a condition. In this anti-tau region, the number of data events ($N^{\text{Data}}_{\text{anti-}}\tau$) is used to estimate the yield in the signal region, weighted by the fake factor, $\mathcal{F}$. Contributions from non-jet

fake sources are estimated using MC simulations and subtracted, yielding the total fake yield:

$$N^{\text{misidentified}}_{\text{SR, }\tau_{\text{lep}}\tau_{\text{had}}} = \mathcal{F} \times \left(N^{\text{anti-}\tau}_{\text{Data}} - N^{\text{anti-}\tau}_{\text{MC, non-jet}} \right). \tag{6.8}$$

To determine the fake factor $\mathcal{F}$, two control regions are defined. The first is enriched in multijet events by inverting the lepton isolation criteria, dominated by gluon-initiated jets. The second is enriched in W+jets events, characterized by quark-initiated jets, by inverting the transverse mass requirement ($m_{\text{T}} > 70\,\text{GeV}$). Both regions are subdivided into τ and anti-τ regions, and the fake factor in each region is derived as the ratio of events passing the τ-lepton selection to those in the anti-tau region. Any contribution in these control regions not originating from jets misidentified as $\tau_{\text{had-vis}}$ is estimated with MC and subtracted before calculating the fake factor:

$$\mathcal{F}_i = \frac{N^{\tau,i}_{\text{Data}} - N^{\tau,i}_{\text{MC, non-jet}}}{N^{\text{anti-}\tau,i}_{\text{Data}} - N^{\text{anti-}\tau,i}_{\text{MC, non-jet}}}. \tag{6.9}$$

The fake factors from the W+jets ($\mathcal{F}_W$) and multijet ($\mathcal{F}_{\text{QCD}}$) regions are combined in a linear combination, where the weights account for the expected R_{QCD} ratio, representing the fraction of multijet events in the misidentified background. This ratio varies across categories and for 1-prong and 3-prong $\tau_{\text{had-vis}}$ decays, typically ranging between 0.02 and 0.11. The fake factors $\mathcal{F}_W$ and $\mathcal{F}_{\text{QCD}}$ differ by approximately 30%, and the final fake factor used in the analysis is expressed as:

$$\mathcal{F} = (1 - R_{\text{QCD}})\mathcal{F}_W + R_{\text{QCD}}\mathcal{F}_{\text{QCD}}. \tag{6.10}$$

The R_{QCD} factor is determined by scaling the events in the multijet control region by the ratio between the number of events where the light lepton passes the isolation requirements and the number where it does not. This ratio, called the isolation factor, is measured in a dedicated region enriched in multijet events, where the light lepton and the $\tau_{\text{had-vis}}$ have the same electric charge. The R_{QCD} factor is calculated separately for 1-prong and 3-prong $\tau_{\text{had-vis}}$ decays, and it is binned in the p_{T} of the $\tau_{\text{had-vis}}$ and the azimuthal angle between the $\tau_{\text{had-vis}}$ candidate and the $E^{\text{miss}}_{\text{T}}$, $\Delta\phi(\tau_{\text{had-vis}}, E^{\text{miss}}_{\text{T}})$.

The fake factors are calculated for each signal region category, for both 1-prong and 3-prong $\tau_{\text{had-vis}}$, and as a function of p_{T}.

Residual contributions from events with real $\tau_{\text{had-vis}}$ candidates in the control region are subtracted using MC simulations, accounting for approximately 18% of the selected events.

The uncertainties in the background estimation using the Fake-Factor method arise from various sources. One major source of uncertainty comes from the statistical limitations in the control regions where the individual fake factors, $\mathcal{F}_{QCD}$ and $\mathcal{F}_W$, are derived.

Additionally, the uncertainty on the R_{QCD} factor accounts for the statistical uncertainties on its determination, as well as the systematic uncertainties related to the possible differences in the fake factors derived from same-sign and opposite-sign event selections.

There is also an uncertainty related to the MC subtraction used in the calculations of F_{QCD}, F_W, and R_{QCD}. This uncertainty is evaluated by varying the MC contribution by 10%, a conservative estimate based on the uncertainties in the cross-sections of the processes subtracted in the MC. Other uncertainties include non-closure systematics and the statistical uncertainty in the anti-$\tau_{\text{had-vis}}$ region, where the fake factors are applied.

6.6.2.3 Misidentified background in the $\tau_{\text{had}}\tau_{\text{had}}$ channel

For the $\tau_{\text{had}}\tau_{\text{had}}$ channel, the process is more complex, as both hadronic τ candidates may be misidentified. The data-driven method for this channel is an extension of the fake factor approach used in the $\tau_\ell\tau_{\text{had}}$ channel but adapted to account for the possibility of two misidentified τ leptons. The primary source of such backgrounds is from QCD multijet events, where one or both jets are wrongly reconstructed as $\tau_{\text{had-vis}}$ candidates. The background contribution to the signal region is derived from events where one of the $\tau_{\text{had-vis}}$ candidates fails the medium identification criteria but passes the loose criteria, ensuring that the events are still included in the analysis. The presence of real hadronic τ leptons in these regions is estimated using simulated events and subtracted from the data.

The fake factors used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel are adapted from control regions defined in the $\tau_\ell\tau_{\text{had}}$ channel, specifically from W+jets enriched samples. Modifications are made to the criteria on $\tau_{\text{had-vis}}$ candidates to match the selection used in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, with adjustments to the transverse momentum and pseudorapidity conditions to ensure sufficient statistics. The fake factors are parametrized as functions of p_{T} and η, reflecting the dependence of the misidentification rate on these variables.

To evaluate systematic uncertainties associated with the background from misidentified τ_{had} candidates, two alternative sets of fake factors are calculated. These are derived from control regions with two τ_{had} leptons: in the first set, the $|\Delta\eta_{\tau\tau}|$ requirement is reversed, and in the second, the opposite-charge condition between the two τ_{had} leptons is inverted. The differences between these alternative fake factors and the nominal estimation provide a measure of the uncertainties related to the composition of the misidentified τ background. In addition, statistical uncertainties on the fake factors and those arising from the chosen parametrization are also considered.

Finally, the accuracy of the misidentified background estimation is validated in dedicated control regions for each channel. For the $\tau_{\text{had}}\tau_{\text{had}}$ channel, a region is defined by inverting the requirement on the $\Delta\eta$ between the $\tau_{\text{had-vis}}$ candidates. In the $\tau_\ell\tau_{\text{had}}$ channel, a validation region is constructed using events with a light

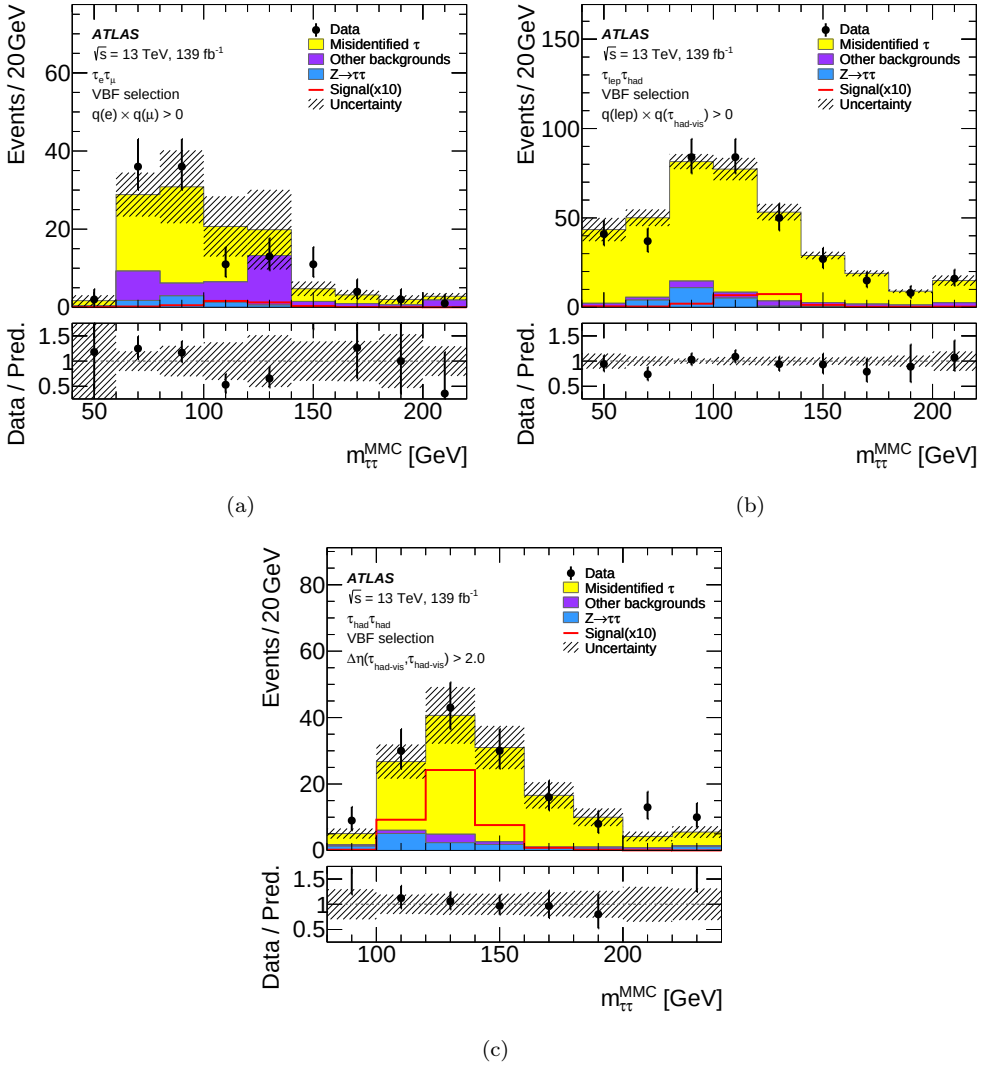


Figure 6.11: Validation of the data-driven estimation for processes where jets are misidentified as τ_{had}-vis in the VBF categories is shown in three scenarios: (a) events in the $\tau_{\text{had}}\tau_{\text{had}}$ final state with a separation $\Delta\eta(\tau_{\text{had}}\tau_{\text{had}}) > 2.0$, (b) events in the $\tau_{\text{lep}}\tau_{\text{had}}$ final state where the light lepton and the τ_{had} have the same charge, and (c) events with same-charge leptons in the $\tau_e\tau_\mu$ final state. In all cases, the hashed band illustrates the statistical uncertainties due to the limited size of the simulated samples, as well as the systematic uncertainties from the data-driven estimation process.

lepton and a $\tau_{\text{had-vis}}$ with the same charge, while for the $\tau_e\tau_\mu$ channel, same-charge light lepton pairs are selected. The validation regions demonstrate good agreement between data and the predicted background, as shown in Figure 6.11.

6.6.2.4 Combination of the misidentified background templates in the VBF categories

The fine phase space binning of the VBF categories in the analysis can introduce unphysical statistical fluctuations in the misidentified background templates, leading to instabilities in the fit that extracts the signal strengths. This issue particularly

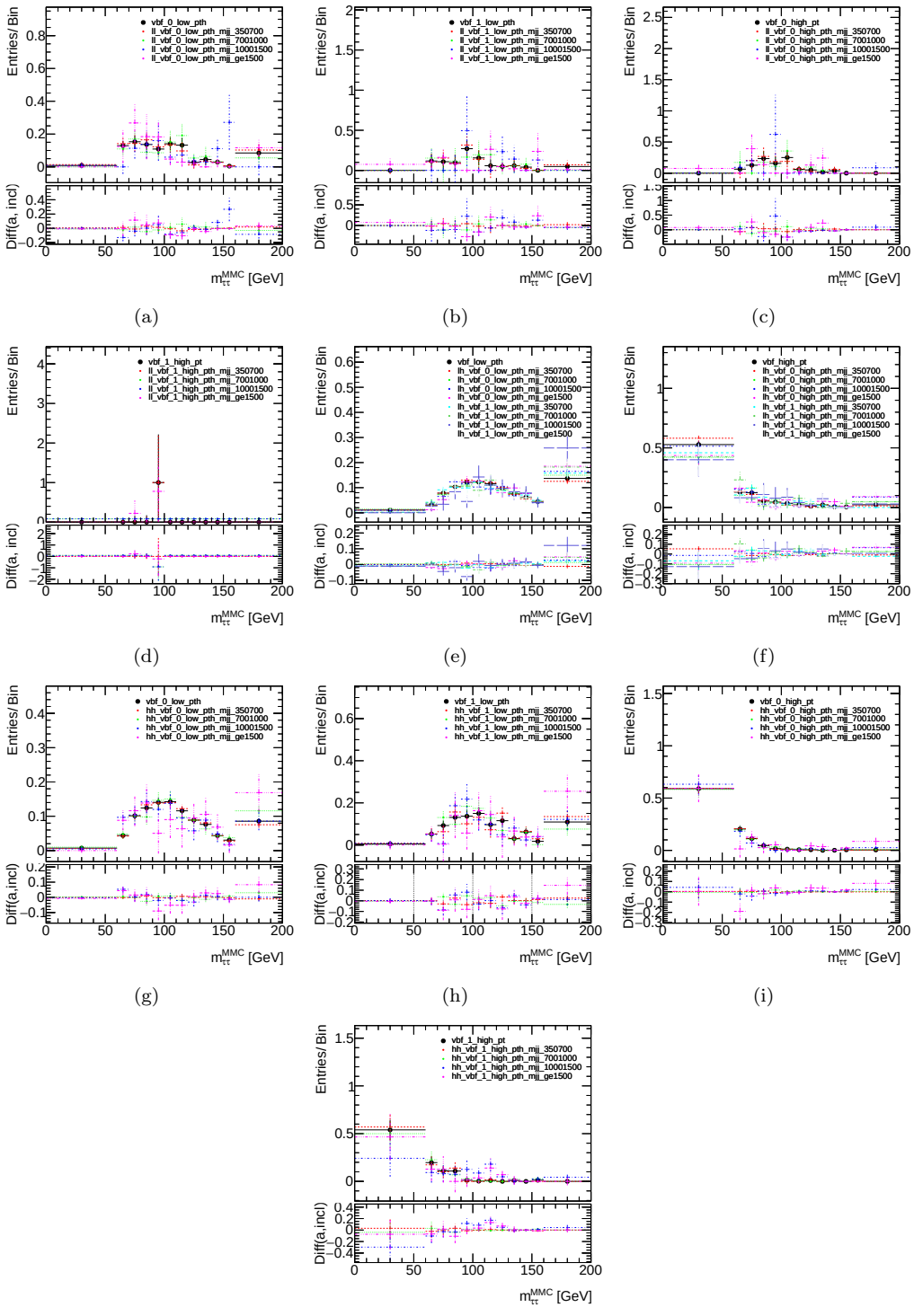


Figure 6.12: Comparison of misidentified background templates before and after the combination across different categories in the VBF phase space. The templates are shown for the three analysis channels.

affects the VBF categories with low event yields, at high p_{T}^{H} and m_{jj}, where the misidentified background is small. However, statistical fluctuations in the misidentified background templates can still have a significant impact on the analysis.

To mitigate this issue, the misidentified background templates are combined across different bins in the VBF categories, leading to inclusive templates that are normalized to the total misidentified background yield in each of the VBF categories. Before the combination, the misidentified background templates were compared across the different categories to ensure that their shapes are qualitatively consistent. In the $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ and the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channels, the misidentified background templates are combined across the different m_{jj} bins, resulting in four templates corresponding to the low and high p_{T}^{H} regions in the two VBF categories defined by the VBF tagger, VBF_0 and VBF_1. In the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel, the misidentified background templates are combined across the different m_{jj} bins and the two VBF_0 and VBF_1 categories, resulting in two templates corresponding to the low and high p_{T}^{H} regions. The original misidentified background templates and the combined templates are shown in Figure 6.12.

6.6.3 Other backgrounds

The top-quark-related backgrounds are negligible in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel, with the exception of the $t\bar{t}H$ categories, and their contribution is estimated using MC simulations. In the $\tau_e\tau_\mu$ and $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels, the top-quark-related backgrounds represent a significant contribution to the total background, and the estimation is performed correcting the MC normalization with data. A set of control regions is defined starting from the different preselections used in the analysis and inverting the veto on b-tagged jets (except for the $t\bar{t}H$ control region, defined through the inversion of cuts on the multiclassifier scores, as described in section 6.5.1.4). These control regions, enriched in top-quark events, are used to derive normalization factors through a simultaneous fit to the data that includes the signal regions. These normalization factors scale the top-quark background prediction. The purity of the top-quark-related processes in these control regions is higher than 70%. Separate control regions are defined for the three analysis channels and for the different preselections (boost, VBF, VH and $t\bar{t}H$) targeting the different Higgs production modes, resulting in 12 top-quark control regions in total.

The rest of the backgrounds, such as diboson or $H \to WW^*$, is estimated using MC simulations, without data-driven corrections.

6.7 Systematic uncertainties

Systematic uncertainties from various sources influence the $H \to \tau\tau$ cross-section measurements. They affect both the signal and background estimation and can be classified based on their origin: experimental uncertainties, which arise from detector

performance and reconstruction; theoretical uncertainties related to background modeling; and theoretical uncertainties affecting the signal prediction. Each type of uncertainty requires specific treatments to propagate its effect into the analysis. Typically, systematic uncertainties are quantified by evaluating the $\pm 1\sigma$ variations and then propagating these to the final distributions used in the statistical fit.

These systematic uncertainties can also be classified into two categories based on their effects on either event weights or kinematic properties of the objects in the events.

The first type of uncertainty, those affecting event weights, modifies how events are weighted within a given category. To account for this, alternative sets of weights are applied for each event, and the resulting distributions are compared to the nominal case. These shifts can change not only the normalization of distributions but also the shape and relative contribution of different processes. Since the weighting variations are often dependent on specific kinematic variables, they introduce both normalization and shape uncertainties into the final result.

The second category includes uncertainties that affect the reconstructed kinematic properties of events, such as the 4-momenta of particles. These uncertainties require a full reprocessing of the event, recalculating derived quantities such as $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{MMC} under the new conditions. Variations in kinematic quantities can cause event migrations between different reconstructed regions, changing the signal acceptance of the analysis, or modifying the kinematic distributions within a region.

The incorporation of systematic uncertainties in the likelihood fit as nuisance parameters (NPs) is discussed in Section 6.8 and is further detailed in Appendix A. This section discusses the main sources of systematic uncertainties in the $H \to \tau\tau$ analysis.

6.7.1 Theoretical uncertainties on the signal prediction

The theoretical uncertainties on the signal prediction in this $H \to \tau\tau$ analysis follow the guidelines provided by the LHC Higgs Cross-Section Working Group [24, 65]. These recommendations ensure consistency between different Higgs boson analyses and provide a common framework to allow for the combination of results. The theoretical uncertainties cover several sources as the parton distribution functions, the strong coupling constant, α_S, the QCD scale uncertainties, the parton shower modeling and the uncertainties related to the matrix element generator.

The impact of uncertainties associated with the parton PDFs is evaluated using the eigenvector variations of the PDF4LHC_NLO_30 Hessian set [178]. These variations are incorporated by modifying the MC weights through the reweighting scheme implemented within the PowhegBox generator. The resulting changes in the predictions for each eigenvector variation are treated as an independent source of uncertainty in the fit model. The α_s uncertainties are derived by varying the value of α_s around the central value. The PDF and α_s uncertainties are correlated across

different Higgs boson production modes, ensuring that the effects of variations in the parton densities and strong coupling are consistently propagated across all production processes.

For gluon-gluon fusion, QCD scale uncertainties account for variations in the renormalization and factorization scales, impacting the inclusive cross-section and uncertainties related to resummation effects [231]. At low p_{T}^{H} ($p_{\mathrm{T}}^{H} < 200$ GeV), they include migration effects between different jet multiplicity bins (0, 1, or 2 jets) and uncertainties in p_{T}^{H} for events with no jets, as well as events with one or more jets, covering migrations between various p_{T}^{H} bins (0–60 GeV, 60–120 GeV, 120–200 GeV). For events with two jets, the uncertainties also consider migrations between m_{jj} bins (0–350 GeV, 350–700 GeV, 700–1000 GeV, 1000–1500 GeV, 1500–∞ GeV) and around $p_{\mathrm{T}}(\mathrm{Hjj}) = 25$ GeV. At high p_{T}^{H}, the QCD scale uncertainties include normalization and migration effects in different p_{T}^{H} bins (200–300 GeV, 300–450 GeV, 450–650 GeV, 650–∞ GeV) and migration effects between boundaries defined by the ratio of the Higgs boson transverse momentum and that of an additional jet, $p_{\mathrm{T}}^{\mathrm{Hj}}$, and p_{T}^{H} as $p_{\mathrm{T}}^{\mathrm{Hj}}/p_{\mathrm{T}}^{H}$. The calculation of these uncertainties is based on variations of the renormalization and factorization scales and in the Stewart-Tackmann approach, and is described in detail in Ref. [232].

The BDTs used in the VBF and VH categories to define signal-pure regions can mimic third-jet vetoes. The variation of normalization and factorization scales in categories with a third-jet veto can underestimate the QCD scale uncertainties in these categories. To cover this effect, two additional systematic uncertainties are introduced, following the Stewart-Tackmann approach.

For electroweak qqH production, ten independent sources of QCD scale uncertainty are considered. These include uncertainties related to the inclusive cross-section, as well as bin migration uncertainties across different p_{T}^{H} (0–200 GeV, 200–∞ GeV) and m_{jj} bins (0–60 GeV, 60–120 GeV, 120–350 GeV, 350–700 GeV, 700–1000 GeV, 1000–1500 GeV, 1500–∞ GeV). They also consider migration uncertainties around $p_{\mathrm{T}}^{\mathrm{Hjj}} = 25$ GeV and uncertainties related to the requirement of having at least two jets in the final state.

The QCD scale uncertainties for $t\bar{t}H$ production are divided into six independent components. These include one uncertainty related to the inclusive cross-section and five uncertainties related to p_{T}^{H} bin migrations, based on the STXS 1.2 definitions. The bin migrations cover the following p_{T}^{H} ranges: 0–60 GeV, 60–120 GeV, 120–200 GeV, 200–300 GeV, 300–450 GeV, 450–∞ GeV.

Parton shower and matrix element uncertainties are also included for the ggF, VBF, WH, ZH, and $t\bar{t}H$ production modes. These uncertainties are evaluated by generating alternative Monte Carlo samples where the parton shower model is varied while keeping the matrix-element generator fixed. The specifications of the alternative samples are described in Section 4.3.1. The nominal samples use Pythia for the parton shower, while the alternative samples use Herwig7. This allows for an estimation of the uncertainties associated with the modeling of the parton shower in the signal production.

Shape uncertainties are introduced for ggF events in the VBF and VH categories, regions with higher jet multiplicities. These uncertainties account for differences in the matrix element calculations for ggF production when additional jets are present. The nominal ggF predictions are compared with alternative Monte Carlo samples generated using the FxFx jet merging scheme, which includes higher-order corrections in the jet multiplicity. This allows for an estimation of the uncertainties related to the modeling of additional jets.

Matrix element uncertainties are also considered for the remaining production modes, estimated by comparing predictions from different matrix-element generators, such as Powheg and MG5_aMC@NLO, while keeping the parton-shower model fixed. These uncertainties cover the differences in the matrix-element calculations, which can impact the overall signal prediction.

6.7.2 Theoretical uncertainties on the background prediction

Several sources of uncertainty are considered for the $Z \to \tau\tau$ background, including uncertainties from the parton distribution functions, factorization and renormalization scales, the CKKW matching scheme, QCD scale factors (QSF), as well as uncertainties from the underlying event and parton shower models.

The uncertainties from the PDFs are estimated using 100 different weight variations from the Sherpa generator. These variations are built by varying the most significant 100 eigenvectors of the Hessian of the NNPDF3.0nnlo PDF set. These are combined to yield a final PDF uncertainty. The uncertainties due to the missing higher-order corrections are estimated by varying the factorization and renormalization scales by different factors, and the envelope of the variations is taken as the uncertainty. To reflect the differences across various analysis categories, both the normalization and shape components of the PDF uncertainty are decorrelated across final states and signal categories. However, the correlations between the nuisance parameters in the $Z \to \tau\tau$ control region and the signal categories are maintained. These uncertainties are strongly correlated with the normalization factors extracted in the fit that scale the $Z \to \tau\tau$ background in the signal regions and the associated control regions obtained through the embedding method. The decorrelation scheme ensures that scaling factors in specific phase space regions do not affect the normalization of the $Z \to \tau\tau$ background in other regions. The decorrelation between normalization and shape components guarantees that the possible constraints on the normalization components due to the embedding method do not affect the shape effects of the uncertainties.

The $Z \to \tau\tau$ theoretical uncertainties are also decorrelated in the $t\bar{t}H$ categories, where the $Z \to \tau\tau$ background normalization is not corrected through a normalization factor extracted from a region obtained with the embedding method.

The same set of nuisance parameters used for the $Z \to \tau\tau$ background is also applied to $Z \to \ell\ell$ events.

An additional source of uncertainty is applied to account for modeling problems in the electroweak component of the $Z \to \tau\tau$ background in the MC simulation. As described in section 6.6.1.2, the electroweak component is estimated using MC simulation, and a scale factor is derived by comparing the SHERPA prediction with alternative samples generated using different matrix-element generators. An uncertainty is assigned to this scale factor to account for possible deviations. The downward variation is defined as the full difference between the template without the scale factor and the template multiplied by 1.7 (the scale factor determined from the study described above). The upward variation is obtained by symmetrizing the downward variation.

An uncertainty is associated with this scale factor to cover potential deviations. The downward variation was determined by the full difference between the template without the scale factor and the template multiplied by 1.7. The upward variation was obtained by symmetrizing the downward variation.

The uncertainties on the top-quark related background include multiple sources: variations in the PDFs (based on the variations of the 30 most significant eigenvectors of the PDF set), matrix-element generator comparisons (using different generators but the same parton shower), parton shower model comparisons (with the same matrix-element generator), as well as uncertainties from initial and final state radiation. The latter two are incorporated using event weights obtained from reweighting the nominal predictions. The inclusion of these uncertainties is especially important in the $t\bar{t}H$ categories, where the top-quark related backgrounds are a significant contribution to the total background.

6.7.3 Experimental uncertainties

In the $H \to \tau\tau$ analysis, several experimental uncertainties arise from different detector components and data-driven techniques. Among the primary uncertainties, the jet energy scale, the τ-lepton identification efficiency, the τ-lepton energy scale, and misidentification rates for τ-leptons have substantial impacts on the final results. Additionally, uncertainties related to the methods used for background estimation need to be accounted for.

Jet-related uncertainties are an essential part of the analysis. The jet energy scale uncertainty arises from multiple sources, including the calibration of jets *in situ*, pile-up effects, high-momentum extrapolations, and the variation in response between jets initiated by quarks and gluons. The magnitude of this uncertainty is approximately 1% in the central region of the detector, covering a broad transverse momentum range of 250 GeV $< p_{\mathrm{T}} <$ 2000 GeV.

The jet energy scale uncertainties are described using a set of 30 uncorrelated nuisance parameters, each representing a different source of uncertainty. Five nuisance parameters are introduced to account for uncertainties in the jet energy scale determination related to jet flavor. These include the flavor composition, which reflects the fraction of jets originating from quarks and gluons; the flavor

response, capturing uncertainties in the response of gluon-initiated jets; the b-jet response, representing uncertainties specific to b-jets response; the punch-through correction, addressing uncertainties in a dedicated calibration correction; and the single-particle response, a high-p_{T} uncertainty derived from test-beam and single-particle measurements.

Other four nuisance parameters cover uncertainties in the jet energy scale related to the pile-up conditions, including the pile-up offset, which accounts for the uncertainty in the pile-up modeling in the MC simulation; the N_{PV} offset, which addresses the uncertainty in the number of primary vertices in the MC simulation; the ρ term, describing the uncertainty in the p_{T} density modeling in the simulation and the pile-up p_{T} term, which captures the uncertainty in residual p_{T} dependencies.

Six nuisance parameters are used to capture uncertainties in the jet energy scale associated with η-intercalibration. These account for statistical variations and systematic effects arising from changes in the MC generator configurations, pile-up simulation conditions, and event topologies. Additionally, four non-closure uncertainties are included: three related to jets in different p_{T} and η regions and one addressing uncertainties in the Tile calorimeter calibration for data collected in 2018.

The remaining 15 nuisance parameters are defined through an eigenvector decomposition of jet energy scale uncertainties obtained from in-situ measurements. The most significant eigenvectors determine the primary variations, while the remaining contributions are consolidated into a residual uncertainty. This analysis employs the *category-reduction* scheme, grouping uncertainties by their source, including detector-related, statistical, modeling-induced, and mixed uncertainties. Two nuisance parameters cover the uncertainties in the JVT and the fJVT efficiencies. The detailed description of the jet energy scale uncertainties is provided in Ref. [212].

Uncertainties in the jet energy resolution, indirectly affects the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement, which in turn impacts the invariant mass, $m_{\tau\tau}$. The degraded resolution leads to a broader $Z \to \tau\tau$ resonance peak in the $m_{\tau\tau}$ distribution, influencing the region most sensitive to the Higgs signal. This uncertainty is momentum-dependent, ranging from 24% at low momentum (20 GeV) to 6% at higher momentum (300 GeV). A set of 12 NPs, resulting from the eigenvector decomposition of the jet energy resolution uncertainties, is used to describe the variations in the resolution. These uncertainties are also described in Ref. [212].

The uncertainties in the $E_{\mathrm{T}}^{\mathrm{miss}}$ are primarily associated with the soft term, as separate uncertainties for hard objects contributing to the $E_{\mathrm{T}}^{\mathrm{miss}}$ are treated independently. These soft-term uncertainties are evaluated by varying its energy scale and resolution, with the latter involving independent variations parallel and perpendicular to the p_{T} direction of the hard objects.

Uncertainties related to τ-leptons also play a significant role, affecting the identification, the energy scale determination, the trigger efficiency and the efficiency

of the eBDT used to reject electrons misidentified as τ-leptons.

The uncertainties in the τ-lepton energy scale are modeled using five nuisance parameters, each corresponding to a distinct source of uncertainty. Two parameters are based on results from dedicated tag-and-probe studies, one originates from measurements of the τ-lepton energy-to-momentum ratio, and the remaining two are derived from comparisons between reconstructed and truth-level τ-lepton energies in MC simulations, excluding detector effects.

A set of nuisance parameters are used to describe the uncertainties in the τ-lepton trigger efficiencies. They are decorrelated across different data-taking periods (2015–2017 and 2018) and between statistical and systematic components. The uncertainties are evaluated using dedicated tag-and-probe studies to compare the efficiencies in data and MC simulations. An additional NP is included to account for the uncertainties in the identification efficiencies for the p_{T} range above 100 GeV, not covered by the tag-and-probe studies.

Additional NPs cover uncertainties in the efficiency of the eBDT used to reject electrons misidentified as τ-leptons and on the τ-lepton reconstruction efficiency. The latter is evaluated comparing the efficiencies between different simulation samples.

Uncertainties in the τ-lepton identification efficiency are also considered. They are derived by comparing the efficiencies in data and MC simulations, and they are decorrelated according to the τ-lepton decay mode and the p_{T} regime.

The τ-lepton identification efficiency uncertainty ranges from 2% to 6%, while the uncertainty in τ-lepton energy scale lies between 1% and 4%, depending on the τ-lepton momentum and track multiplicity. Moreover, the efficiency of the boosted decision tree used to reject electrons misidentified as τ-leptons has an associated uncertainty of 1% to 2%, which varies depending on the τ-lepton decay mode and kinematics. In terms of trigger performance, the uncertainty in τ-lepton trigger efficiency is relatively small, between 1% and 1.5%, but still a non-negligible contributor to the total systematic uncertainty. These uncertainties are evaluated using MC simulation and are a significant component of the overall τ-lepton reconstruction uncertainty.

Uncertainties in muon reconstruction and identification include contributions from the momentum scale and resolution, as well as trigger, reconstruction, and isolation scale factors. Specific uncertainties involve the track resolution in the ID and MS, momentum scale, and two charge-dependent momentum scale corrections. Additionally, statistical and systematic uncertainties affect the trigger, reconstruction, and isolation scale factors. These uncertainties have minor impact on the $H \to \tau\tau$ analysis.

Uncertainties related to electrons include contributions from the energy scale, energy resolution, and scale factors for trigger, reconstruction, identification, and isolation. The uncertainties due to the energy scale and resolution of electrons are incorporated into the analysis through two nuisance parameters. These nuisance parameters combine the uncertainties across different η regions and energy ranges.

These uncertainties arise from factors such as limited knowledge of detector material, shower shape modeling, electronic noise, and energy loss before calorimeters.

Uncertainties in the scale factors for trigger, reconstruction, identification and isolation are also considered. Except for identification SFs, these uncertainties are simplified into single nuisance parameters by adding all sources of systematic and statistical uncertainties in quadrature, correlating them across p_{T} and η bins, as in the case of the electron energy scale and resolution uncertainties.

Identification SFs required a more detailed treatment due to strong constraints observed with simpler schemes. A set of 12 nuisance parameters account for statistical uncertainties in different p_{T} and η bins, while 16 additional parameters represent systematic uncertainties. The latter are derived by decomposing the covariance matrix of systematic uncertainties into uncorrelated sources. The statistical uncertainties for the identification SFs are divided into barrel and endcap regions in η and further segmented into six p_{T} bins, covering ranges from 15 GeV to ≥ 80 GeV. The full description of the electron uncertainties is provided in Ref. [204].

Uncertainties on the flavor tagging efficiencies are also considered in the analysis, and have a small impact on the final results. These uncertainties are larger for the top-quark related backgrounds, but are still a minor contribution to the total systematic uncertainty. Uncertainties in b-jet tagging efficiencies are accounted for through variations of scale factors that are derived from an eigenvector decomposition of the covariance matrix of systematic and statistical uncertainties. These uncertainties include contributions for correctly tagging b-jets, misidentifying c-jets as b-jets, and misidentifying light-flavor jets as b-jets. Additionally, two nuisance parameters address extrapolations: one for high-p_{T} jets and another for extending SFs from c-jets to hadronic τ-lepton jets. These uncertainties are described in Ref. [215].

The uncertainty in the integrated luminosity measurement is an additional systematic affecting the signal and the background processes. The absolute luminosity calibration of the LUCID detector is determined using dedicated van der Meer scans, where the beams are scanned across each other. The uncertainty in the absolute luminosity determination is $\pm 0.83\%$, and its driven by the uncertainty on the extrapolation of the calibration from the low-μ conditions of the van der Meer scans to the conditions of the data-taking periods. It is also affected by magnetic non-linearities effects affecting the beam positions and effects of beam-beam interactions [172]. The uncertainty impacts the normalization of signal and background Monte Carlo samples; however, it does not apply to background processes whose normalizations are constrained by measurements in dedicated control regions.

Background estimation methods, particularly those relying on data-driven techniques, introduce specific uncertainties that have been already described in previous sections. In the $\tau_e\tau_\mu$ channel, uncertainties arise from the statistical limitations of the efficiency measurements, variations due to the number of b-tagged jets, and processes involving real τ-leptons. For the $\tau_\ell\tau_{\mathrm{had}}$ channel, the fake factor

method brings uncertainties from the statistical uncertainties in the fake factors, the flavor composition of the jets, and uncertainties associated with processes where a real τ-lepton is present. The $\tau_{\text{had}}\tau_{\text{had}}$ channel has similar uncertainties, including those related to the statistical precision of the fake factors and the background composition. These uncertainties are described in detail in Section 6.6.2.

Moreover, a conservative normalization systematic uncertainty of 1% is introduced in the $Z \to \tau\tau$ control regions to account for potential mismodeling in the kinematic variables used to define the control regions due to the usage of the embedding method. This mismodeling might result in migration of events between categories. The assigned uncertainty is treated as uncorrelated across the different analysis regions.

6.8 Statistical analysis

The analyses presented in this chapter are designed to measure several parameters of interest (POI) related to the Higgs boson production. In the case of the STXS measurement, the analysis aims to measure an inclusive cross-section $\sigma(pp \to H \to \tau\tau)$, cross-sections for the different production modes ($\sigma^{\text{ggF}}_{H\to\tau\tau}$, $\sigma^{\text{VBF}}_{H\to\tau\tau}$, $\sigma^{VH}_{H\to\tau\tau}$, $\sigma^{t\bar{t}H}_{H\to\tau\tau}$), and the cross-sections in the different STXS bins. In the case of the fiducial differential cross-section measurement, the analysis aims to measure the cross-section in the different bins of the observables that are unfolded. In both cases, the analysis is performed using a binned maximum likelihood fit to the invariant mass of the $\tau\tau$ system, $m_{\tau\tau}$, in the different analysis categories. The likelihood function used in the fit is defined as the product of Poisson probabilities across all the bins of the input distributions, including the signal and control regions:

$$\mathcal{L}(\vec{n}|\vec{\mu},\vec{\theta},\vec{\lambda},\vec{\gamma}) = \prod_{r\in\text{regions}}\prod_{i\in\text{bins}} \text{Pois}(n_{r,i}|\vec{\mu},\vec{\theta},\vec{\lambda},\vec{\gamma})\mathcal{L}_\gamma(\gamma_{r,i}|m_{r,i}) \prod_p \text{Gauss}(\theta^0_p|\theta_p,\sigma^0_p). \tag{6.11}$$

In this equation, $\vec{n}$ is the vector of observed events in each bin of the input distributions, while $n_{r,i}$ is the number of observed events in a given bin i of a given region r. The signal yields are scaled by the signal strength modifiers, which are the parameters of interest in the fit, represented by the vector $\vec{\mu}$. The $Z \to \tau\tau$ and top-quark related backgrounds are scaled by normalization factors that are also treated as free parameters in the fit, represented by the vector $\vec{\lambda}$. The systematic uncertainties are included in the likelihood function introducing nuisance parameters that are constrained by Gaussian terms that multiply the Poisson likelihood components. The NPs are represented by the vector $\vec{\theta}$, and are constrained by Gaussian terms with mean θ^0_p and width equal to the systematic uncertainty associated with the NP, σ_p. Finally, the likelihood function depends on the parameters $\vec{\gamma}$ that are used to model the statistical uncertainties in the background sample sizes. These parameters are constrained by Poissonian-like terms, $\mathcal{L}_\gamma(\gamma_{r,i}|m_{r,i})$, to reflect the statistical nature of the uncertainties.

The term $\text{Pois}(n_{r,i}|\vec{\mu},\vec{\theta},\vec{\lambda},\vec{\gamma})$ represents the Poisson probability for the number of observed events in bin i of region r given the expected number of events in that bin, which is a function of the signal strength modifiers, the normalization factors and the nuisance parameters. The expected number of events in bin i of region r is given by $\sum_k \mu_k s_{r,i,k}(\vec{\theta}) + \gamma_{r,i} b_{r,i}(\vec{\theta},\vec{\lambda})$, where the sum runs over the different signal processes and $b_{r,i}(\vec{\theta},\vec{\lambda})$ is the expected background yield in that bin. The $\gamma_{r,i}$ parameter is used to model the statistical uncertainty in the background sample size in each bin, and is considered correlated across the different background processes. The Poisson probability is given by:

$$\text{Pois}(n_i|\vec{\mu},\vec{\theta},\vec{\lambda},\vec{\gamma}) = \frac{\left(\sum_k \mu_k s_{r,i,k}(\vec{\theta}) + \gamma_{r,i} b_{r,i}(\vec{\theta},\vec{\lambda})\right)^{n_i}}{n_i!} e^{-\left(\sum_k \mu_k s_{r,i,k}(\vec{\theta}) + \gamma_{r,i} b_{r,i}(\vec{\theta},\vec{\lambda})\right)}. \tag{6.12}$$

The model is fitted to the data by maximizing the likelihood function in order to extract the free parameters of the model, that include the signal strength modifiers and the normalization factors. A full description of the profile likelihood method can be found in Appendix A. The next sections will describe the likelihood models that are used in the STXS and the fiducial differential cross-section measurements to extract the parameters of interest.

6.8.1 Fit model for the STXS measurement

The likelihood function used in the STXS measurement incorporates several signal and control regions for various Higgs production modes. The event categorization has been described in previous sections, but a brief summary of the final categories is provided here. The categories used in the statistical analysis are shown in Figure 6.13, together with the background normalization factors that are treated as free parameters in the fit. The signal regions considered in the fits are also shown in Figure 6.14, together with the different STXS bins that each signal region targets.

The Boost category includes six signal regions per channel, defined by the jet multiplicity, N_{jets} and the transverse momentum of the Higgs boson, p_{T}^{H}. Each of these SRs has a corresponding control region that targets $Z \to \tau\tau$ plus jets, with kinematic embedding described in section 6.6.1.1. Additional CRs are enriched in top-quark pair production ($t\bar{t}$) for the $\tau_e\tau_\mu$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels. These top-quark CRs are defined inclusively for the Boost category. Altogether, this setup yields a total of 38 categories in the Boost region when summing the signal and control regions across the three channels.

For the VH category, two signal regions are constructed for each channel by dividing the inclusive VH category into high and low BDT score regions. Here, the BDT is trained to enhance separation between VH production and other processes. Similarly, CRs are defined to capture $Z \to \tau\tau$ plus jets for each signal region, and a top-quark CR is defined inclusively for the VH production, leading to 14 categories for this production mode.

In the VBF category, sixteen signal regions per channel are defined, following the selection described in section 6.5.1.2. Control regions analogous to those used in the Boost and VH categories are also constructed to account for the $Z \to \tau\tau$ plus jets. As in the Boost and VH categories, top-quark CRs are defined inclusively for the VBF category. Summed across channels, this results in 98 VBF categories.

For the $t\bar{t}H$ process, six signal regions and two control regions are introduced in the $\tau_{\text{had}}\tau_{\text{had}}$ channel using the output of the multiclassifier described in Section 6.4.3. As described in Section 6.5.1.4, in order to build the SRs three phase space regions are defined based on p_{T}^{H}, and these regions are further split into two categories: a sideband region covering bins of the $m_{\tau\tau}$ distribution with low signal purity, and a window region with high signal purity. The control regions are inclusive for the $t\bar{t}H$ process, and aim to control the background normalization of the $t\bar{t}$ and $Z \to \tau\tau$ processes.

In the signal regions, the binned $m_{\tau\tau}$ distribution is used to extract information about the Higgs boson production cross-section. The MMC method is employed to reconstruct $m_{\tau\tau}$ in the Boost and VH categories, while the collinear mass is used for the VBF and $t\bar{t}H$ categories in cases where the MMC technique fails.

The control regions are used to improve constraints on background processes by employing a single-bin histogram that counts the number of events within the relevant CR. No shape information is used in the CRs to constraint the background processes. The data in low purity bins of the CRs also help to constraint the normalization of the background processes and the systematic uncertainties associated to them. The normalization factors are treated as free parameters in the fit, and one normalization factor is defined for each CR. A scheme showing the fit model used in the analysis, including the signal and control regions and the associated normalization factors is shown in Figure 6.13.

Three distinct measurements are performed within this framework. First, the total cross-section for $pp \to H \to \tau\tau$ is estimated using a single parameter of interest, corresponding to the cross-section times the Higgs boson branching fraction to $\tau\tau$. In this case, the relative contributions from different production modes are fixed according to the SM predictions. This fit model is usually referred to as the 1-POI fit.

Second, cross-sections for individual Higgs production modes, including ggF, VBF, VH, and $t\bar{t}H$, are determined by introducing four POIs. The event yields are parameterized separately for each production mode. In this case, the model is called the 4-POI fit.

Finally, a fit is performed to extract 18 POIs corresponding to cross-sections for different bins within the STXS stage 1.2 framework, receiving the name of the 18-POI fit. These include six bins for ggF, three for $t\bar{t}H$, and nine for VBF plus $qq \to V(\to qq)H$. The bins were selected by combining some originally proposed STXS bins to maximize sensitivity and account for the available signal sample size. Specifically, the $qq \to V(\to qq)H$ process is measured for dijet masses between 60 GeV and 120 GeV, while the VBF contribution is measured for events where m_{jj}

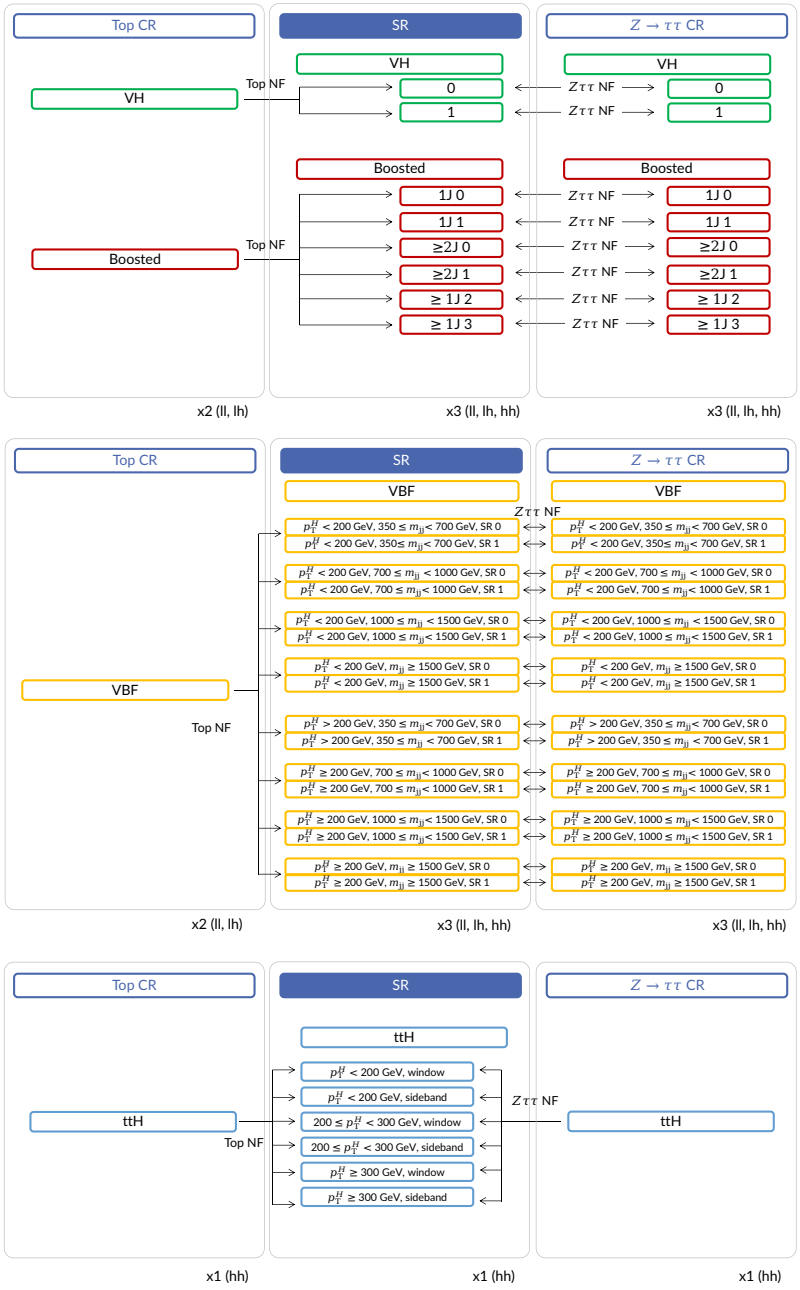


Figure 6.13: Schematic summary of the fit model used in the analysis. All regions which are used directly in the combined fit are indicated. They are grouped by topology (Boosted (red), VBF (orange), VH (green) and ttH (blue)). The $t\bar{t}H$ signal and control regions are only defined for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The $Z \to \tau\tau$ control regions are defined in each channel. The top-quark control regions in Boosted, VBF and VH are not defined for the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The arrows indicate the free floating normalization factors which are acting on various regions.

exceeds 350 GeV. The 18 parameters to be measured can be seen in Figure 6.14, together with the corresponding signal regions designed to enhance the sensitivity to them. All the POIs have dedicated SRs except for the one corresponding to the ggF contribution with $N_{\text{jets}} \geq 2$ and $m_{\text{jj}} \geq 350$ GeV, which populates the VBF_0 SRs.

6.8.1.1 Simplifying the likelihood model: pruning, symmetrization and smoothing

The large number of signal and control regions, as well as the number of normalization factors, nuisance parameters and parameters of interest in the fit, leads to a complex fit model, whose minimization can be computationally expensive and challenging. Furthermore, the implementation of the STXS measurement requires defining separate samples for each of the signal processes (i.e for each of the STXS bins). In many cases, due to the limited size of the available statistics in the MC samples, the shape effects introduced by the NPs in the likelihood model are significantly impacted by statistical fluctuations. As a result, systematic variations with a small impact can be influenced by a statistical noise that can be larger than the actual systematic effect. This noise can lead to instabilities in the likelihood

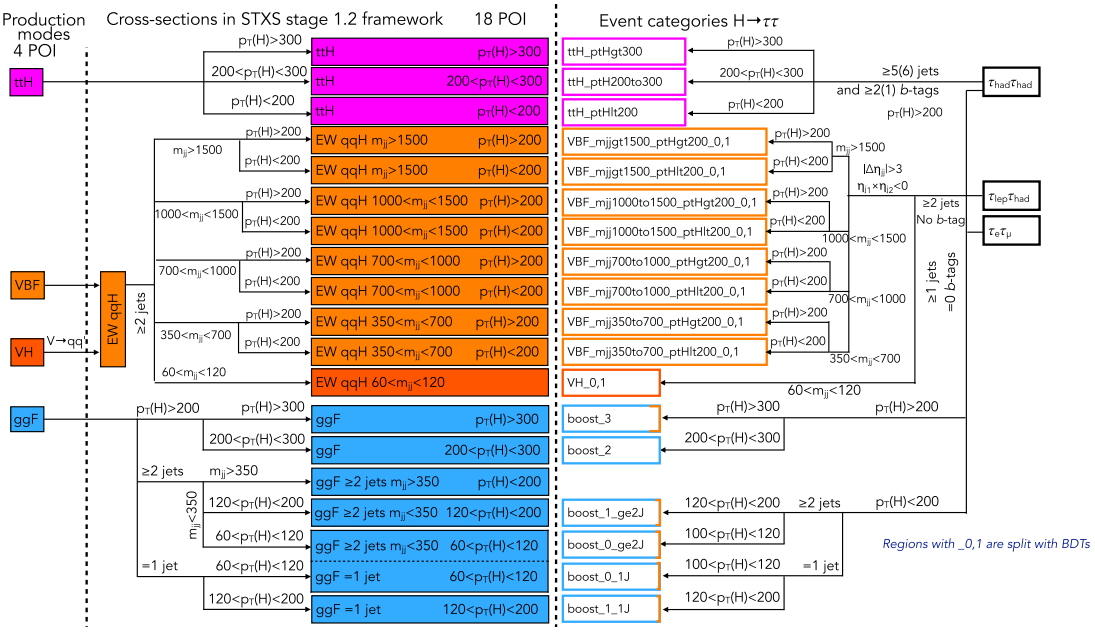


Figure 6.14: Schematic view of the fit model showing at the left the definition within the STXS framework of the 18 POIs that are measured in the analysis and on the right the signal categorization performed to enhance the sensitivity to the different POIs.

minimization and can increase artificially the impact of some systematic uncertainties in the fit results. The analysis incorporates several techniques to mitigate these issues, simplifying when possible the likelihood model and suppressing the noise introduced by statistical fluctuations in the systematic templates that are used to create the likelihood model, while preserving the meaningful shape variations.

The first step of the simplification process is the so-called pre-pruning, where samples and systematic uncertainties that do not have a significant impact on the fit model are removed. This step reduces significantly the computational cost of the negative log-likelihood minimization, eliminating also systematic uncertainties with high statistical fluctuations. Samples with an expected event yield below 0.01 events are pruned both in signal and control regions. If the sample is not removed, but the expected event yield is less than 1% of the expected signal yields (total yields) in the corresponding signal region (control region), then the systematic uncertainties associated with this sample are pruned. Furthermore, if the relative statistical uncertainty in the sample yields exceeds 10%, the kinematic systematic uncertainties are also pruned, as they are considered to be artificially inflated by the statistical fluctuations in the MC samples.

After pre-pruning, the systematic uncertainties retained in the likelihood model undergo symmetrization and smoothing for each input sample. These steps are necessary to mitigate unphysical fluctuations in the systematic templates, which can arise from statistical variations in the MC samples and potentially lead to instabilities in the fit.

The process begins by addressing any negative bin entries in the systematic templates. If such an issue is detected, the smaller variation is corrected by inverting its sign. Next, the templates are examined to identify cases where both the $+1\sigma$ and -1σ variations shift in the same direction. When this occurs, the sign of the smaller variation is again inverted to ensure proper symmetry.

Once these corrections are made, a smoothing procedure is applied to the templates to reduce statistical noise. After smoothing, the templates are rechecked to ensure that no remaining bins exhibit the same-direction shifts for the $+1\sigma$ and -1σ variations. If any such cases are found, they are corrected accordingly.

In regions dominated by limited statistics, such as VH and $t\bar{t}H$, a full symmetrization procedure is applied, as systematic uncertainties play a less significant role in these categories. In these regions, all bins are symmetrized by centering their variations around zero, ensuring stability.

Finally, pruning is applied after symmetrization and smoothing. The shape and/or normalization component of each NP, as well as the statistical uncertainty in any given bin, is removed from the fit model according to specific criteria:

- Bin-by-bin statistical uncertainties on background predictions are removed if the relative uncertainty in a bin is less than 1%.

- The normalization component of an NP is removed if its relative uncertainty on the event yield is less than 1%.

- The shape component of an NP is removed if the relative difference across all bins is less than 1%, or if the systematic uncertainties across all bins of the $m_{\tau\tau}$ distribution, binned into $[0, 90, 110, 140, \infty]$ GeV, agree within one standard deviation of their statistical uncertainty.

6.8.2 Fit model for the fiducial differential cross-section measurement

The likelihood function for the differential cross-section measurement incorporates data from both signal regions and control regions to constrain the background contributions and extract the Higgs boson signal, as in the STXS measurement. The differential cross-section measurement is focused on the VBF production mode.

Each analysis channel includes eight SRs, which are defined by splitting the events based on the BDT output into high- and low-purity regions (VBF_0 and VBF_1), as well as the binning detailed in Table 6.9. To further constrain the backgrounds, two control regions targeting the $Z \to \tau\tau$ plus jets process are introduced for each channel, but without splitting based on binning of the variables to be unfolded. These CRs correspond to the inclusive VBF_0 and VBF_1 signal regions and are constructed using the kinematic embedding method. Additionally, a control region enriched in $t\bar{t}$ events is defined for each of the $\tau_e\tau_\mu$ and $\tau_{\text{lep}}\tau_{\text{had}}$ channels.

The $m_{\tau\tau}$ distribution serves as the primary observable for extracting information about the cross-section with respect to the SM prediction, while single-bin distributions are used for each CR to constrain background normalizations. At detector level, the $m_{\tau\tau}$ distributions are produced for each bin of the observables.

The differential cross-section measurements are performed using the profile-likelihood unfolding method in the signal regions, with the fit parameters corresponding to the Higgs production cross-sections in each bin at particle level. The profile likelihood unfolding method is described in detail in Appendix A. No regularization is applied in the fit since the bin-to-bin migrations are small for all observables under consideration. The unfolding procedure is tested for potential biases arising from the choice of the MC model, ensuring that it can reliably reproduce altered shapes within the statistical precision of the data.

The systematic uncertainties are treated in the same way as in the STXS measurement, introducting nuisance parameters that are constrained by Gaussian terms in the likelihood. The systematic templates are corrected to avoid negative entries and to ensure that the $+1\sigma$ and -1σ variations are in opposite directions. No smoothing is applied to the systematic templates in the case of the unfolded differential cross-section measurement.

6.9 Simplified template cross-section measurement results

This section presents the results of the cross-section measurements using the three different fit models used in the analysis: the 1-POI, 4-POI and 18-POI fits.

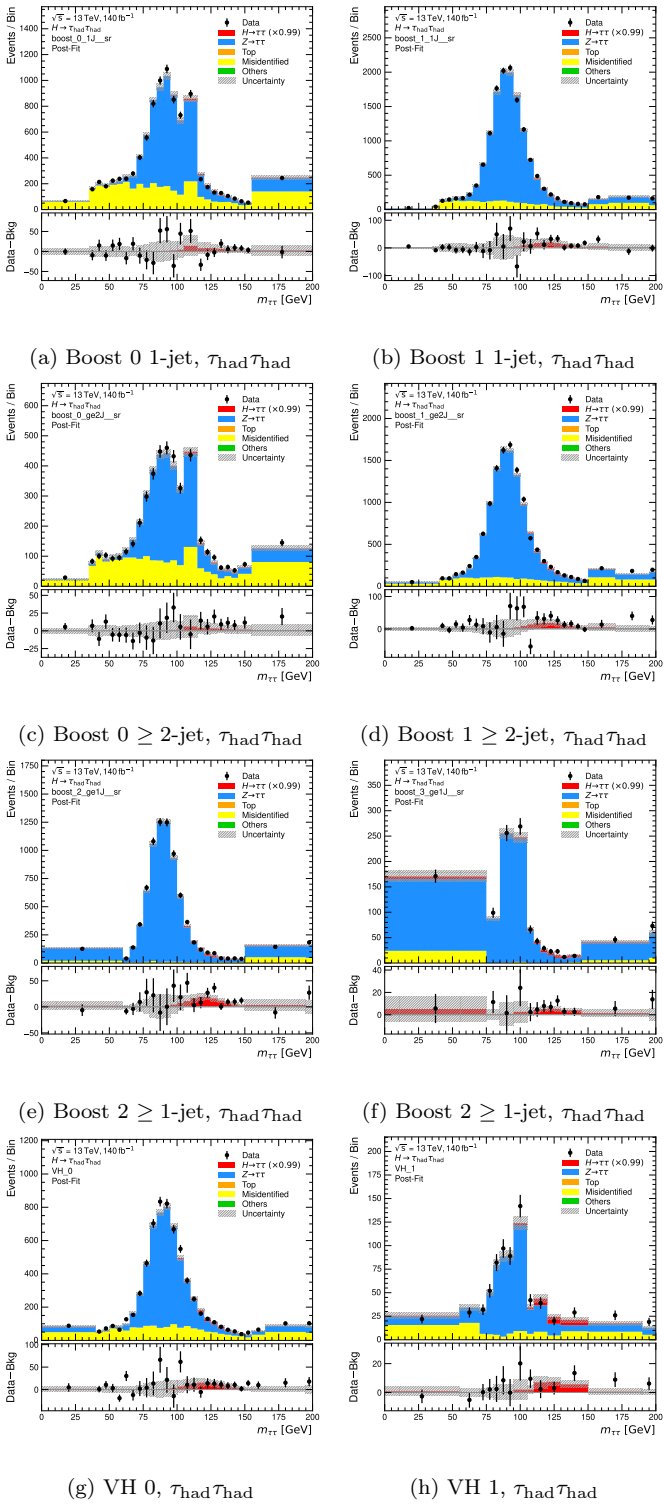
6.9.1 Post-fit distributions

In each of the signal regions of the analysis the $m_{\tau\tau}$ distribution is fitted to the data in order to extract the signal strength modifiers that scale the signal yields with the different signal parametrizations. The binning of the $m_{\tau\tau}$ distribution is optimized in each signal region to guarantee that the statistical uncertainties on the background sample size are below 20% in each bin. The binning algorithm uses as input the $m_{\tau\tau}$ distribution up to 200 GeV with a bin width of 10 GeV. The algorithm starts checking if the last bin of the distribution has a statistical uncertainty below 20%. If this is not the case, the last bin is merged with the previous one and the statistical uncertainty is recomputed. This process is repeated until the statistical uncertainty of the last bin is below 20%. Then the algorithm moves to the previous bin and repeats the process. This is done iteratively until the first bin of the distribution is reached. If the first bin has a statistical uncertainty above 20%, the first bin is merged with the second one. The binning algorithm is applied separately to the bins below and above 110 GeV. In this way, the algorithm ensures that the $Z \to \tau\tau$ and the signal components are kept in separate bins, since bins below 110 GeV, dominated by the $Z \to \tau\tau$ background, are never merged with bins above 110 GeV, where the $H \to \tau\tau$ contribution is expected. Maintaining the statistical uncertainties on the background sample size below 20% in each bin ensures that the signal strength measurements are not biased by the statistical fluctuations in the background sample size.

The post-fit $m_{\tau\tau}$ distributions in the different signal regions of the analysis are shown in Figures 6.15-6.23. The binning of the distributions corresponds to the optimized binning described above. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The agreement between the data and the prediction is generally good. The normalization factors and nuisance parameters that affect the signal and background predictions are discussed in the next section.

6.9.2 Background normalization factors

This section shows the best-fit values of the background normalization factors that scale the $Z \to \tau\tau$ and top-quark related backgrounds in the analysis categories. The normalization factors are shown for the $\tau_{\text{had}}\tau_{\text{had}}$, $\tau_{\text{lep}}\tau_{\text{had}}$, and $\tau_e\tau_\mu$ channels in Figures 6.25, 6.26, and 6.27, using the three different fit setups. All the top-quark NFs are compatible with unity within one standard deviation. The $Z \to \tau\tau$ normalization factors are also compatible with unity within one standard



(a) Boost 0 1-jet, $\tau_{\text{had}}\tau_{\text{had}}$

(b) Boost 1 1-jet, $\tau_{\text{had}}\tau_{\text{had}}$

(c) Boost 0 $\geq$ 2-jet, $\tau_{\text{had}}\tau_{\text{had}}$

(d) Boost 1 $\geq$ 2-jet, $\tau_{\text{had}}\tau_{\text{had}}$

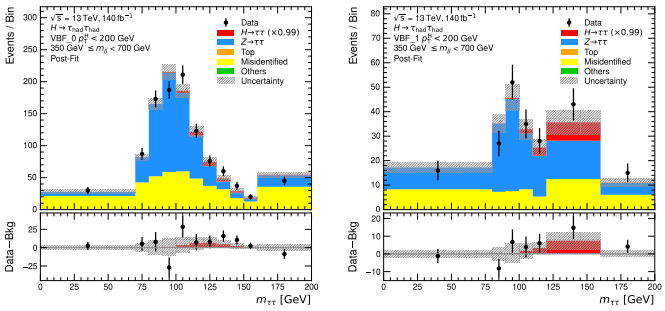
(e) Boost 2 $\geq$ 1-jet, $\tau_{\text{had}}\tau_{\text{had}}$

(f) Boost 2 $\geq$ 1-jet, $\tau_{\text{had}}\tau_{\text{had}}$

(g) VH 0, $\tau_{\text{had}}\tau_{\text{had}}$

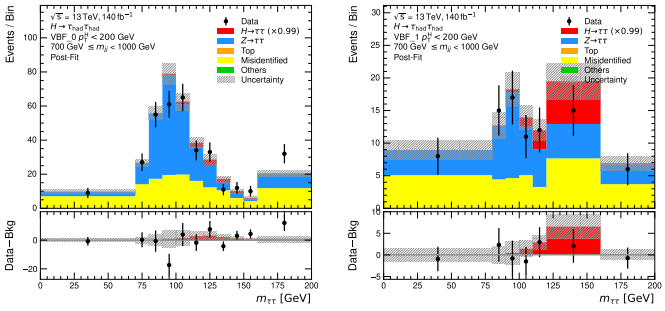
(h) VH 1, $\tau_{\text{had}}\tau_{\text{had}}$

Figure 6.15: Post-fit $m_{\tau\tau}$ distributions boost and VH categories in the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.



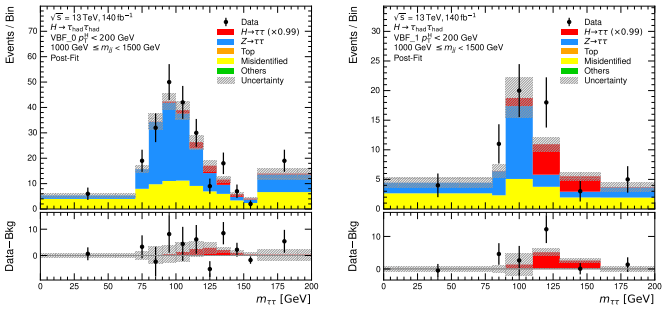
(a) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$

(b) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



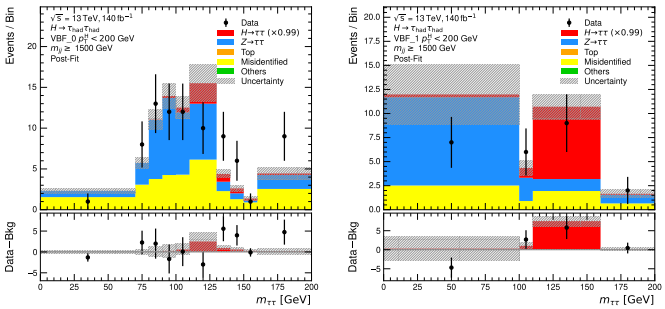
(c) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$

(d) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



(e) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$

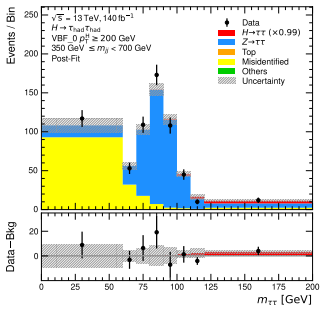
(f) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



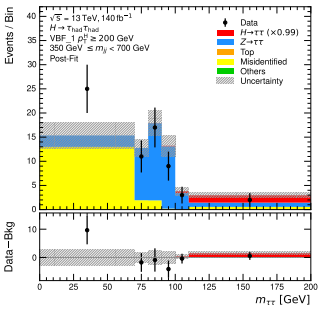
(g) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $m_{\mathrm{jj}} > 1500$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$

(h) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $m_{\mathrm{jj}} > 1500$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$

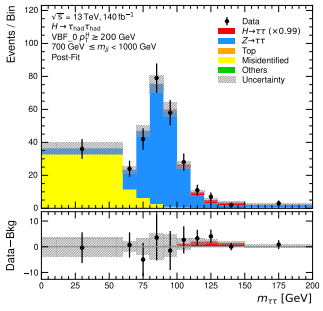
Figure 6.16: Post-fit $m_{\tau\tau}$ distributions $p_{\mathrm{T}}^{H} < 200$ GeV VBF categories in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.



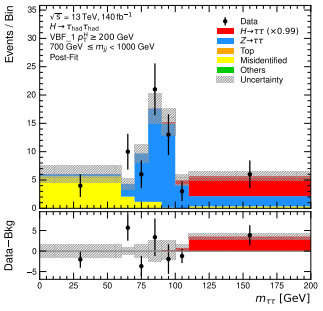
(a) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



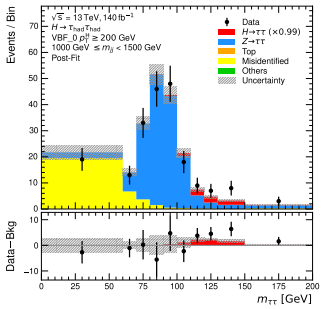
(b) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



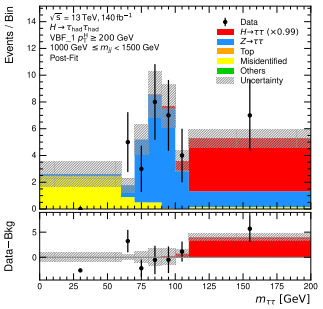
(c) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



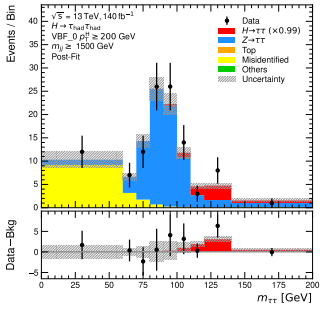
(d) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



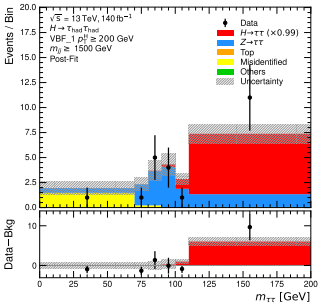
(e) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



(f) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$

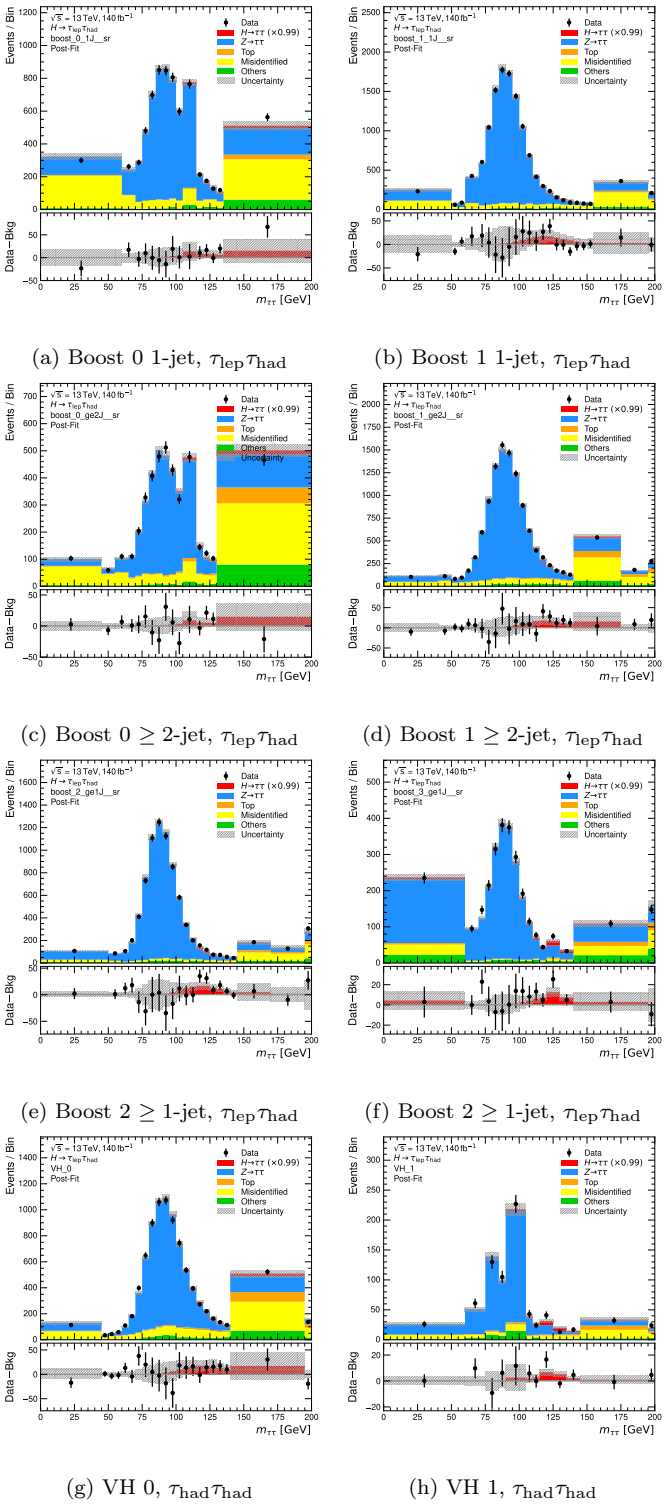


(g) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $m_{\mathrm{jj}} > 1500$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$



(h) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $m_{\mathrm{jj}} > 1500$, $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$

Figure 6.17: Post-fit $m_{\tau\tau}$ distributions $p_{\mathrm{T}}^{H} \geq 200$ GeV VBF categories in the $\tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.



(a) Boost 0 1-jet, $\tau_{\text{lep}}\tau_{\text{had}}$

(b) Boost 1 1-jet, $\tau_{\text{lep}}\tau_{\text{had}}$

(c) Boost 0 $\geq$ 2-jet, $\tau_{\text{lep}}\tau_{\text{had}}$

(d) Boost 1 $\geq$ 2-jet, $\tau_{\text{lep}}\tau_{\text{had}}$

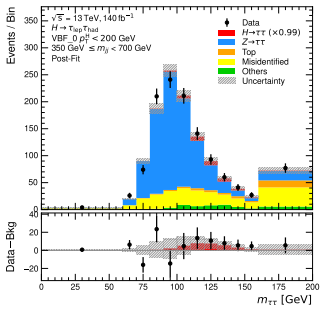
(e) Boost 2 $\geq$ 1-jet, $\tau_{\text{lep}}\tau_{\text{had}}$

(f) Boost 2 $\geq$ 1-jet, $\tau_{\text{lep}}\tau_{\text{had}}$

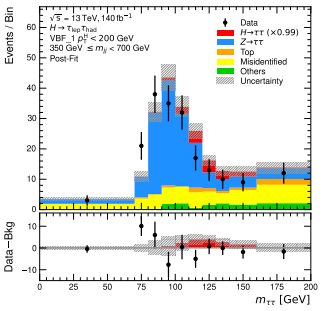
(g) VH 0, $\tau_{\text{had}}\tau_{\text{had}}$

(h) VH 1, $\tau_{\text{had}}\tau_{\text{had}}$

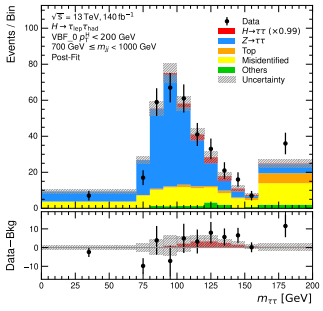
Figure 6.18: Post-fit $m_{\tau\tau}$ distributions boost and VH categories in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.



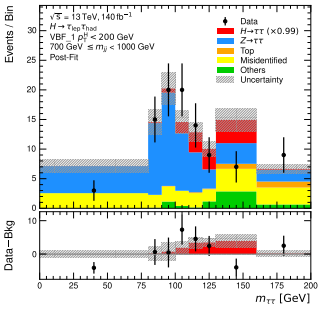
(a) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



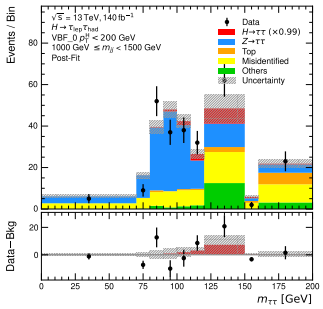
(b) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



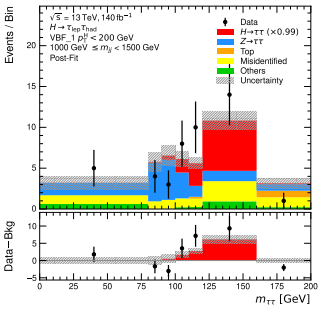
(c) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



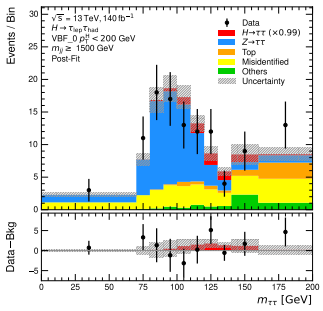
(d) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



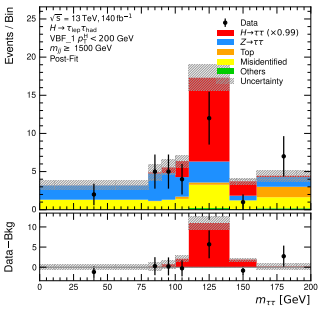
(e) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



(f) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$

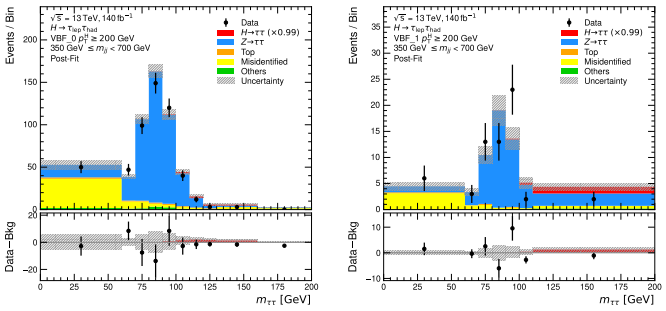


(g) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $m_{\mathrm{jj}} > 1500$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



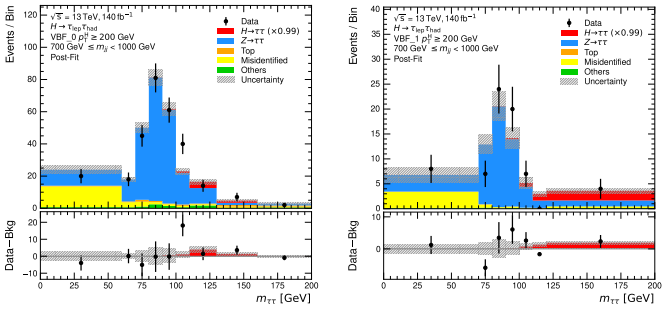
(h) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $m_{\mathrm{jj}} > 1500$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$

Figure 6.19: Post-fit $m_{\tau\tau}$ distributions $p_{\mathrm{T}}^{H} < 200$ GeV VBF categories in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.



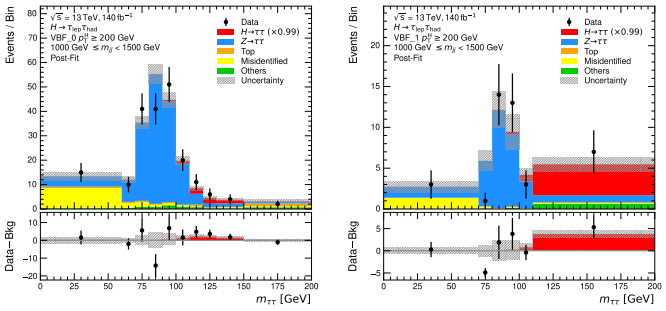
(a) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$

(b) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



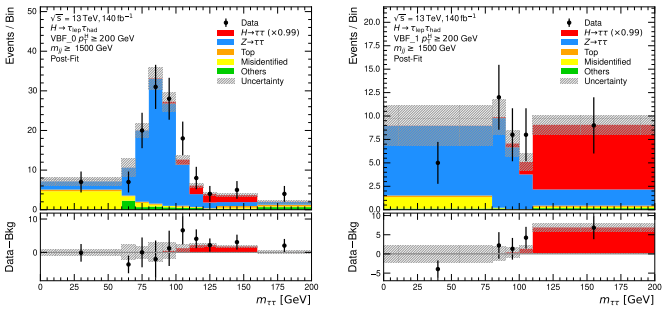
(c) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$

(d) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



(e) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$

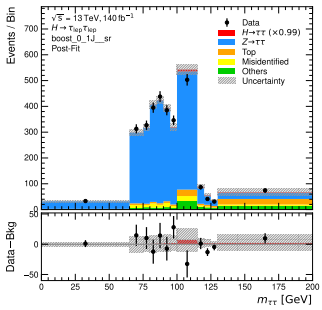
(f) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



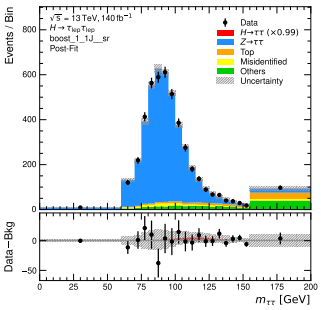
(g) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $m_{\mathrm{jj}} > 1500$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$

(h) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $m_{\mathrm{jj}} > 1500$, $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$

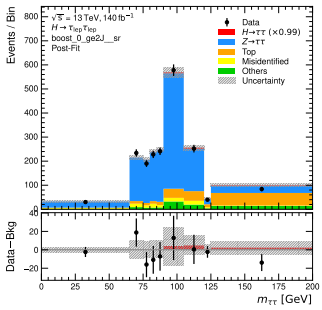
Figure 6.20: Post-fit $m_{\tau\tau}$ distributions $p_{\mathrm{T}}^{H} \geq 200$ GeV VBF categories in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.



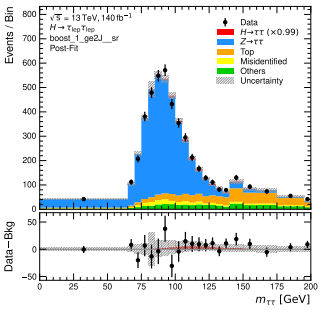
(a) Boost 0 1-jet, $\tau_e\tau_\mu$



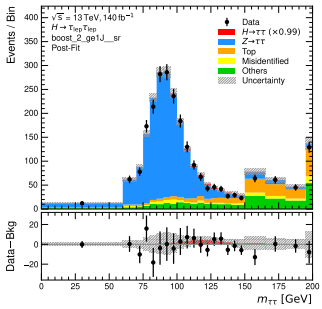
(b) Boost 1 1-jet, $\tau_e\tau_\mu$



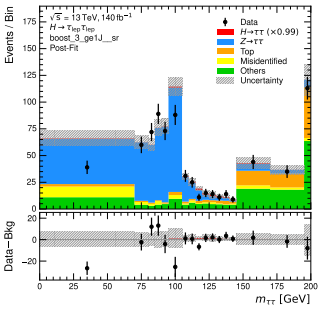
(c) Boost 0 $\geq$ 2-jet, $\tau_e\tau_\mu$



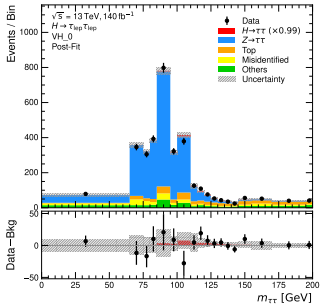
(d) Boost 1 $\geq$ 2-jet, $\tau_e\tau_\mu$



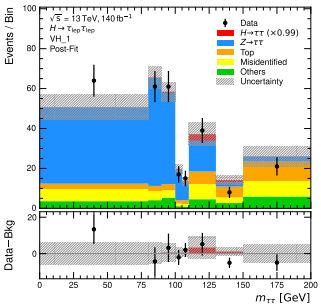
(e) Boost 2 $\geq$ 1-jet, $\tau_e\tau_\mu$



(f) Boost 2 $\geq$ 1-jet, $\tau_e\tau_\mu$

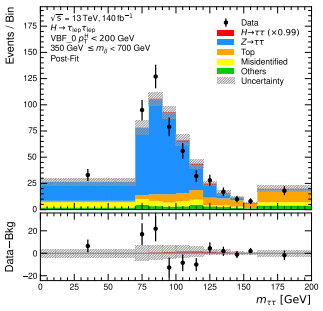


(g) VH 0, $\tau_{\text{had}}\tau_{\text{had}}$

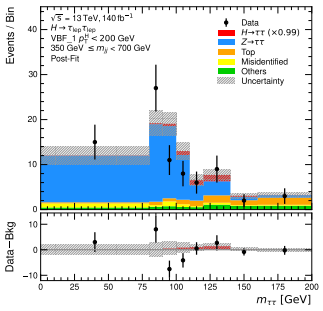


(h) VH 1, $\tau_{\text{had}}\tau_{\text{had}}$

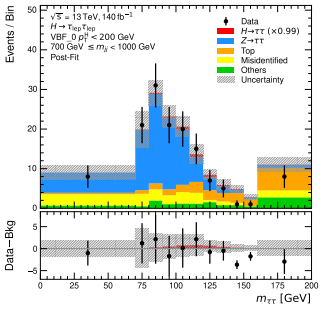
Figure 6.21: Post-fit $m_{\tau\tau}$ distributions boost and VH categories in the $\tau_e\tau_\mu$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.



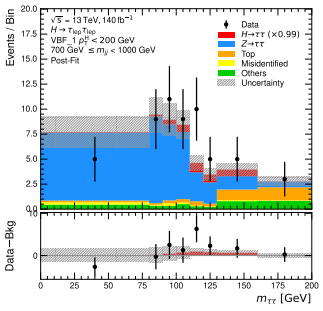
(a) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_e\tau_\mu$



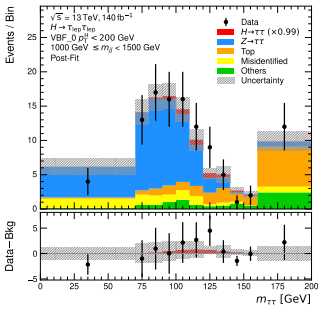
(b) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_e\tau_\mu$



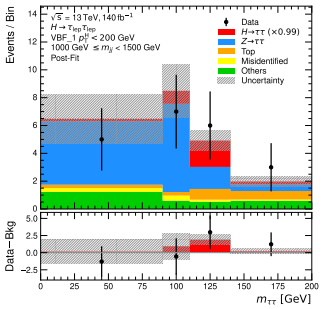
(c) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_e\tau_\mu$



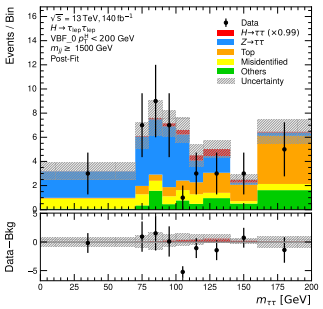
(d) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_e\tau_\mu$



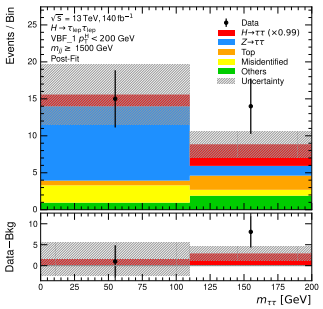
(e) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_e\tau_\mu$



(f) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_e\tau_\mu$

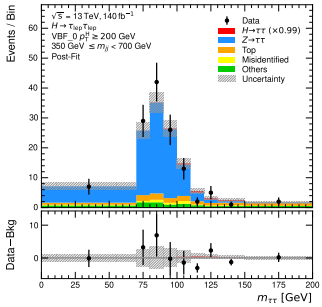


(g) VBF 0, $p_{\mathrm{T}}^{H} < 200$, $m_{\mathrm{jj}} > 1500$, $\tau_e\tau_\mu$

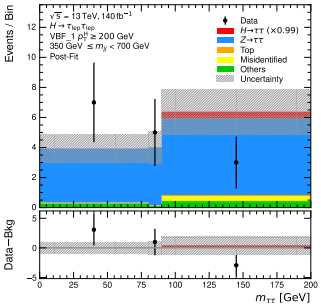


(h) VBF 1, $p_{\mathrm{T}}^{H} < 200$, $m_{\mathrm{jj}} > 1500$, $\tau_e\tau_\mu$

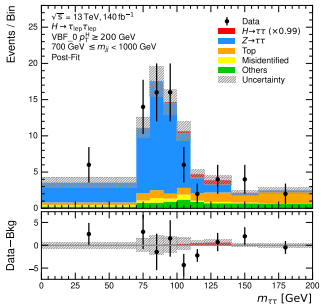
Figure 6.22: Post-fit $m_{\tau\tau}$ distributions $p_{\mathrm{T}}^{H} < 200$ GeV VBF categories in the $\tau_e\tau_\mu$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.



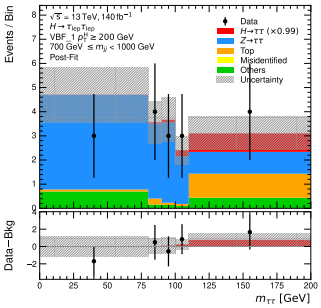
(a) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_e\tau_\mu$



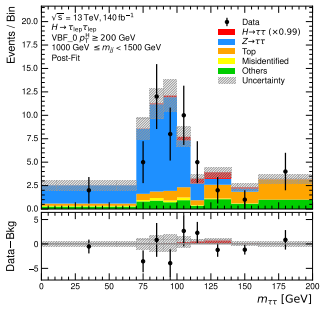
(b) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $350 < m_{\mathrm{jj}} < 700$, $\tau_e\tau_\mu$



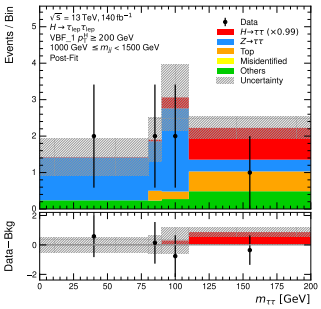
(c) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_e\tau_\mu$



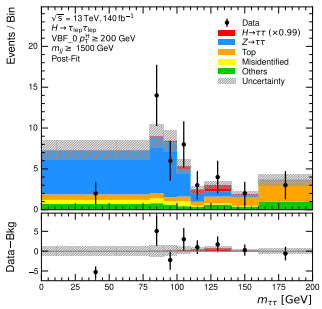
(d) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $700 < m_{\mathrm{jj}} < 1000$, $\tau_e\tau_\mu$



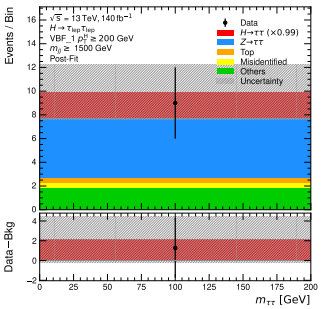
(e) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_e\tau_\mu$



(f) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $1000 < m_{\mathrm{jj}} < 1500$, $\tau_e\tau_\mu$



(g) VBF 0, $p_{\mathrm{T}}^{H} > 200$, $m_{\mathrm{jj}} > 1500$, $\tau_e\tau_\mu$



(h) VBF 1, $p_{\mathrm{T}}^{H} > 200$, $m_{\mathrm{jj}} > 1500$, $\tau_e\tau_\mu$

Figure 6.23: Post-fit $m_{\tau\tau}$ distributions $p_{\mathrm{T}}^{H} \geq 200$ GeV VBF categories in the $\tau_e\tau_\mu$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The binning of the distributions corresponds to the optimized binning used in the statistical analysis to ensure that the statistical uncertainties on the background sample size are below 20% in each bin.

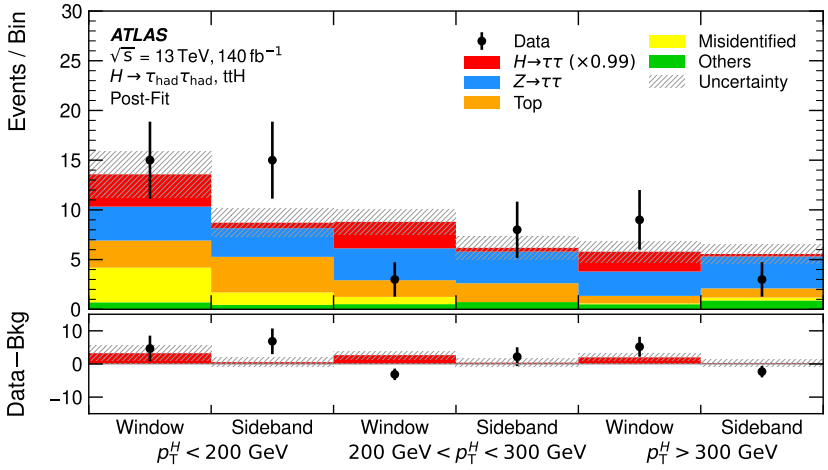


Figure 6.24: Post-fit yields in the $t\bar{t}H$ categories of the $\tau_{\text{had}}\tau_{\text{had}}$ channel. The signal yields are scaled by the signal strength obtained from the 1-POI fit. The window categories contains events with $\text{m}_{\tau\tau}$ in the range [100, 150] GeV, while events outside of this mass region are included in the sideband region.

deviation for the boost and VH categories. In the VBF categories, the $Z \to \tau\tau$ normalization factors show a trend towards lower values at high m_{jj}, independently of the analysis channel or the p_{T}^H regime. The normalization factors correct for the m_{jj} mismodeling of the $Z \to \tau\tau$ background predicted by the simulation with the help of the embedding method. The normalization factors are generally in agreement between the VBF_0 and VBF_1 categories that are defined for the same p_{T}^H and m_{jj} regime.

The uncertainties on the normalization factors are dominated by systematic uncertainties in the boost categories, due to the larger dataset. In the VH, VBF and $t\bar{t}H$ categories, the statistical uncertainties dominate. The best-fit values of the normalization factors are well in agreement between the three fit setups. The changes in the signal parametrization do not affect the background estimation.

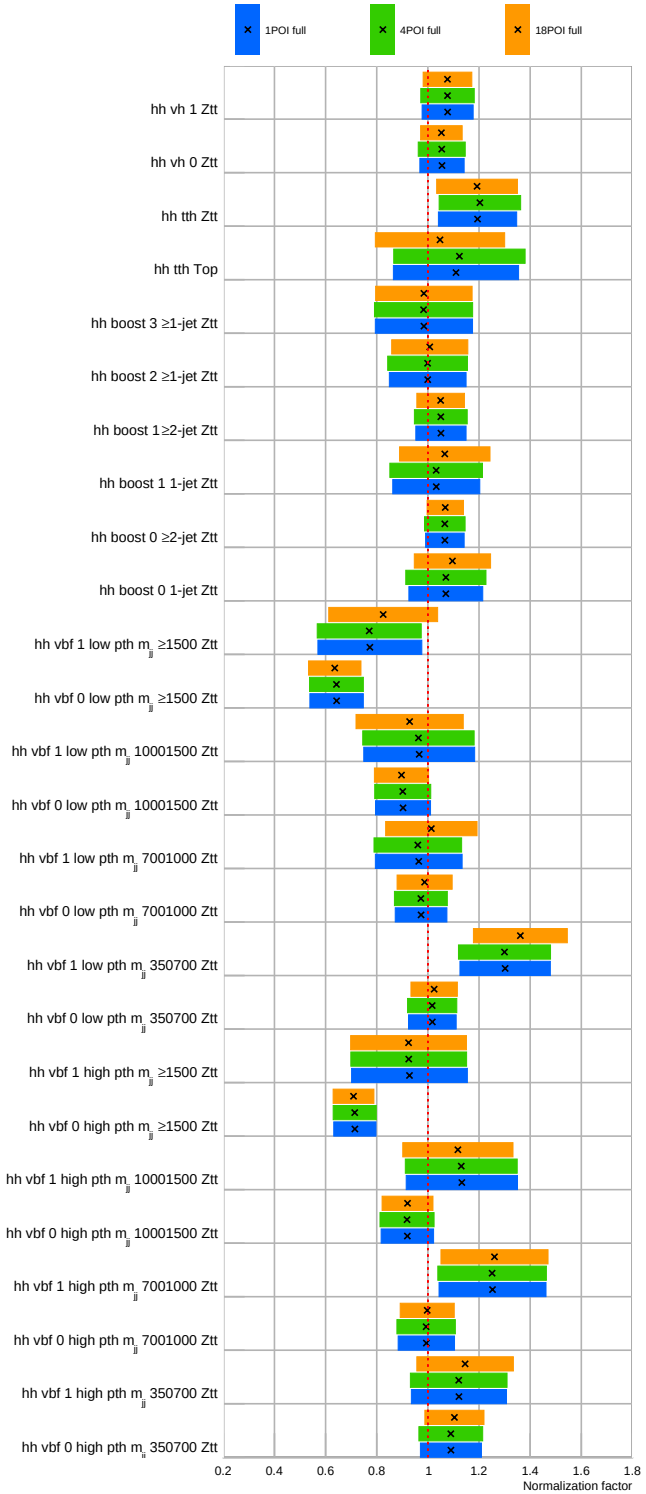


Figure 6.25: Background normalization factors for the top-quark and $Z \to \tau\tau$ backgrounds in the $\tau_{\text{had}}\tau_{\text{had}}$ channel.

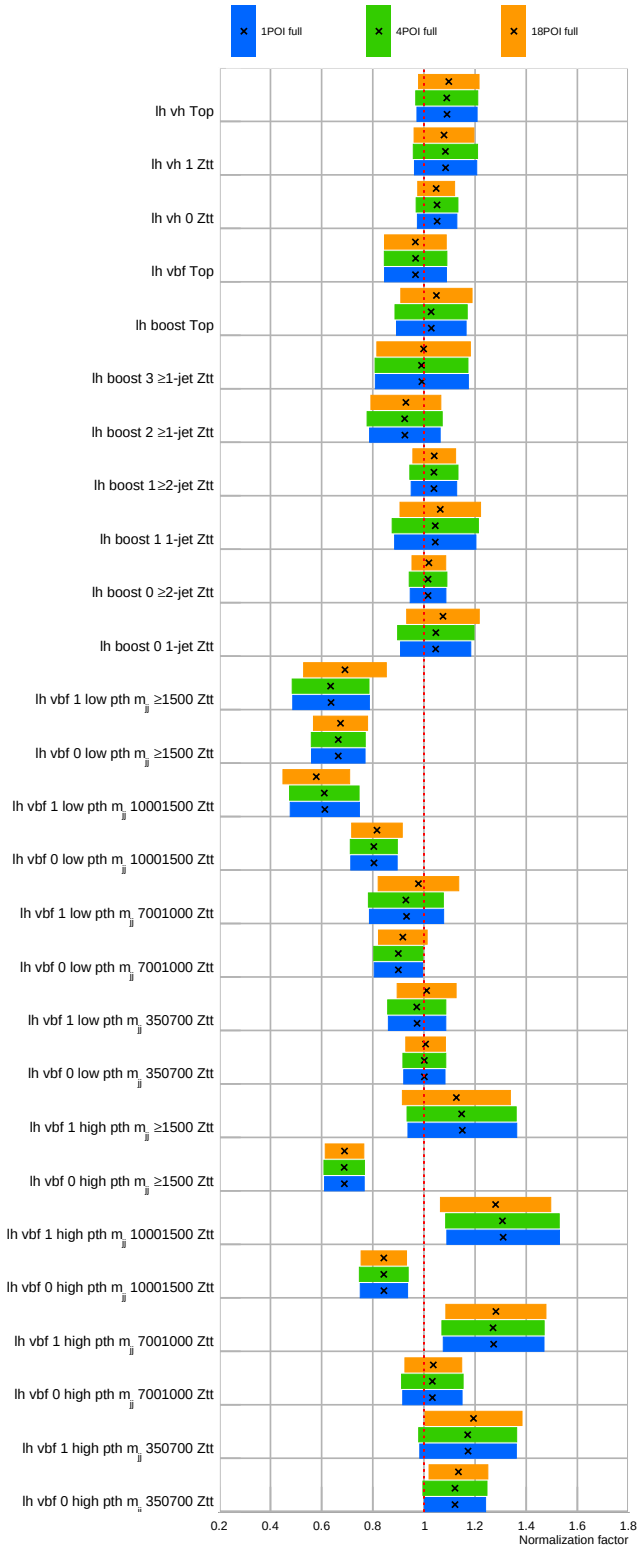


Figure 6.26: Background normalization factors for the top-quark and $Z \to \tau\tau$ backgrounds in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel.

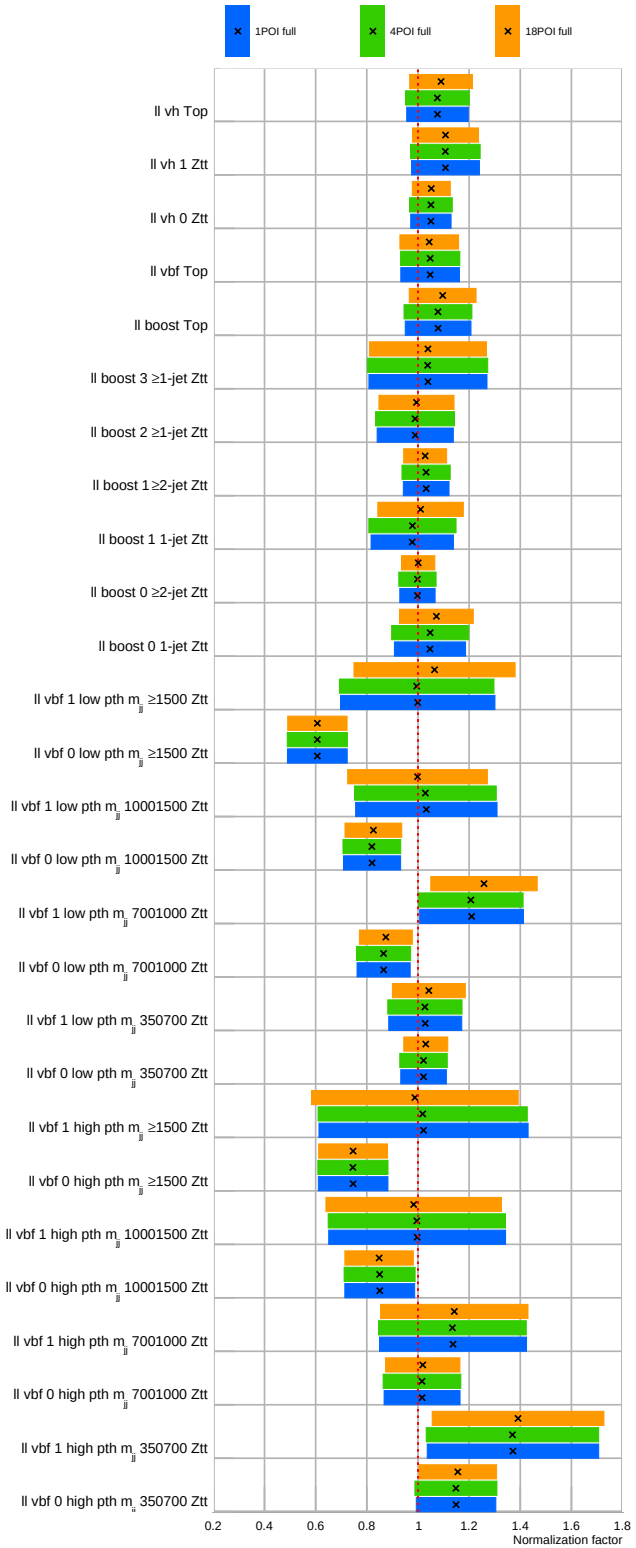


Figure 6.27: Background normalization factors for the top-quark and $Z \to \tau\tau$ backgrounds in the $\tau_e\tau_\mu$ channel.

6.9.3 Nuisance parameters

This section describes the behavior of the nuisance parameters that pass the pruning criteria. Nuisance parameters are introduced in the likelihood function to account for systematic uncertainties, and while they are not the primary focus of the fit, they play a crucial role in the fit results. Each nuisance parameter is constrained by additional functions incorporated into the likelihood model, which draw on information from auxiliary measurements, theoretical calculations, or dedicated studies. The fit may determine that the best-fit value of a nuisance parameter differs from the nominal value based on auxiliary measurements; in this case, the parameter is said to be *pulled* by the fit. Additionally, if the analysis data are sensitive to the systematic uncertainty represented by a nuisance parameter, the fit may reduce its pre-fit uncertainty, resulting in a *constraint* on the parameter.

The following paragraphs describe the behavior of the nuisance parameters related to the systematic uncertainties in the three fit setups. The parameters are grouped by the type of systematic uncertainty they represent, and the behavior of the parameters is shown for the combined fit that includes the three analysis channels.

Figure 6.28 presents the best-fit values of the nuisance parameters related to systematic uncertainties in parton shower and hadronization modeling, the choice of matrix element generator, and the Stewart-Tackmann ad-hoc uncertainties, which were introduced to account for potential mismodeling of third-jet veto effects in the analysis. The parameters do not exhibit significant pulls, and their best-fit values are consistent with the pre-fit values within one standard deviation. Additionally, the uncertainties on these nuisance parameters are not constrained by the data. Overall, the behavior of the nuisance parameters is consistent across the three signal parameterizations used in the fit.

Figure 6.29 presents the best-fit values of the nuisance parameters related to PDF uncertainties and QCD scale uncertainties in the signal simulation. The PDF-related nuisance parameters show no significant pulls or constraints, with best-fit values consistent with the pre-fit values. Similarly, the QCD scale uncertainties are generally in good agreement with the pre-fit values. However, a pull is observed in the 18-POI fit parameterization for the nuisance parameter associated with uncertainties in the migrations below and above the $p_{\mathrm{T}}{}^{Hjj} = 25$ GeV threshold for ggF events. This parameter particularly affects the shape of the ggF signal templates in the VBF categories, especially in the ggF-dominated VBF_0 categories.

The nuisance parameters associated with systematic uncertainties in the theoretical predictions for top-quark background processes are presented in Figure 6.30a. PDF-related nuisance parameters show no significant pulls or constraints from the fit, as can be seen in Figure 6.30b. The most substantial constraints are found for parameters related to the choice of matrix element generator and final state radiation modeling in the non-$t\bar{t}H$ categories. These systematics have a large normalization effect on the top-quark background and are strongly constrained by

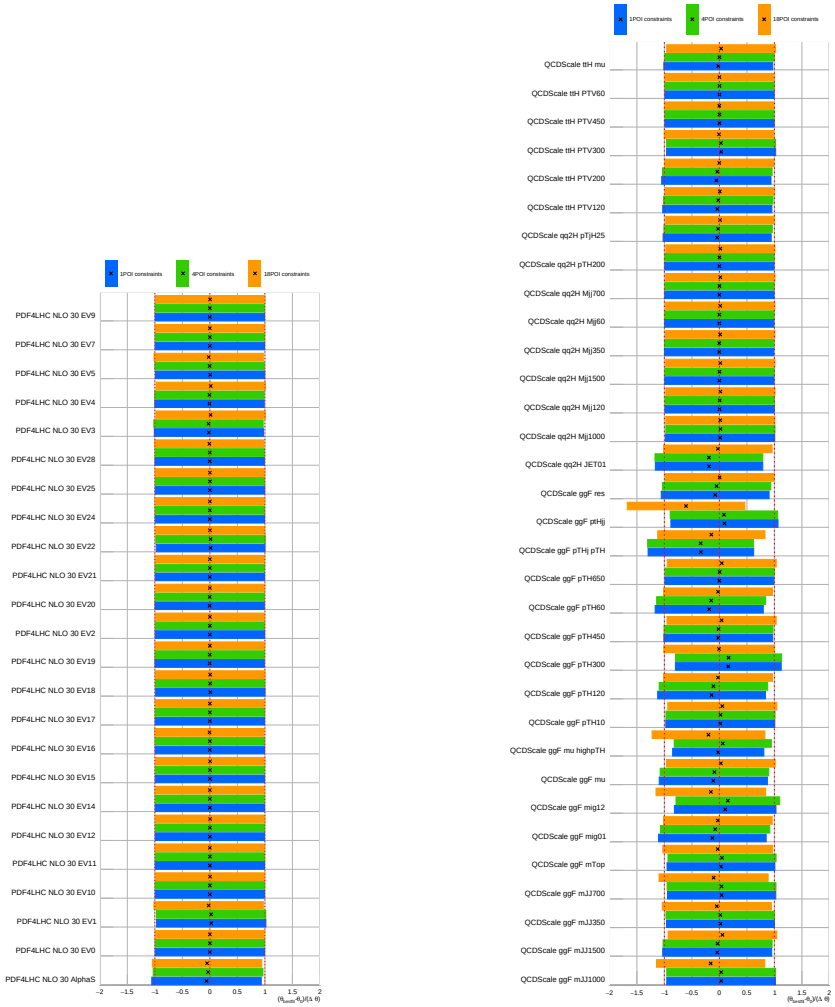


Figure 6.28: Nuisance parameters associated with signal theory uncertainties related to the parton shower and hadronization modeling, the choice of matrix element generator, and the Stewart-Tackmann ad-hoc uncertainties. The best-fit values of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.

the high-statistics top-quark control regions, which are effective in reducing these uncertainties. Additionally, these nuisance parameters exhibit strong correlations with the top-quark background normalization factors.

Figures 6.31 to 6.34 display the best-fit values of the nuisance parameters associated with theoretical uncertainties in the prediction of the $Z \to \tau\tau$ background. These uncertainties arise from various sources, including the value of the strong coupling constant, the matching scale in the parton shower, the ad-hoc treatment of the electroweak component of the $Z \to \tau\tau$ background in the analysis, the propagation of PDF uncertainties in the simulation, and variations in the renormalization and factorization scales. For PDF and QCD scale uncertainties, separate nuisance parameters are used to handle their shape and normalization components. This approach helps prevent correlations between the normalization uncertainties and the $Z \to \tau\tau$ normalization factors from influencing the background shape uncertainties. Additionally, these uncertainties are decorrelated across different analysis categories, as each category employs distinct normalization factors. While shape-related uncertainties are often pruned during the fit, the normalization uncertainties generally remain and are constrained by the data, thanks to the statistical strength of the object-level embedding technique used to construct the $Z \to \tau\tau$ control regions.

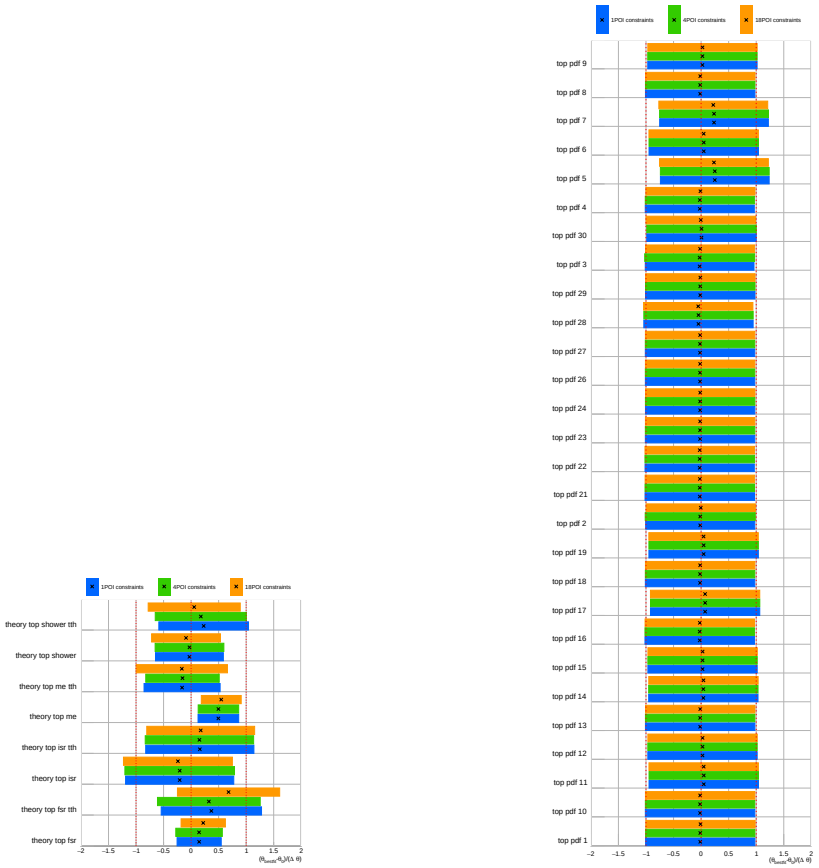
The most significant pull is observed for the nuisance parameter related to the treatment of electroweak $Z \to \tau\tau$ production in the analysis. This parameter addresses the mismodeling of the electroweak component in the $Z \to \tau\tau$ background



(a) (b)

Figure 6.29: Nuisance parameters associated with (a) PDF uncertainties in the signal simulation and (b) signal uncertainties estimated through variations of the QCD renormalization and factorization scales. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.

simulation. Its pull allows the fit to reduce the number of $Z \to \tau\tau$ events attributed to QCD production and increase those attributed to electroweak production.



(a) (b)

Figure 6.30: Nuissance parameters associated with (a) theoretical uncertainties in the top-quark background prediction and (b) PDF uncertainties in the top-quark background simulation. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.

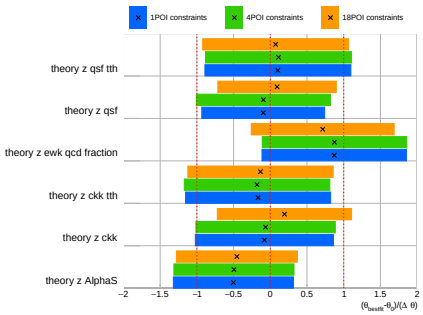
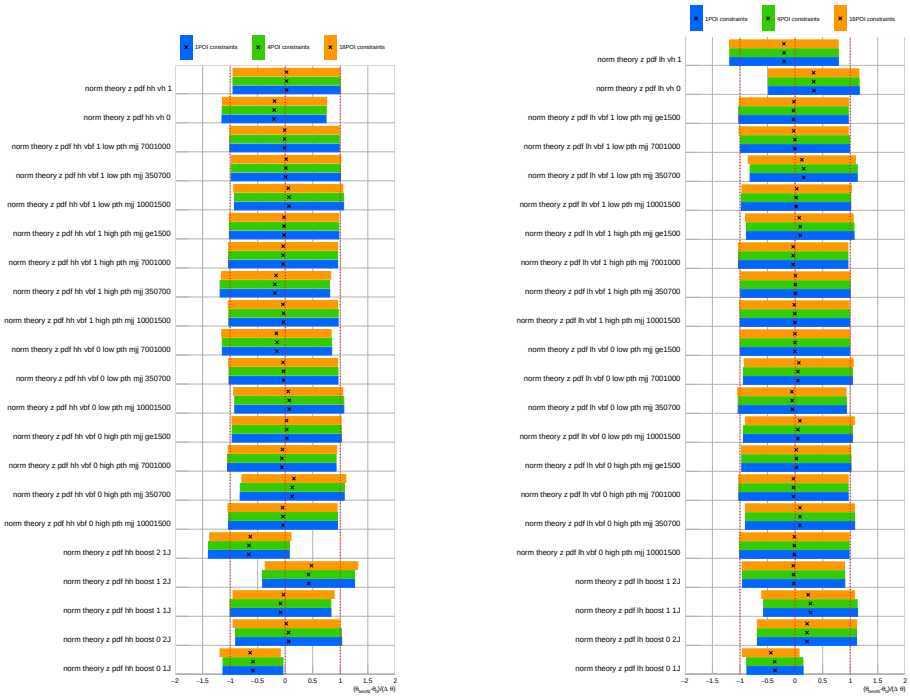
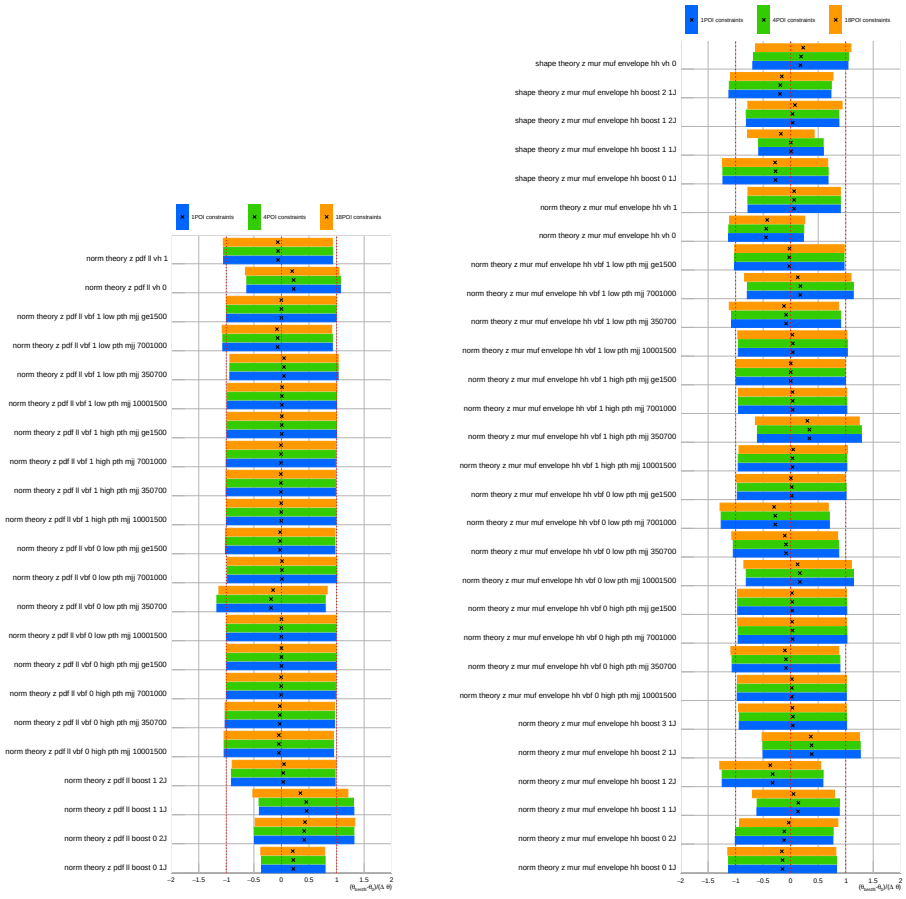


Figure 6.31: Nuisance parameters associated with theoretical uncertainties in the $Z \to \tau\tau$ background prediction. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.



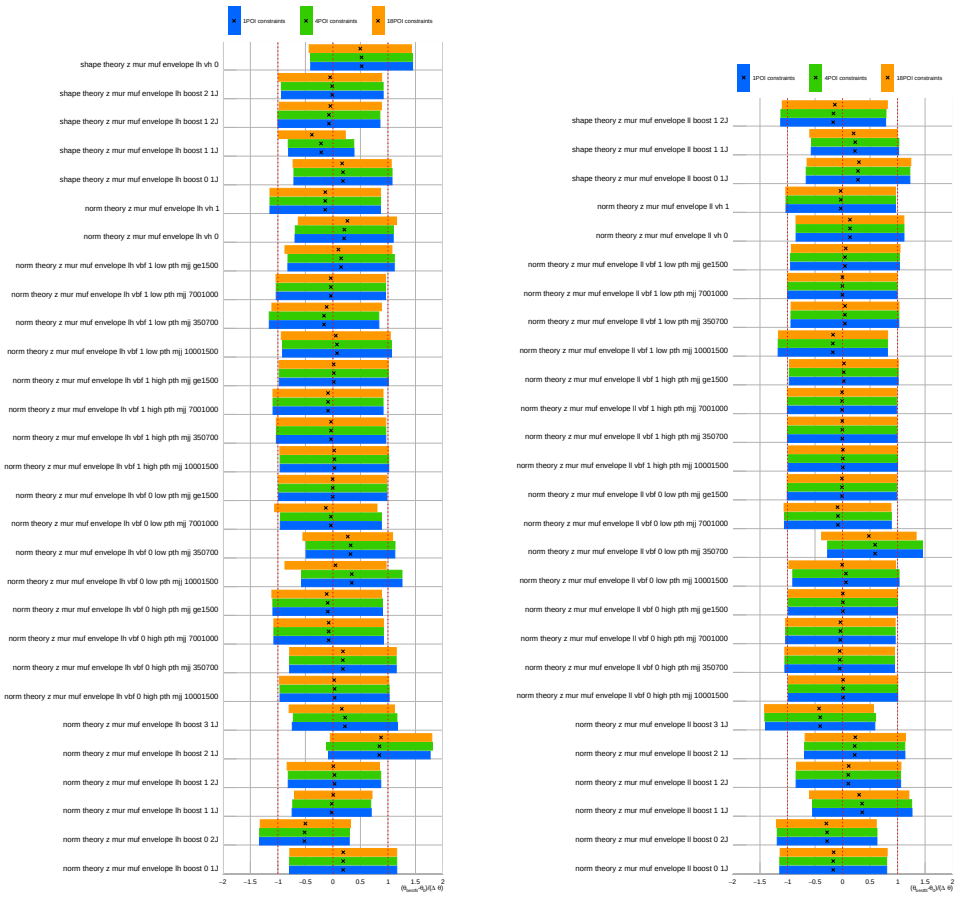
(a) (b)

Figure 6.32: Nuissance parameters associated with PDF uncertainties in the $Z \to \tau\tau$ background simulation in the (a) $\tau_{\text{had}}\tau_{\text{had}}$ and (b) $\tau_{\text{lep}}\tau_{\text{had}}$ channels. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.



(a) (b)

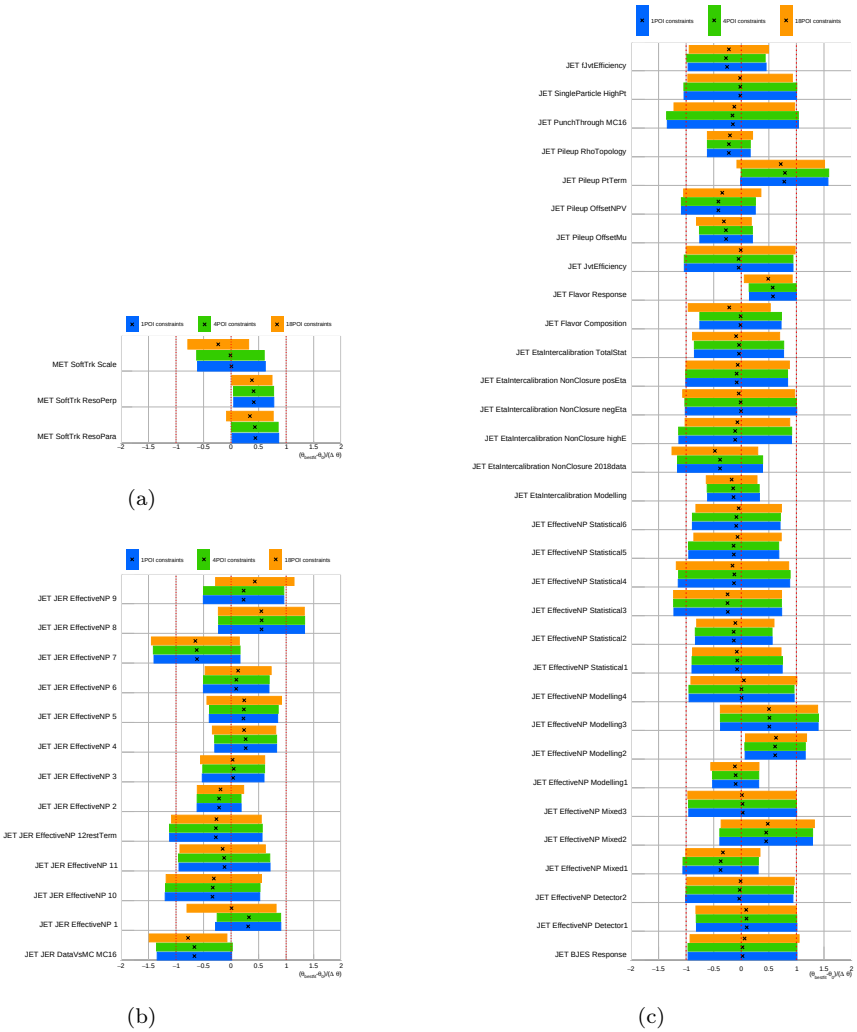
Figure 6.33: Nuisance parameters associated with (a) PDF uncertainties in the $Z \to \tau\tau$ background simulation in the $\tau_e\tau_\mu$ channel and (b) the missing higher-order terms in the $Z \to \tau\tau$ background prediction in the $\tau_{\text{had}}\tau_{\text{had}}$, calculated through variations of the renormalization and factorization scales. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.



(a) (b)

Figure 6.34: Nuisance parameters associated with the missing higher-order terms in the $Z \to \tau\tau$ background prediction in the (a) $\tau_{\text{lep}}\tau_{\text{had}}$ and (b) $\tau_e\tau_\mu$ channels, calculated through variations of the renormalization and factorization scales. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.

Jet energy scale and JER uncertainties play a crucial role in the $H \to \tau\tau$ analysis, which uses mainly the MMC algorithm to reconstruct the Higgs boson mass. This algorithm is sensitive to the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, which is influenced by the jet energy resolution. The fit constrains the nuisance parameters associated with JER uncertainties, since the $m_{\tau\tau}$ distribution is sensitive to these uncertainties. If the JER is worse than expected, the $m_{\tau\tau}$ distribution will broaden. Additionally, a reduced JER can cause pile-up jets to be promoted above energy thresholds, affecting the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation and, consequently, the mass reconstruction. The pulls and constraints on the JER-related nuisance parameters primarily arise from the boosted categories, where the large dataset allows for the exploitation of shape information to constrain and pull these parameters. Pulls and constraints are also observed for the MET nuisance parameters. The best-fit values of these nuisance parameters can be seen in Figure 6.35.



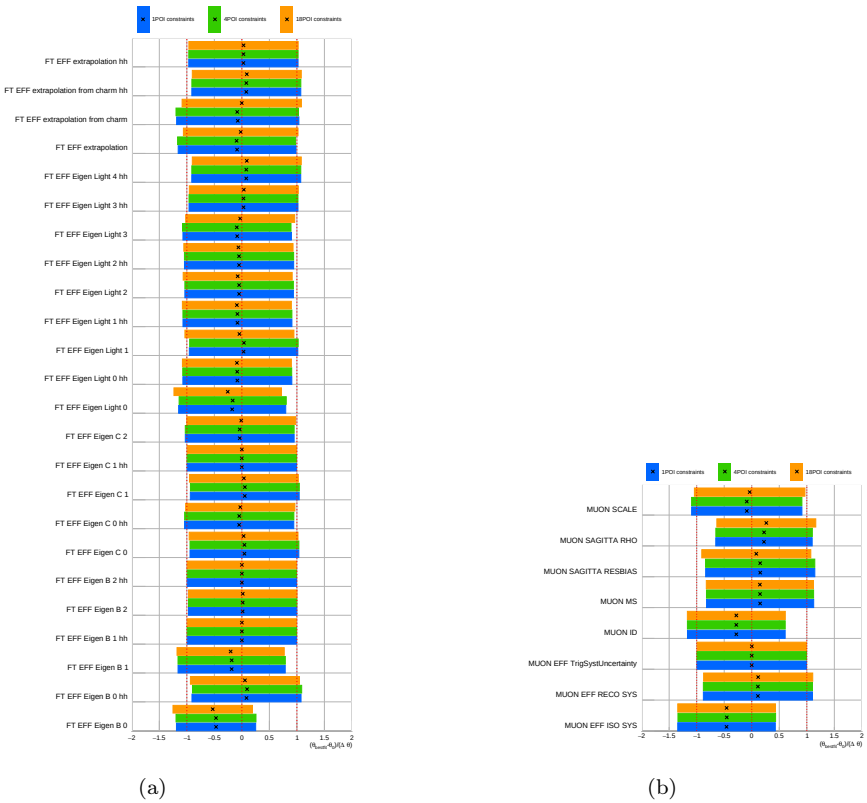
(a) (b) (c)

Figure 6.35: Nuisance parameters associated with (a) MET uncertainties, (b) JER uncertainties, and (c) jet energy scale uncertainties. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.

Figure 6.36 presents the nuisance parameters associated with flavor-tagging uncertainties, as well as muon reconstruction, isolation, trigger, identification efficiencies, and the muon momentum scale. None of these parameters show significant pulls or constraints from the fit. The most notable constraint is observed in the b-tagging efficiency nuisance parameters, due to their influence on the normalization of the $t\bar{t}$ background. This constraint arises from the analysis's sensitivity, particularly in the top-quark control regions, which are used to better constrain the normalization of these background sources.

In Figure 6.37, the nuisance parameters related to the modeling uncertainties for electrons (a), as well as τ-leptons (b), are shown. For the electron uncertainties—especially those tied to object identification with corrections for kinematic effects—there are only minor pulls and constraints. On the other hand, some of the simplified nuisance parameters related to muons exhibit pulls, though these primarily influence the $\tau_e\tau_\mu$ channel, which has limited impact on the overall results. As such, these pulls are of little consequence on the final outcome and do not require further analysis.

A few NPs associated with the τ-lepton energy scale show both pulls and constraints, with the strongest pull and constraint linked to the τ-lepton reconstruction

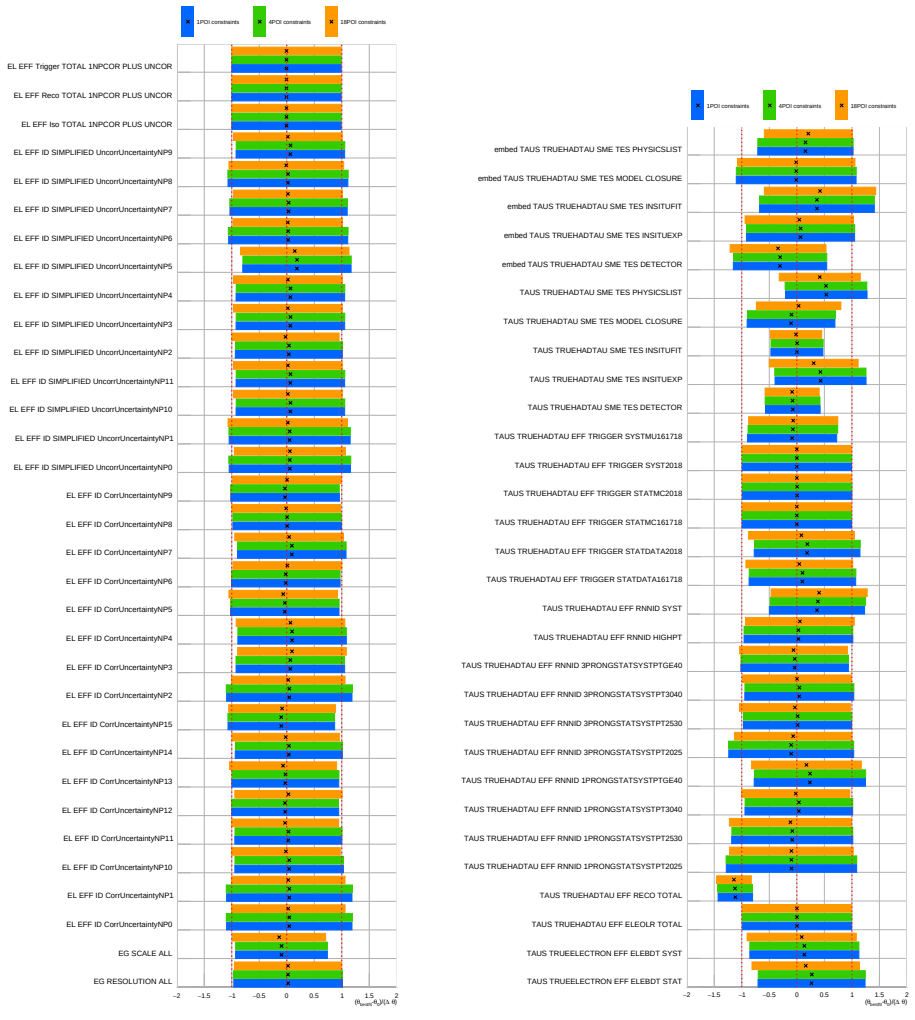


(a) (b)

Figure 6.36: Nuisance parameters associated with (a) flavor-tagging uncertainties and (b) muon reconstruction, isolation, trigger, identification efficiencies, and the muon momentum scale. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.

efficiency. This uncertainty is based entirely on simulated data, and thus does not benefit from auxiliary measurements that could correct for discrepancies between simulation and real data. The reliance on simulation is due to the absence of a high-purity event selection for $\tau_{\text{had-vis}}$ that would enable a tag-and-probe method. However, this analysis provides such a selection through object-based embedded control regions, making it more sensitive to this uncertainty than the auxiliary measurement itself.

To address this limitation, the prior for this NP is inflated by a factor of four within the likelihood model. This inflation effectively removes the Gaussian constraint on the best-fit value, allowing the NP to behave as a free parameter. As a result, the likelihood fit can directly measure the parameter and account for differences between the simulation and observed data.

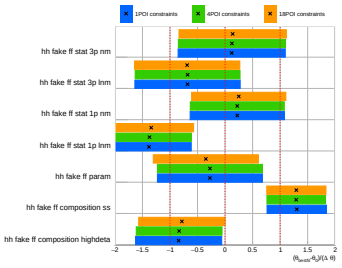


(a) (b)

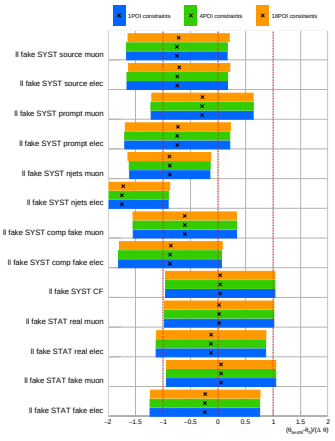
Figure 6.37: Nuisance parameters associated with (a) electron modeling uncertainties and (b) τ-lepton modeling uncertainties. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.

Figure 6.38 displays the nuisance parameters associated with the data-driven estimation of the misidentified background. Significant pulls and constraints are observed for these parameters, which is expected given that the uncertainties are conservatively derived ad-hoc from the data. These uncertainties are specifically designed to account for any mismodeling in the misidentified background predictions compared to the observed data. However, despite their sizable pulls and constraints, these parameters have a limited effect on the signal extraction.

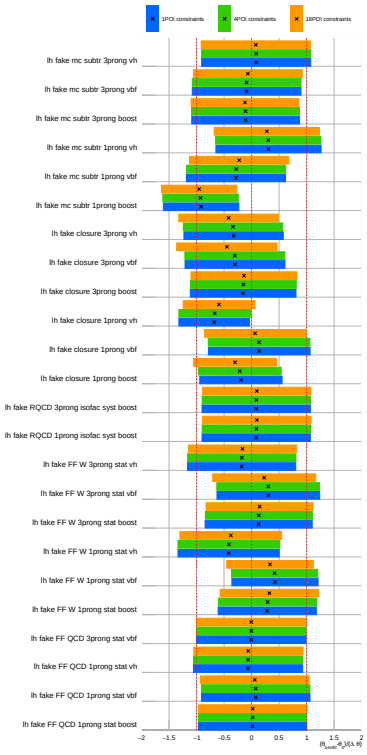
In general, all the background-related nuisance parameters show good consistency between the three fit setups.



(a)



(b)



(c)

Figure 6.38: Nuisance parameters associated with the data-driven estimation of the misidentified background. The best-fit values and the post-fit uncertainties of the nuisance parameters are shown for the 1-POI, 4-POI, and 18-POI fit setups.

6.9.4 Inclusive cross-section measurements

The 1-POI fit is used to extract the inclusive $pp \to H \to \tau\tau$ cross-section using 140 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV. The best-fit value of the signal cross-section over the SM prediction is $0.99^{+0.13}_{-0.11} = 0.99 \pm 0.07(\text{stat.})^{+0.10}_{-0.09}(\text{syst.})$. This signal strength corresponds to an inclusive cross-section of

$$\sigma(pp \to H \to \tau\tau) = 3.14 \pm 0.22(\text{stat.})^{+0.32}_{-0.29}(\text{syst.})\ \text{pb} = 3.14^{+0.41}_{-0.35}\ \text{pb}, \qquad (6.13)$$

compatible within uncertainties with the SM prediction of 3.17 ± 0.09 pb [24]. The p-value for the compatibility of the result with the SM hypothesis is 0.97, indicating a very good agreement with the SM prediction. The inclusive cross-section measurement is dominated by systematic uncertainties.

In order to evaluate the impact of the different systematic uncertainty sources, groups of nuisance parameters associated with the same systematic source are defined. Firstly, the nominal fit is performed in order to derive the uncertainties on the POI, $\Delta\mu$. Then, the fit is repeated, fixing the nuisance parameters associated with a given systematic source to their best-fit values. The resulting uncertainty, $\Delta\mu'$, is usually smaller[2] than $\Delta\mu$, as the fit is performed taking into account fewer systematic uncertainties. However, the difference the two uncertainties is an indication of the impact of the systematic source on the POI. The grouped impact is defined as the quadratic difference between $\Delta\mu$ and $\Delta\mu'$, $\sqrt{(\Delta\mu)^2 - (\Delta\mu')^2}$.

The grouped impact of the systematic uncertainties in the 1-POI fit is shown in Table 6.10. The source of systematic uncertainties having the highest impact are the theoretical uncertainties in the signal prediction, followed by the statistical uncertainties in the background sample size and the uncertainties in the jet energy scale and resolution and in the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

The results from the 1-POI and 4-POI fits are shown in Figure 6.39 in terms of the measured signal strengths. The 4-POI fit is used to extract the signal strengths for the ggF, VBF, VH, and $t\bar{t}H$ production modes in a simultaneous fit. The observed cross-section for the ggF production mode is

$$\sigma_{H\to\tau\tau}^{\mathrm{ggF}} = 2.5 \pm 0.4(\mathrm{stat.})^{+0.8}_{-0.6}(\mathrm{syst.})\ \mathrm{pb} = 2.5^{+0.9}_{-0.7}\ \mathrm{pb}, \tag{6.14}$$

which is compatible with the SM prediction of 2.77 ± 0.09 pb. The measurement is dominated by systematic uncertainties, with the largest contributions coming from the theoretical uncertainties in the signal prediction, followed by the uncertainties in the jet energy scale and resolution.

The VBF production mode is measured to have a cross-section of

$$\sigma_{H\to\tau\tau}^{\mathrm{VBF}} = 0.23 \pm 0.03(\mathrm{stat.})^{+0.3}_{-0.2}(\mathrm{syst.})\ \mathrm{pb} = 0.23 \pm 0.04\ \mathrm{pb}, \tag{6.15}$$

which is compatible with the SM prediction of 0.220 ± 0.005 pb. The measurement is dominated by systematic uncertainties, with the largest contributions coming from the theoretical uncertainties in the signal prediction, followed by the uncertainties in the jet energy scale and resolution.

For the VH production mode, the measured cross-section is

$$\sigma_{H\to\tau\tau}^{VH} = 0.11^{+0.07}_{-0.06}(\mathrm{stat.})^{+0.04}_{-0.04}(\mathrm{syst.})\ \mathrm{pb} = 0.11^{+0.08}_{-0.07}\ \mathrm{pb}, \tag{6.16}$$

[2] In some cases, $\Delta\mu'$ can be larger than $\Delta\mu$. This case usually arises when the set of nuisance parameters that are fixed have a very small impact on the POIs, specially when the nuisance parameters are strongly correlated with other sources of systematic uncertainties. The fits can converge to slightly different local minima, leading to unphysical results where the uncertainty on the POI is larger when the nuisance parameters are fixed.

Production mode	Combined	ggF	$t\bar{t}H$	VBF	VH
Best-fit value	0.99	0.92	0.77	1.04	0.98
Total uncertainty	$\pm$0.12	$\pm$0.30	$\pm$0.97	$\pm$0.18	$\pm$0.66
Statistical uncertainty	$\pm$0.07	$\pm$0.15	$\pm$0.82	$\pm$0.13	$\pm$0.55
Total systematic uncertainty	$\pm$0.09	$\pm$0.25	$\pm$0.51	$\pm$0.12	$\pm$0.36
Background samples size	$\pm$0.03	$\pm$0.09	$\pm$0.31	$\pm$0.04	$\pm$0.26
Theoretical uncertainty in signal	$\pm$0.08	$\pm$0.18	$\pm$0.12	$\pm$0.11	$\pm$0.14
Jet and $E_{\mathrm{T}}^{\mathrm{miss}}$	$\pm$0.03	$\pm$0.12	$\pm$0.15	$\pm$0.03	$\pm$0.12
Hadronic τ-lepton decays	$\pm$0.02	$\pm$0.04	$\pm$0.09	$\pm$0.01	$\pm$0.04
Misidentified background	$\pm$0.02	$\pm$0.04	$\pm$0.05	$\pm$0.02	$\pm$0.11
Luminosity	–	$\pm$0.01	$\pm$0.01	$\pm$0.01	$\pm$0.01
Theoretical uncertainty in top-quark processes	–	$\pm$0.01	$\pm$0.31	–	$\pm$0.02
Theoretical uncertainty in Z + jets processes	$\pm$0.01	$\pm$0.02	$\pm$0.08	–	$\pm$0.02
Flavour tagging	$\pm$0.01	$\pm$0.02	$\pm$0.05	$\pm$0.01	$\pm$0.01
Electrons and muons	$\pm$0.01	$\pm$0.02	$\pm$0.02	$\pm$0.01	$\pm$0.02

Table 6.10: Summary of the different sources of uncertainty in the ratio of $\sigma_H \times B(H \to \tau\tau)$ to the SM expectation for each production mode and for the combined cross-section measurement. Experimental uncertainties for reconstructed objects combine efficiency and energy/momentum scale and resolution uncertainties. The background samples size uncertainties include the bin-by-bin statistical uncertainties in the simulated backgrounds as well as statistical uncertainties in misidentified backgrounds, which are estimated using data. Entries with no significant impact on the measurement are denoted by '–'.

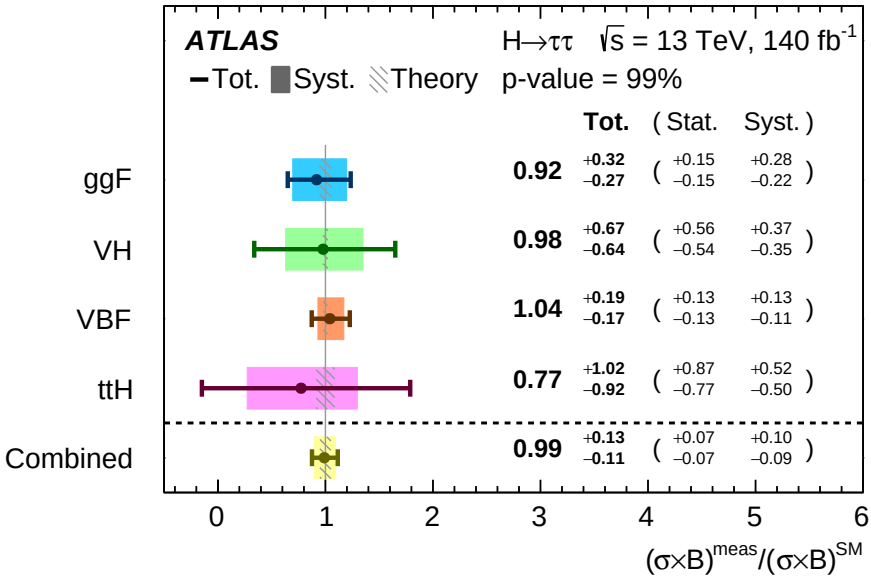


Figure 6.39: The measured values of $\sigma_H \times B(H \to \tau\tau)$ relative to the SM expectations in the 1-POI (labelled as 'Combined') and per-production-mode (4-POI) fit. The total uncertainty in each measurement is indicated by the error bar, with the contribution from the systematic uncertainty indicated by the coloured band.

compatible with the SM prediction of 0.117 ± 0.003 pb. The measurement is strongly dominated by data statistical uncertainties, followed by the uncertainties on the background sample size.

Finally, the $t\bar{t}H$ production mode is measured to have a cross-section of

$$\sigma^{t\bar{t}H}_{H\to\tau\tau} = 0.02^{+0.03}_{-0.02}(\text{stat.}) \pm 0.02(\text{syst.})\ \text{pb} = 0.02^{+0.03}_{-0.02}\ \text{pb}, \tag{6.17}$$

which is also in good agreement with the SM prediction of 0.031 ± 0.003 pb. This measurement is strongly dominated by data statistical uncertainties, and therefore can be improved with more data. The systematic uncertainties are dominated by the statistical uncertainties due to the background sample size, theoretical uncertainties in the top-quark background prediction and the signal prediction, and the uncertainties in the jet energy scale and resolution.

The breakdown of the systematic uncertainties in the signal strength measurements for the 4-POI fit is shown in Table 6.10. The p-value for the compatibility of the result with the SM hypothesis is 0.99, showing a very good compatibility with the SM prediction.

6.9.5 STXS measurements

The 18-POI fit is used to extract the STXS measurements for the $H \to \tau\tau$ analysis in the 18 STXS bins targeted by the analysis. The measured signal strengths for the 18 STXS bins are shown in Figure 6.40. The measured differential cross-sections

are shown in Figure 6.41. The correlation between the 18 POIs is shown in Figure 6.42.

The p-value for the compatibility of the result with the SM hypothesis is 0.06, indicating a reasonable agreement with the SM prediction. Three out of four VBF bins at $p_{\mathrm{T}}^{H} < 200$ GeV show a signal strength below the SM prediction, while the ggF contribution with ≥ 2 jets, $p_{\mathrm{T}}^{H} < 200$ GeV and $m_{\mathrm{jj}} \geq 350$ GeV (VBF-like ggF) shows a signal strength above the SM prediction. This POI is strongly anti-correlated with the VBF bins at $p_{\mathrm{T}}^{H} < 200$ GeV, since the VBF-like ggF contributions contaminates predominantly the VBF_0 categories. The signals are indistinguishable through the $m_{\tau\tau}$ distribution, used to extract the signal strength, leading to the big anti-correlation of up to 70% for the lowest m_{jj} bin. Further studies demonstrated that if the VBF-like ggF contribution is fixed to the SM prediction, the p-value for the compatibility of the result with the SM hypothesis increases to 0.12. The same p-value is obtained if the VBF-like ggF contribution is merged together with the VBF contributions in the categories where the VBF-like ggF contribution is significant (i.e. if the ratio between the VBF-like ggF and VBF contributions is fixed to the SM prediction, but the total cross-section is determined by the fit in m_{jj} bins). These results are shown in Appendix B.

The breakdown of the systematic uncertainties in the signal strength determination for the 18-POI fit is shown in Tables 6.11, 6.12, and 6.13. The results are strongly limitated by statistical uncertainties in the case of the VBF, VH (referred together as qqH in the STXS framework), and $t\bar{t}H$ production modes. The situation is different for the ggF production mode, where the systematic uncertainties dominate the measurement. The largest contributions come from the theoretical uncertainties in the signal prediction and the statistical uncertainties due to the background sample size. The contributions from $E_{\mathrm{T}}^{\mathrm{miss}}$, jet energy scale and resolution and hadronic τ-lepton decays (dominated by the τ energy scale and the trigger and reconstruction efficiencies) are also significant.

Another method to evaluate the impact of the systematic uncertainties is the so-called ranking plot. The impact $\Delta\mu$ of each NP, θ, is defined as the difference between the best-fit value of the POI in the nominal fit and the best-fit value of the POI in the fit where the NP is fixed to $\hat{\theta} \pm x$, being $\hat{\theta}$ the best-fit value of the NP and x the uncertainty on the NP. For each of the NPs, the fit is performed with $x = \pm 1$, resulting in the so-called pre-fit impact. Then, the fit is repeated taking into account the possible constraints in the NP uncertainty, with $x = \Delta\hat{\theta}$, resulting in the post-fit impact, which can be smaller than the pre-fit impact. The ranking plot shows the pre-fit and post-fit impact of the NPs with the largest impact on each of the POIs.

The ranking for the four POIs that scale the $p_{\mathrm{T}}^{H} < 200$ GeV VBF contributions are shown in Figure 6.43. The systematic uncertainties with the largest impact on the signal strengths are theoretical uncertainties derived through variations of the QCD scales. Specifically relevant is the QCD scale uncertainty in the migration of ggF events above and below $p_{\mathrm{T}}(\mathrm{Hjj}) = 25$ GeV. This NP is highly correlated with

Production mode	ggF	ggF	ggF	ggF	ggF	ggF
N_{jets}	1	≥ 2	≥ 1	≥ 1	≥ 1	≥ 2
p_{T}^{H} (GeV)	120–200	120–200	200–300	>300	60–120	0–200
m_{jj} (GeV)	-	0–350	-	-	<300	≥350
Best-fit value	0.35	0.53	0.99	1.51	0.50	5.09
Total uncertainty	±0.61	±0.75	±0.37	±0.55	±0.89	±2.79
Statistical uncertainty	±0.38	±0.49	±0.28	±0.44	±0.52	±1.65
Total systematic uncertainty	±0.49	±0.57	±0.25	±0.28	±0.72	±2.24
Samples size	±0.26	±0.29	±0.12	±0.14	±0.44	±1.04
Theoretical uncertainty in signal	±0.06	±0.28	±0.12	±0.25	±0.15	±1.77
Jet and $E_{\text{T}}^{\text{miss}}$	±0.26	±0.28	±0.15	±0.11	±0.37	±0.90
Hadronic τ-lepton decays	±0.09	±0.08	±0.04	±0.07	±0.13	±0.43
Misidentified background	±0.08	±0.11	±0.05	±0.04	±0.22	±0.69
Luminosity	±0.02	±0.01	±0.02	±0.03	±0.07	±0.16
Theoretical uncertainty in top-quark processes	±0.01	±0.03	±0.02	±0.03	±0.08	±0.13
Theoretical uncertainty in Z + jets processes	±0.10	±0.09	±0.02	±0.02	±0.33	±0.17
Flavour tagging	±0.01	±0.01	±0.01	±0.02	±0.04	±0.07
Electrons and muons	±0.03	±0.06	±0.01	±0.01	±0.05	±0.26

Table 6.11: Summary of the different sources of uncertainty in the ratio of $\sigma_H \times B(H \to \tau\tau)$ to the SM expectation for each ggF STXS bin. Experimental uncertainties for reconstructed objects combine efficiency and energy/momentum scale and resolution uncertainties. Samples size includes the bin-by-bin statistical uncertainties in the simulated backgrounds as well as statistical uncertainties in misidentified τ backgrounds, which are estimated using data. Entries with no significant impact on the measurement are denoted by '–'.

Production mode	qqH	qqH	qqH	qqH	qqH	qqH	qqH	qqH	qqH
p_{T}^{H} (GeV)	-	0-200	0–200	0–200	0–200	≥200	≥200	≥200	≥200
m_{jj} (GeV)	60–120	350–700	700–1000	1000–1500	≥1500	350–700	700–1000	1000–1500	≥1500
Best-fit value	0.94	-0.96	-0.24	1.68	0.12	-1.16	0.98	1.40	1.29
Total uncertainty	±0.66	±1.24	±0.85	±0.58	±0.34	±0.84	±0.68	±0.53	±0.37
Statistical uncertainty	±0.56	±0.82	±0.62	±0.49	±0.29	±0.65	±0.63	±0.50	±0.34
Total systematic uncertainty	±0.37	±0.95	±0.58	±0.32	±0.18	±0.52	±0.25	±0.19	±0.15
Samples size	±0.27	±0.55	±0.31	±0.15	±0.09	±0.46	±0.16	±0.10	±0.05
Theoretical uncertainty in signal	±0.14	±0.75	±0.43	±0.25	±0.12	±0.17	±0.18	±0.16	±0.15
Jet and $E_{\mathrm{T}}^{\mathrm{miss}}$	±0.12	±0.38	±0.21	±0.14	±0.04	±0.13	±0.07	±0.04	±0.03
Hadronic τ-lepton decays	±0.04	±0.14	±0.09	±0.04	±0.02	±0.03	±0.02	±0.01	±0.02
Misidentified background	±0.12	±0.22	±0.14	±0.08	±0.05	±0.09	±0.02	±0.02	±0.02
Luminosity	±0.02	±0.02	±0.02	±0.03	±0.01	±0.02	±0.02	±0.02	±0.02
Theoretical uncertainty in top-quark processes	±0.03	±0.08	±0.03	±0.01	±0.02	±0.02	±0.05	±0.01	±0.01
Theoretical uncertainty in Z + jets processes	±0.02	±0.07	±0.04	±0.02	±0.01	±0.01	±0.01	±0.01	±0.01
Flavour tagging	±0.01	±0.03	±0.02	±0.02	±0.01	±0.03	±0.02	±0.01	±0.01
Electrons and muons	±0.02	±0.13	±0.07	±0.04	±0.02	±0.01	±0.02	±0.01	±0.01

Table 6.12: Summary of the different sources of uncertainty in the ratio of $\sigma_H \times B(H \to \tau\tau)$ to the SM expectation for each VBF and VH STXS bin. Experimental uncertainties for reconstructed objects combine efficiency and energy/momentum scale and resolution uncertainties. Samples size includes the bin-by-bin statistical uncertainties in the simulated backgrounds as well as statistical uncertainties in misidentified τ backgrounds, which are estimated using data. Entries with no significant impact on the measurement are denoted by '–'.

Production mode	$t\bar{t}H$	$t\bar{t}H$	$t\bar{t}H$
p_{T}^{H} (GeV)	0–200	200–300	$\geq$300
Best-fit value	2.1	-2.2	3.6
Total uncertainty	±1.7	±1.2	±2.6
Statistical uncertainty	±1.5	±0.9	±2.4
Total systematic uncertainty	±0.8	±0.7	±1.0
Samples size	±0.5	±0.5	±0.5
Theoretical uncertainty in signal	±0.3	±0.2	±0.5
Jet and $E_{\mathrm{T}}^{\mathrm{miss}}$	±0.5	±0.2	±0.2
Hadronic τ-lepton decays	±0.1	±0.1	±0.2
Misidentified background	±0.1	±0.1	±0.1
Luminosity	–	–	±0.1
Theoretical uncertainty in top-quark processes	±0.4	±0.4	±0.6
Theoretical uncertainty in Z + jets processes	±0.1	±0.1	±0.2
Flavour tagging	±0.1	–	±0.1
Electrons and muons	–	–	–

Table 6.13: Summary of the different sources of uncertainty in the ratio of $\sigma_H \times B(H \to \tau\tau)$ to the SM expectation for each $t\bar{t}H$ STXS bin. Experimental uncertainties for reconstructed objects combine efficiency and energy/momentum scale and resolution uncertainties. Samples size includes the bin-by-bin statistical uncertainties in the simulated backgrounds as well as statistical uncertainties in misidentified τ backgrounds, which are estimated using data. Entries with no significant impact on the measurement are denoted by '–'.

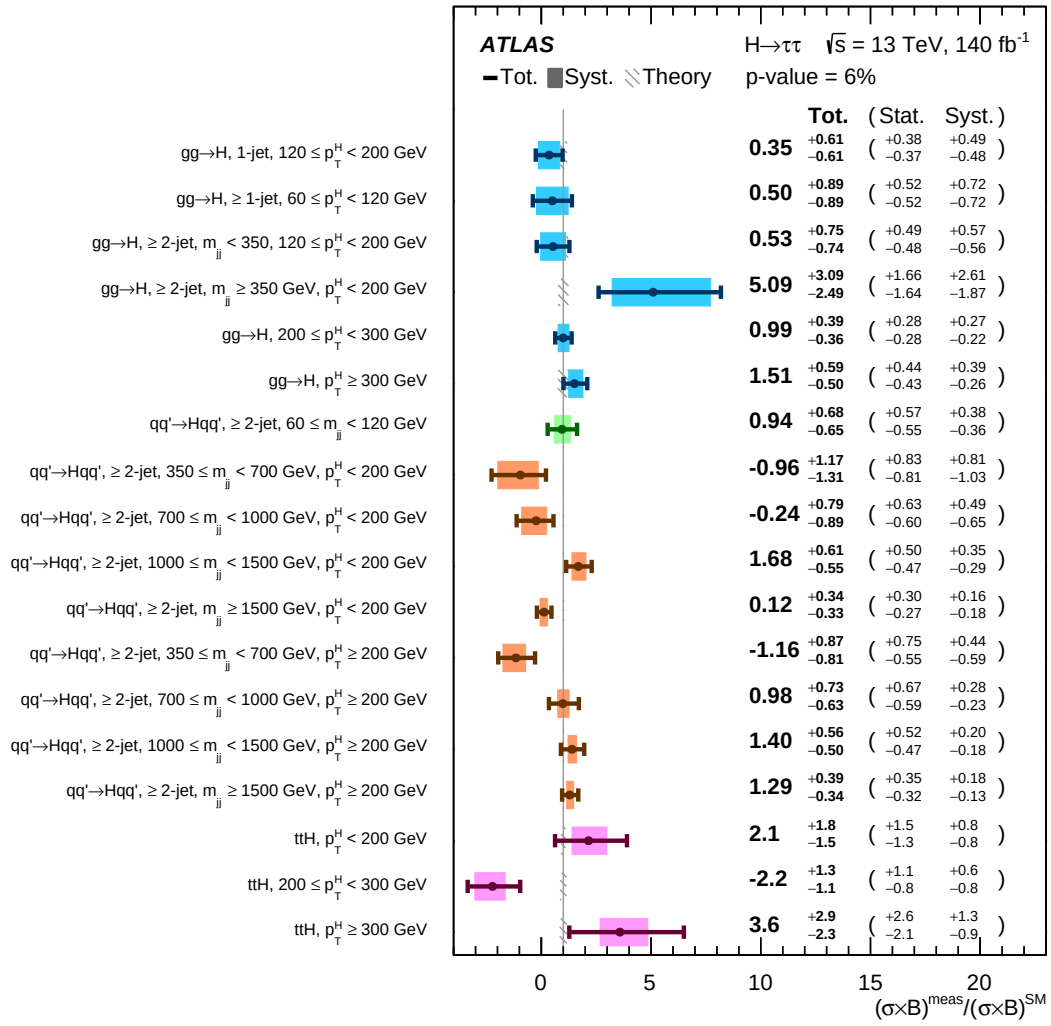


Figure 6.40: The measured values for $\sigma_H \times B(H \to \tau\tau)$ relative to the SM expectations in the simplified template cross-section measurement. The error bars and shaded areas show the total uncertainty and systematic uncertainty in the measurements, respectively.

the VBF-like ggF POI, as pulls in the NP change the shape of the VBF-like ggF template in the VBF_0 categories. A change in the VBF-like ggF POI best-fit value highly affects the VBF signal strengths at $p_{\mathrm{T}}^{H} < 200$ GeV, due to the aforementioned anti-correlation between these signals. Other systematic uncertainties with a large impact on the signal strengths are the ad-hoc Stewart-Tackmann uncertainties, the uncertainty due to the parton showering model, the JER uncertainties and the uncertainties on the $Z(\to \tau\tau) + \text{jets}$ normalization factors.

The ranking plots for the VBF POIs with $p_{\mathrm{T}}^{H} \geq 200$ GeV are shown in Figure 6.44. The systematic uncertainty related to the parton shower has a very large impact, specially in the highest m_{jj} bins. This behavior is not new, and was observed before in Ref. [222] when evaluating the impact of the systematic uncertainties in the VBF inclusive cross-section. The VBF tagger BDT, used to

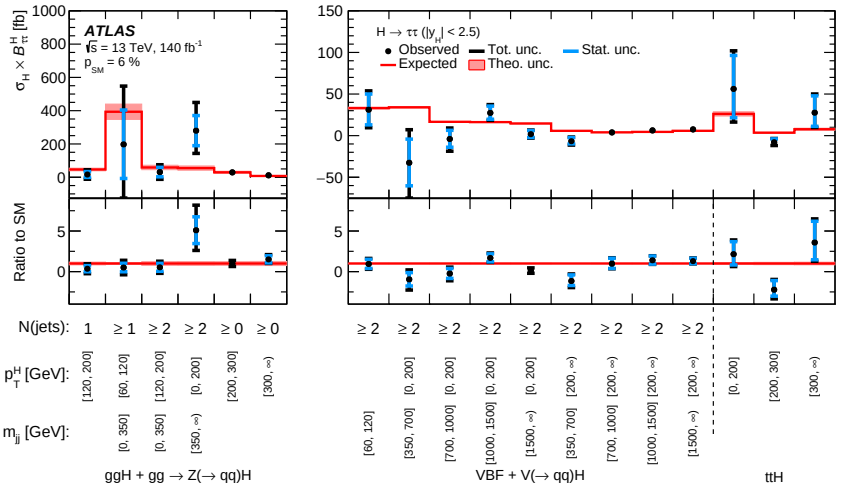


Figure 6.41: The measured values for $\sigma_H \times B(H \to \tau\tau)$ in the simplified template cross-section measurement. The total uncertainty in the measurement is indicated by the black error bars, with the individual contribution from the statistical uncertainty in blue.

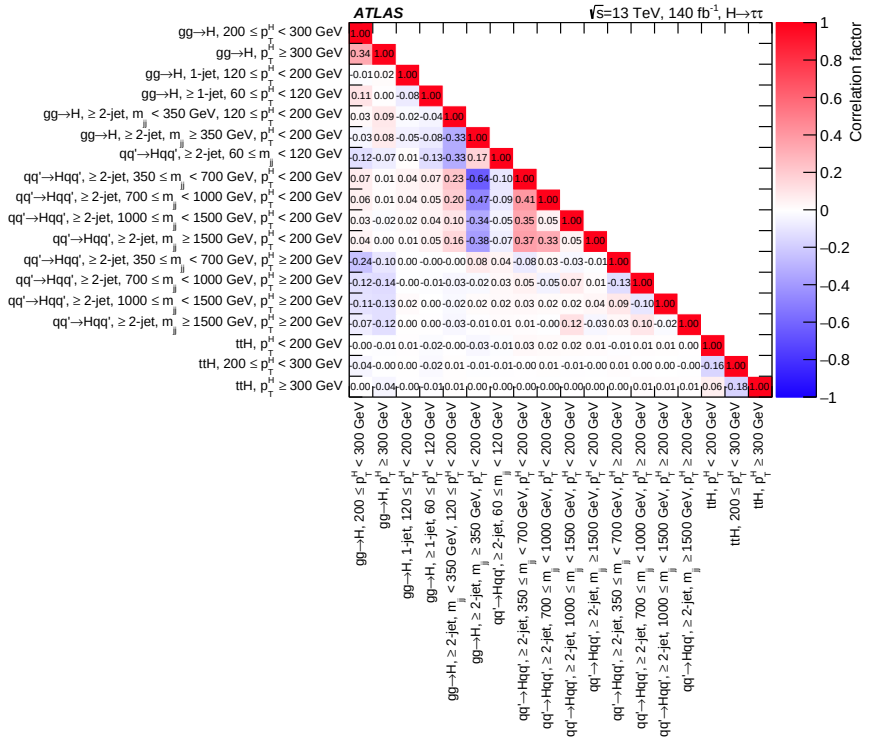


Figure 6.42: Measured linear correlation factors between each pair of parameters of interest in the simplified template cross-section measurement.

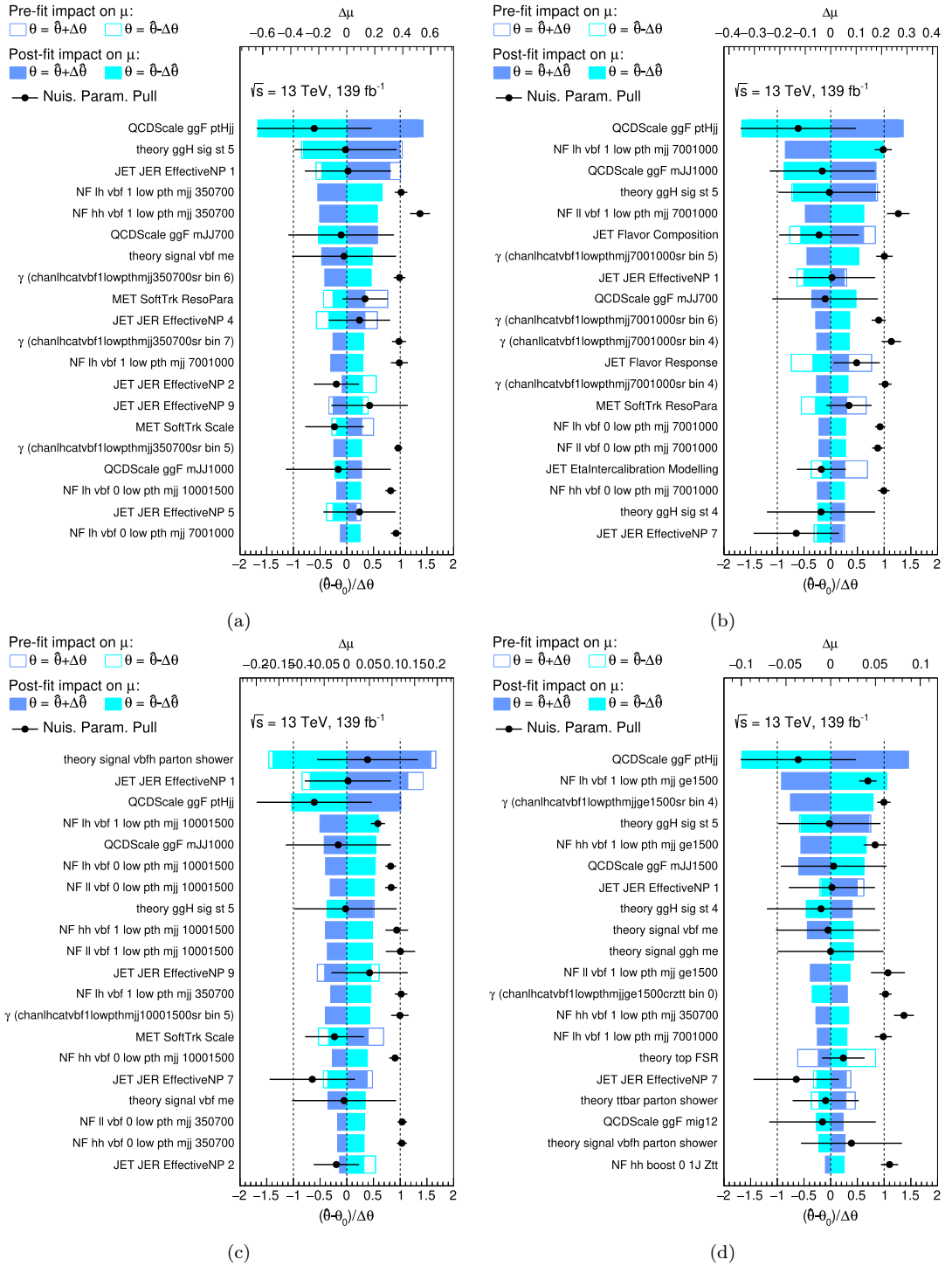


Figure 6.43: Ranking plots for the VBF POIs scaling the signal contributions at $p_{\mathrm{T}}^{H} < 200$ GeV and (a) 350 GeV $\leq m_{\mathrm{jj}} <$ 700 GeV, (b) 700 GeV $\leq m_{\mathrm{jj}} <$ 1000 GeV, (c) 1000 GeV $\leq m_{\mathrm{jj}} <$ 1500 GeV, and (d) $m_{\mathrm{jj}} \geq$ 1500 GeV. The 20 NPs and background NFs with the largest impact on the signal strengths are shown. The empty (filled) rectangles represent the pre-fit (post-fit) impact of the parameters on the signal strengths. Each impact is calculated as the difference between the best-fit value of the signal strength in the nominal fit and the best-fit value of the signal strength in the fit where the NP is shifted by its pre-fit (post-fit) uncertainty. The black points and error bars represent the post-fit values and uncertainties of the NPs. The markers of the background NFs show the pull of the NFs with respect to the nominal value of 1.

define the VBF_0 and VBF_1 categories, can mimic the effect of a third-jet veto, used to suppress the background contributions. Variations in the parton shower model can affect sizeably the VBF tagger output, leading to a large impact on the cross-section measurements. As in the previous case, the QCD scale uncertainties, the ad-hoc Stewart-Tackmann uncertainties and the JER uncertainties have a sizeable impact on all the signal strengths.

Figure 6.45 shows the ranking plots for the $t\bar{t}H$ POIs. The systematic uncertainties with the largest impact on the signal strengths are related to the theoretical uncertainties in the top-quark background prediction, including the uncertainties on the parton shower model, the matrix element calculation and the modeling of the initial and final state radiation. The uncertainty on the top-quark normalization factor is also large, especially in the lowest p_{T}^{H} bin. The uncertainty on the matrix element calculation on the $t\bar{t}H$ signal prediction is also significant in the three p_{T}^{H} bins. Other systematic uncertainties contributing significantly to the total uncertainty are the ones related to the jet energy scale and resolution and the τ-lepton reconstruction and identification.

6.10 Unfolded fiducial differential cross-section measurement results

The measured fiducial differential cross-sections as functions of key kinematic variables are presented in Figure 6.46. The precision of these measurements, evaluated on a per-bin basis, generally falls within the range of 25% to 50%. However, it is important to note that in the lowest p_{T}^{H} bin, the uncertainty escalates significantly, reaching as high as 300%. Across the majority of bins, the statistical uncertainties dominate the total experimental uncertainty, particularly at higher values of p_{T}^{H} and $p_{\mathrm{T}}^{j_0}$.

In terms of systematic uncertainties, the leading contributions vary depending on the specific bin being analyzed. The primary sources of systematic uncertainty include the limited statistical precision of the background model, as well as experimental uncertainties associated with jet and τ-lepton reconstruction. Additionally, uncertainties related to the estimation of the fake-lepton background also contribute significantly to the overall uncertainty budget.

To interpret the results, the data are compared against SM predictions generated using the Powheg + Pythia8 [163, 167, 168], Powheg + Herwig7 [164] and MadGraph5_aMC@NLO + Pythia8 [166] event generators. These simulations account for the Higgs boson production mechanisms: ggF, VBF, VH and $t\bar{t}H$. The theoretical cross-sections for each production mode are combined according to their SM branching ratios. It is worth noting that the theoretical uncertainties associated with these predictions are substantially smaller than the experimental uncertainties observed in the measurements.

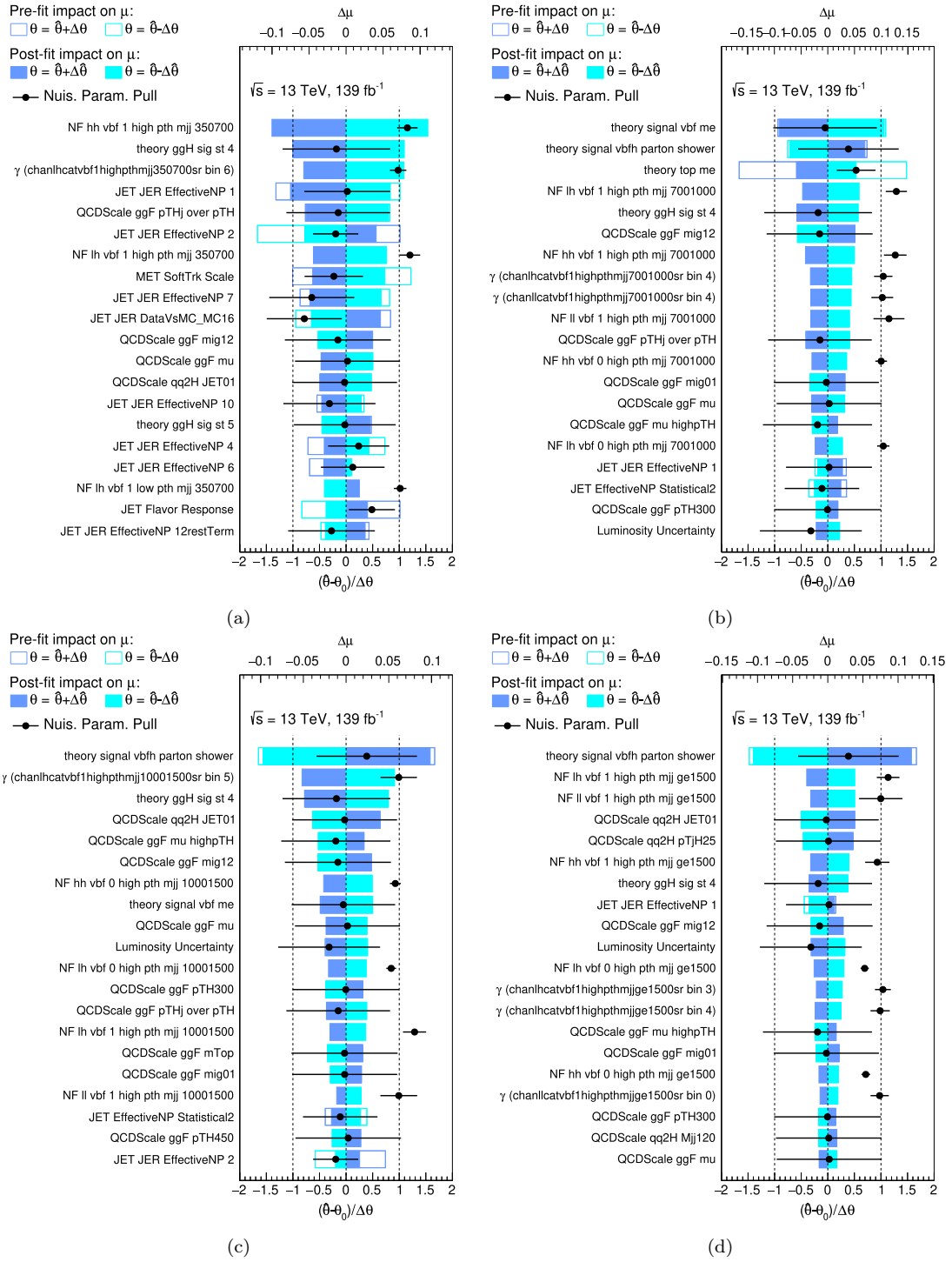


Figure 6.44: Ranking plots for the VBF POIs scaling the signal contributions at $p_{\mathrm{T}}^{H} \geq 200$ GeV and (a) 350 GeV $\leq m_{\mathrm{jj}} <$ 700 GeV, (b) 700 GeV $\leq m_{\mathrm{jj}} <$ 1000 GeV, (c) 1000 GeV $\leq m_{\mathrm{jj}} <$ 1500 GeV, and (d) $m_{\mathrm{jj}} \geq$ 1500 GeV. Only the The 20 NPs and background NFs with the largest impact on the signal strengths are shown. The empty (filled) rectangles represent the pre-fit (post-fit) impact of the parameters on the signal strengths. Each impact is calculated as the difference between the best-fit value of the signal strength in the nominal fit and the best-fit value of the signal strength in the fit where the NP is shifted by its pre-fit (post-fit) uncertainty. The black points and error bars represent the post-fit values and uncertainties of the NPs. The markers of the background NFs show the pull of the NFs with respect to the nominal value of 1.

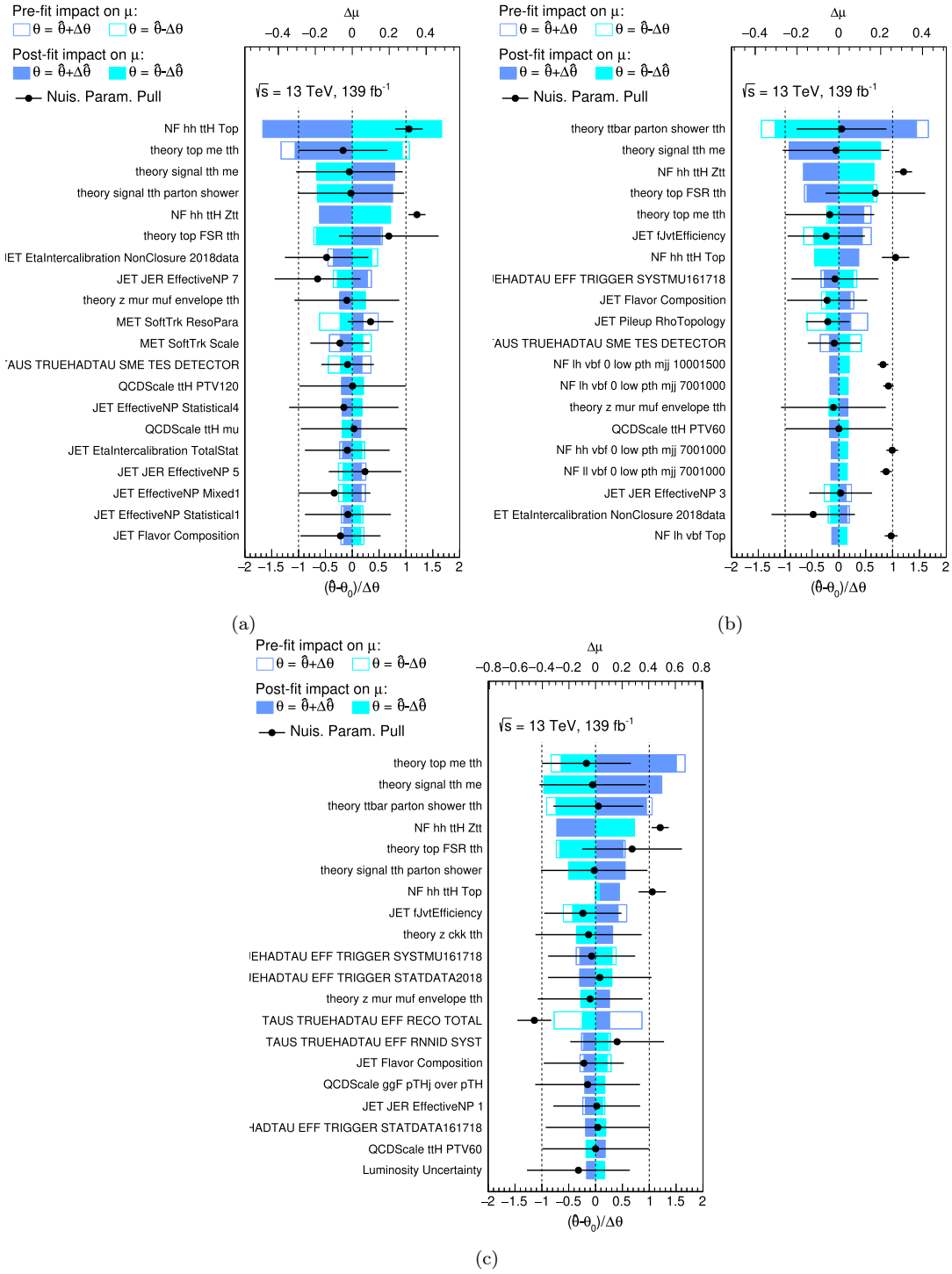


Figure 6.45: Ranking plots for the $t\bar{t}H$ POIs scaling the signal contributions and (a) $p_{\mathrm{T}}^{H} < 200$ GeV, (b) 200 GeV $\leq p_{\mathrm{T}}^{H} < 300$ GeV, and (c) $p_{\mathrm{T}}^{H} \geq 300$ GeV. The 20 NPs and background NFs with the largest impact on the signal strengths are shown. The empty (filled) rectangles represent the pre-fit (post-fit) impact of the parameters on the signal strengths. Each impact is calculated as the difference between the best-fit value of the signal strength in the nominal fit and the best-fit value of the signal strength in the fit where the NP is shifted by its pre-fit (post-fit) uncertainty. The black points and error bars represent the post-fit values and uncertainties of the NPs. The markers of the background NFs show the pull of the NFs with respect to the nominal value of 1.

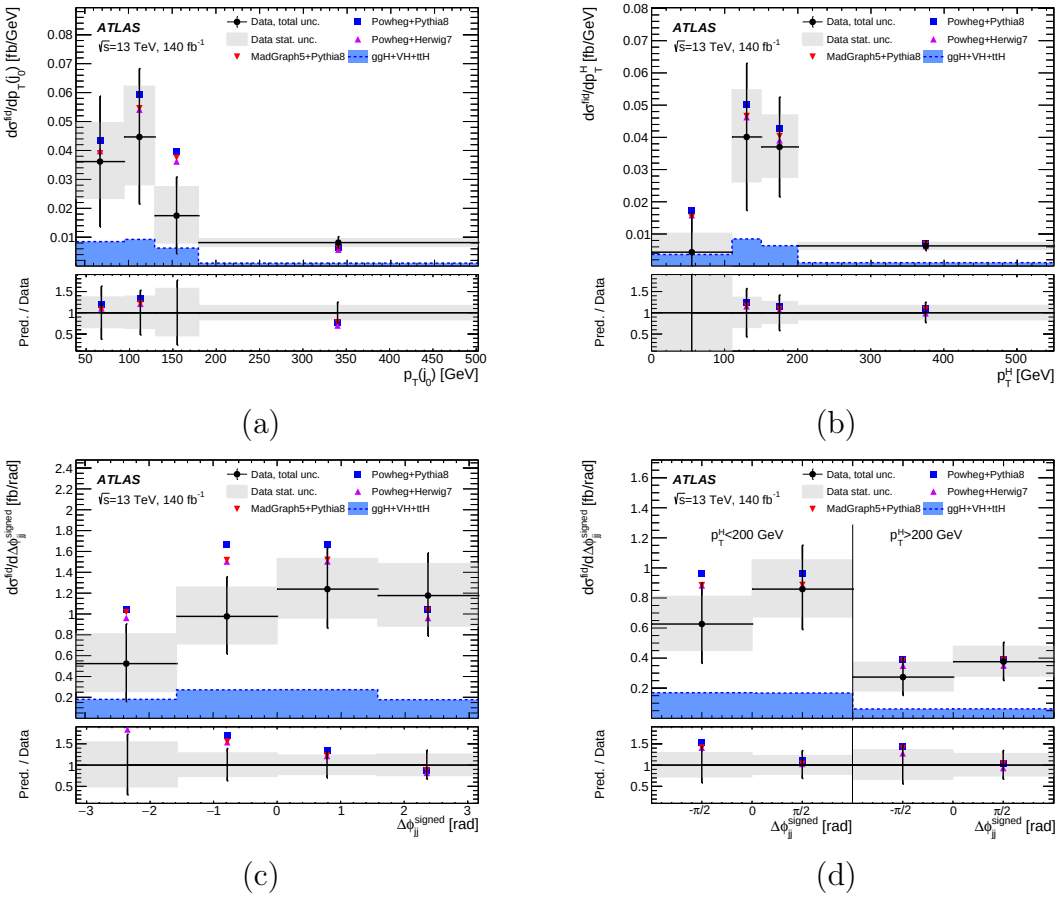


Figure 6.46: Measured fiducial differential cross-sections for (a) $p_{\mathrm{T}}^{j_0}$, (b) p_{T}^{H}, (c) $\Delta\phi_{jj}^{\mathrm{signed}}$, and (d) $\Delta\phi_{jj}^{\mathrm{signed}}$ vs p_{T}^{H}. The measurements are compared with SM predictions from various generators. The shaded box around the data points represents the statistical uncertainty, while the error bars indicate the total uncertainty. The ggF, VH, and $t\bar{t}H$ contributions predicted by Powheg + Pythia8 are also shown. The lower panels show the ratio of the different predictions to the data.

The measured cross-sections are in agreement with the SM predictions from the different generators. In most cases, the predictions from all generators fall within one standard deviation of the measured values. Only in a few specific bins are deviations larger than one standard deviation observed.

Figure 6.46 displays the fiducial differential cross-sections for (a) $p_{\mathrm{T}}^{j_0}$, (b) p_{T}^{H}, (c) the azimuthal angle separation between jets $\Delta\phi_{jj}^{\mathrm{signed}}$, and (d) $\Delta\phi_{jj}^{\mathrm{signed}}$ as a function of p_{T}^{H}. These measurements are overlaid with particle-level SM predictions obtained using the three aforementioned generators, considering contributions from the VBF, ggF, VH, and $t\bar{t}H$ Higgs boson production modes. Each data point is accompanied by a shaded box representing the statistical uncertainty, while the error bars indicate the total uncertainty, inclusive of both statistical and systematic contributions. For reference, the predictions from Powheg + Pythia8 for the ggF, VH, and $t\bar{t}H$ production modes are also overlaid.

The lower panels of the figure illustrate the ratio of the various predictions to the experimental data. Here, error bars reflect the total uncertainty in the measurements, while shaded bands represent the statistical uncertainty. This visual comparison underscores the strong overall agreement between data and theory, with only minor deviations in select bins.

6.10.1 Interpretation in the Standard Model Effective Field Theory framework

The limitations of the SM have motivated numerous high-energy theories that attempt to address these challenges. However, due to the lack of observed direct evidence of new particles, many of these BSM theories cannot be distinguished experimentally, as they all converge to the SM at low-energy limits, consistent with the decoupling theorem [233]. This theorem establishes that phenomena at a low energy scale E_0 are largely independent of a theory's high-energy dynamics at much higher scales E_1, allowing predictions at E_0 to be made with an EFT that encapsulates the relevant degrees of freedom.

EFTs leverage existing energy hierarchies to simplify theoretical descriptions at a particular scale, omitting components irrelevant to that scale. Two approaches are commonly employed: a top-down approach, which constructs a low-energy effective theory by integrating out high-energy degrees of freedom from a known theory, and a bottom-up approach, where an EFT expansion systematically categorizes potential effects from decoupled high-energy physics without any precondition on its structure. Given that the search for BSM physics at the LHC has not found any direct signals, the approach discussed here adopts the bottom-up method, assuming that any BSM influence would originate from an energy scale Λ above the electroweak scale, where SM fields and symmetries still dominate interactions. However, a top-down approach is also needed to use the bottom-up results to constrain specific BSM models.

The SM Lagrangian $\mathcal{L}_{\text{SM}}$, with mass dimension four, can be systematically expanded by higher-dimensional operators $\mathcal{O}_i^{(d)}$ that preserve SM gauge invariance. This expansion introduces new effective interactions with strengths governed by *Wilson coefficients* $c_i^{(d)}$, associated with each operator, and suppressed by powers of Λ to ensure dimensional consistency. While both Λ and c_i remain experimentally unconstrained, the values of c_i can be extracted from data for a chosen Λ, providing a quantitative measure of potential high-energy dynamics. The modified lagrangian is expressed as:

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_d \sum_i \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathcal{O}_i^{(d)}, \quad \text{for } d > 4. \tag{6.18}$$

The EFT formalism employed here, often referred to as the Standard Model Effective Field Theory, thus accommodates possible BSM effects across a range of scales by systematically including higher-dimensional terms.

In this context, higher-dimension operators become increasingly suppressed as their mass dimension increases, limiting the EFT expansion to the most impactful terms. Dimension-five operators, of which the Weinberg operator is the sole contributor, introduce Majorana neutrino masses and violate lepton number conservation; this operator's associated scale is expected to be well beyond the reach of LHC studies and is thus usually neglected for collider data interpretations. At dimension six, approximately 3000 operators contribute, although constraints from proton decay—which limit baryon and lepton number violation—reduce this number significantly. When $\Delta B = \Delta L = 0$ is imposed, the number of operators decreases to 2499. A further reduction is achieved by assuming a global $U(3)^5$ flavor symmetry, which treats the three generations of quarks and leptons as indistinguishable under an effective operator. Dimension-seven operators reintroduce lepton and baryon number violations and are therefore expected to be negligible at the LHC. Dimension-eight operators, numbering around 45000, are further suppressed by Λ^2 relative to dimension-six terms and are often considered subleading, though they may contribute in specific cases [234]. Once the baryon and lepton number conservation constraints are applied, the number of dimension-six operators is reduced [225].

For practical analysis, a choice of operator basis is required to fully capture the most general description of BSM effects in SMEFT. The Warsaw basis [225] is commonly selected for dimension-six operators due to its established Monte Carlo implementations [235], enabling precise predictions for measurable observables.

From all the independent operators when baryon number conservation is assumed, only a subset directly associated with the HVV vertex, those contributing to Higgs-gauge boson interactions, are considered experimentally accessible within the sensitivity limits of this measurement. These include three CP-conserving operators: $H^\dagger H W^n_{\mu\nu} W^{n\mu\nu}$, $H^\dagger H B_{\mu\nu} B^{\mu\nu}$, and $H^\dagger \tau^n H W^n_{\mu\nu} B^{\mu\nu}$, along with their CP-violating counterparts. The six corresponding Wilson coefficients—c_{HW}, c_{HB}, c_{HWB}, $c_{H\tilde{W}}$, $c_{H\tilde{B}}$, and $c_{H\tilde{W}B}$, represent the effective coupling strengths for each operator and are expressed in terms of c_i/Λ^2, where the new physics scale Λ is assumed to be 1 TeV. The STXS framework is primarily based on CP-even observables. Constraints on CP-even coefficients are generally stronger than those on CP-odd coefficients, for which the sensitivity remains relatively limited. These coefficients, summarized in Table 6.14, quantify potential deviations from the SM in the HVV interactions.

In this study, differential cross-sections are used to capture the sensitivity of unfolded kinematic distributions to BSM effects. For each kinematic bin, the cross-section is parameterized with respect to the Wilson coefficients, incorporating both linear and quadratic terms to describe interference and purely BSM contributions, respectively:

$$\sigma_{\text{SM+EFT}} \propto |\mathcal{M}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{M}_i|^2 = |\mathcal{M}_{\text{SM}}|^2 + 2\sum_i \frac{c_i}{\Lambda^2}\mathcal{R}(\mathcal{M}^*_{\text{SM}}\mathcal{M}_i) + \sum_{i,j}\frac{c_i c_j}{\Lambda^4}\mathcal{R}(\mathcal{M}^*_i\mathcal{M}_j). \tag{6.19}$$

	CP-even		
Operator $\mathcal{O}_i^{(d=6)}$	$H^\dagger H W^n_{\mu\nu} W^{n\mu\nu}$	$H^\dagger H B_{\mu\nu} B^{\mu\nu}$	$H^\dagger \tau^n H W^n_{\mu\nu} B^{\mu\nu}$
Wilson coefficient	c_{HW}	c_{HB}	c_{HWB}
	CP-odd		
Operator $\mathcal{O}_i^{(d=6)}$	$H^\dagger H \tilde{W}^n_{\mu\nu} W^{n\mu\nu}$	$H^\dagger H \tilde{B}_{\mu\nu} B^{\mu\nu}$	$H^\dagger \tau^n H \tilde{W}^n_{\mu\nu} B^{\mu\nu}$
Wilson coefficient	$c_{H\tilde{W}}$	$c_{H\tilde{B}}$	$c_{H\tilde{W}B}$

Table 6.14: The three CP-even and three CP-odd operators and their corresponding Wilson coefficients which directly impact the interactions between the Higgs boson and vector bosons.

For each kinematic bin, the cross-section as a function of the Wilson coefficients c_i can be parameterized as:

$$\sigma^k_{\text{SM+EFT}} = \sigma^k_{\text{SM}} \left(1 + \alpha_{ik} \frac{c_i}{\Lambda^2} + \beta_{ik} \left(\frac{c_i}{\Lambda^2} \right)^2 \right). \tag{6.20}$$

The coefficients α_{ik} and β_{ik} describe the linear and quadratic dependence on c_i in each bin k. These coefficients are determined from simulated samples generated with different values of c_i, using MadGraph5_aMC@NLO 2.9.5 with the SMEFTsim [235, 236] model. The MC simulations is produced assuming $\Lambda = 1$ TeV, employing the NNPDF3nlo PDF set and showered with Pythia 8.245 using the A14 tune. The resulting changes in cross-sections from these BSM samples are then applied to VBF cross-section predictions obtained from Powheg Box v2, showered with Pythia 8.230. The σ_{SM} cross-section corresponds to the SM prediction from Powheg+Pythia 8. For other Higgs boson production modes, the predictions from the relevant samples described in Chapter 4 are used.

The linear term, determined by α_{ik}, arises from the interference between the SM and BSM matrix elements and scales as $1/\Lambda^2$. The quadratic term, governed by β_{ik}, corresponds purely to BSM contributions and scales as $1/\Lambda^4$. Dimension-8 operators, not considered here, could also contribute at this order. Hence, for a valid interpretation with only dimension-6 operators, the quadratic contributions must remain subdominant to the linear ones. To evaluate this, measurements are performed considering both the full form of Eq. (6.20) and a linearized version excluding the quadratic term for each operator in Table 6.14.

Figure 6.47 illustrates the impact of varying the Wilson coefficients c_{HW} and $c_{H\tilde{W}}$ on the distributions of $\Delta\phi^{\text{signed}}_{jj}$, p^H_{T}, and $\Delta\phi^{\text{signed}}_{jj}$ vs p^H_{T}. The CP-odd operator associated with $c_{H\tilde{W}}$ introduces an asymmetric effect on $\Delta\phi^{\text{signed}}_{jj}$, which is further amplified by the p^H_{T} cut in the two-dimensional distribution. On the other hand, the CP-even operator c_{HW} affects both the $\Delta\phi^{\text{signed}}_{jj}$ and p^H_{T} distributions, with a symmetric impact on $\Delta\phi^{\text{signed}}_{jj}$ due to the CP-even nature of the operator.

The results presented here assume only one non-vanishing Wilson coefficient at a time, while the others are set to zero. To extract the optimal value for each of the six Wilson coefficients under consideration, the unfolded data are

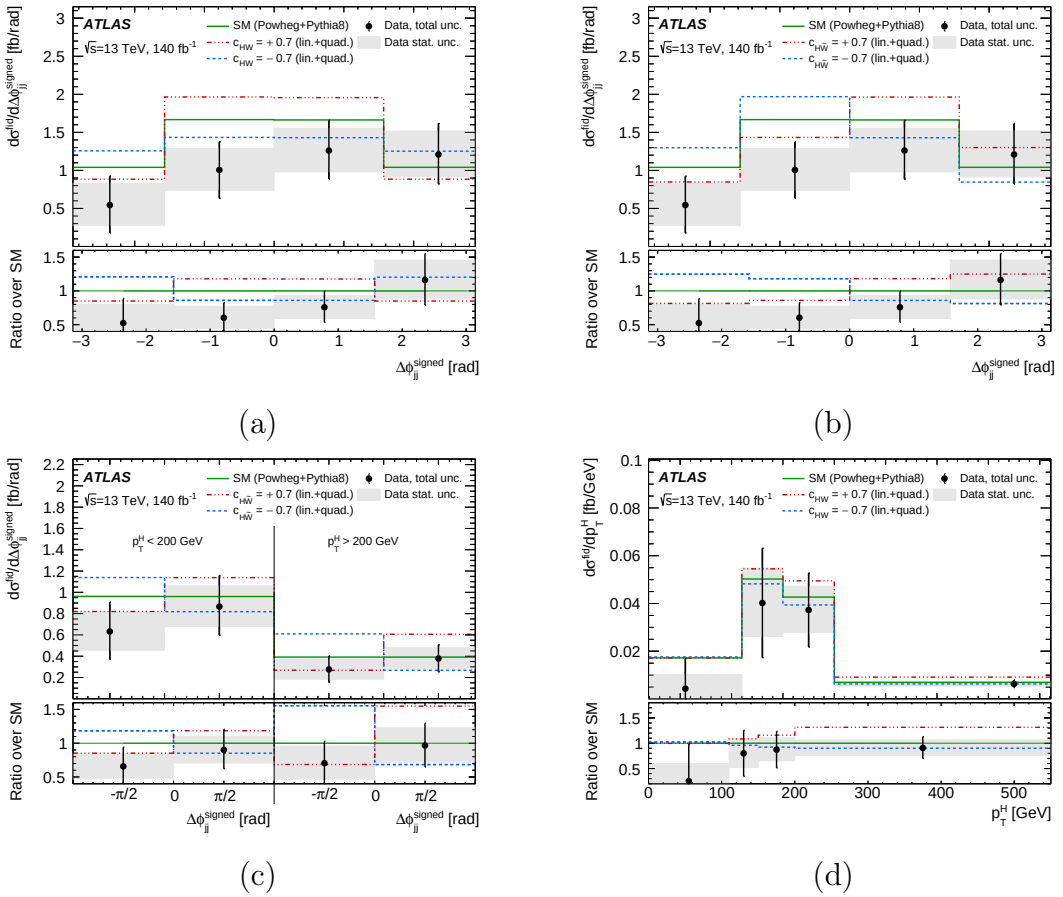
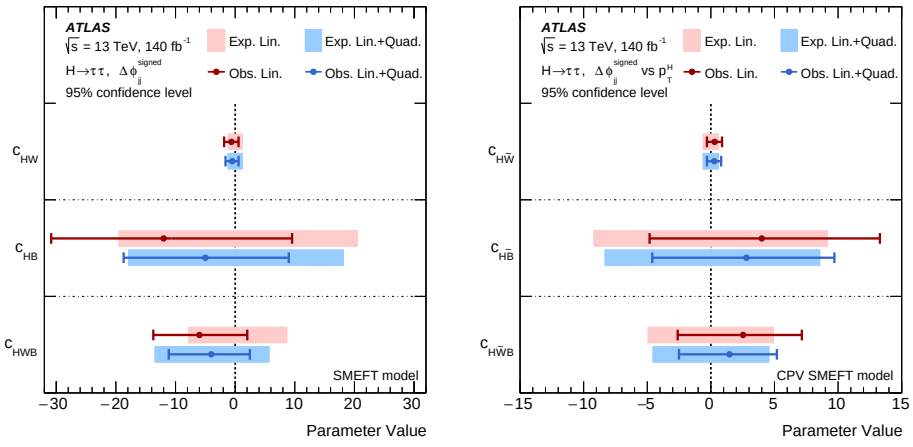


Figure 6.47: The unfolded data compared with the nominal SM predictions from Powheg+Pythia 8, as well as the predictions for several values of Wilson coefficients in the SMEFT, for (a,b) $\Delta\phi_{jj}^{\text{signed}}$, (c) $\Delta\phi_{jj}^{\text{signed}}$ vs p_{T}^{H} and (d) p_{T}^{H}.

combined with the theoretical predictions for the cross-section as a function of the Wilson coefficients. The same profile-likelihood method described in Section 6.8.2 is employed, but only one Wilson coefficient is treated as a free parameter, while the others are set to zero. A scan of the negative log-likelihood is performed for various values of each Wilson coefficient, fixing the predicted cross-section in each bin according to the SMEFT framework.

By examining the deviation of the negative log-likelihood from its minimum, 95% confidence intervals are derived for each Wilson coefficient. This procedure is applied considering both linear terms only, as well as a combination of linear and quadratic terms. For Wilson coefficients associated with CP-even operators, the $\Delta\phi_{jj}^{\text{signed}}$ distribution is used to achieve the best precision, while the $\Delta\phi_{jj}^{\text{signed}}$ vs p_{T}^{H} distribution is utilized for CP-odd operators. These distributions were chosen based on their superior expected sensitivity. The resulting confidence intervals for each of the six Wilson coefficients are displayed in Figure 6.48, comparing both the expected and observed intervals.

The intervals for most coefficients are quite similar when considering only



(a) (b)

Figure 6.48: Observed and expected 95% confidence intervals for the six Wilson coefficients considered in the analysis. Each coefficient is studied individually, with both linear-only and linear-plus-quadratic models used to evaluate sensitivity to $\mathcal{O}(\Lambda^{-4})$ terms. Panel (a) shows results for CP-even operators using the $\Delta\phi_{jj}^{\text{signed}}$ distribution, while panel (b) presents results for CP-odd operators using the $\Delta\phi_{jj}^{\text{signed}}$ vs p_{T}^{H} distribution. All limits are computed at a new physics scale $\Lambda = 1$ TeV.

linear terms or both linear and quadratic terms, except for the c_{HWB} coefficient. This suggests that the SMEFT analysis using only dimension-six operators is largely applicable to the majority of operators under study. For instance, the observed (expected) confidence intervals for c_{HW} are $[-1.85, +0.57]$ ($[-1.17, +1.23]$), and for $c_{H\tilde{W}}$ are $[-0.31, +0.88]$ ($[-0.60, +0.60]$) in the linear scenario, which are the tightest constraints to date, particularly for $c_{H\tilde{W}}$.

Chapter 7

Searches for lepton-flavor-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$

7.1 Analysis motivation and strategy

Measuring with high precision the couplings of the Higgs boson to fermions and gauge bosons is a fundamental test of the consistency of the SM. The exploration of the Higgs sector does not only involve the measurement of the Higgs boson properties, but also the search for new physics beyond the SM linked to the Higgs sector. This chapter presents searches for lepton-flavor-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ final states using 139 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV collected by the ATLAS detector.

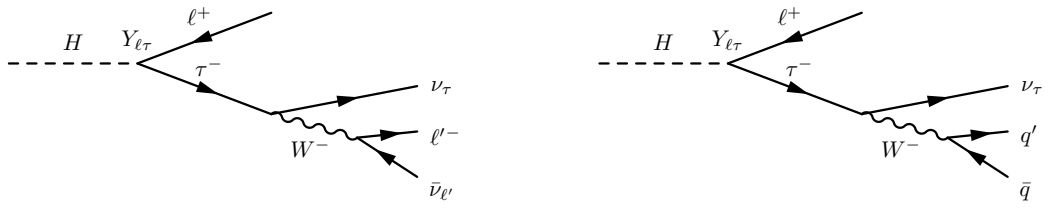
Experimental evidence from neutrino oscillations has already established that lepton flavor is not a strictly conserved symmetry, suggesting that LFV could exist in the charged-lepton sector. An introduction to the theoretical aspects of LFV in the Higgs sector is provided in Section 1.6.

Direct searches are essential for setting experimental limits on LFV in Higgs decays. Low energy experiments looking for processes like $\mu \to e\gamma$, $\tau \to \mu\gamma$, and $\tau \to e\gamma$ have already placed constraints on the branching ratios of possible LFV decays of the Higgs boson. A summary of these constraints is also presented in Section 1.6. However, the indirect constraints on $H \to e\tau$ and $H \to \mu\tau$ are weaker than the direct limits that the LHC experiments can provide.

The direct searches for $H \to e\tau$ and $H \to \mu\tau$ are the best way to probe the off-diagonal Yukawa couplings of the Higgs boson. Previous analyses conducted by ATLAS [84] have already placed constraints on the branching ratios for $H \to e\tau$ and $H \to \mu\tau$ decays, with upper limits of 0.47% and 0.28% at 95% CL, respectively, based on data collected at $\sqrt{s} = 13$ TeV and corresponding to an integrated luminosity of 36.1 fb^{-1}. Similarly, the CMS Collaboration, using 137 fb^{-1} of $\sqrt{s} = 13$ TeV data, has set stricter bounds with $\mathcal{B}(H \to e\tau) < 0.22\%$ and

$\mathcal{B}(H \to \mu\tau) < 0.15\%$ [85]. For the decay $H \to e\mu$, ATLAS reported a branching ratio limit of 6.1×10^{-5} at 95% CL [82], using 139 fb^{-1} of $\sqrt{s} = 13$ TeV data.

One of the main purposes of this thesis is to explore LFV Higgs decays, particularly the $H \to e\tau$ and $H \to \mu\tau$ decay channels, using the full Run 2 dataset collected by the ATLAS experiment. For each decay channel, two primary final states are analyzed: $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$. The former exploits leptonic decays of the τ-lepton, $\tau \to \ell'\nu\bar{\nu}$, while the latter investigates hadronic decays, $\tau \to \text{hadrons} + \nu$. Figure 7.1 illustrates these decay modes, with the off-diagonal Yukawa coupling term, $Y_{\ell\tau}$, mediating LFV interactions.



(a) (b)

Figure 7.1: LFV decay schemes of the Higgs boson for the (a) $\ell\tau_{\ell'}$ and (b) $\ell\tau_{\text{had}}$ final states. The off-diagonal Yukawa coupling term is indicated by the $Y_{\ell\tau}$ symbol.

In the analysis of the $\ell\tau_{\ell'}$ final state, backgrounds from same-flavor lepton pairs, primarily produced by Drell-Yan processes, pose a significant challenge. Consequently, this analysis focuses exclusively on events where the two leptons are of different flavors.

Two distinct strategies are used to estimate the background in the $\ell\tau_{\ell'}$ channel. The first, referred to as the MC-based method, relies on templates generated from simulations, with data used to normalize the main backgrounds, and a data-driven approach employed to estimate the contribution from misidentified leptons. The second strategy, the symmetry-based method, is based on the assumption that the background from prompt leptons (such as those originating from Higgs boson decays or heavy vector bosons) is symmetric with respect to electron-muon exchange. This allows for a direct estimation of the SM background using only data [237], since the prompt lepton background is expected to be symmetric in the SM. For the $\ell\tau_{\text{had}}$ final state, only the MC-based method is applied, where simulation provides the background templates, and a data-driven method estimates misidentified backgrounds.

The two methods serve different purposes. While the MC-based technique aims to measure the branching ratios of $H \to e\tau$ and $H \to \mu\tau$ individually, the symmetry-based approach is directly sensitive to the difference between these branching ratios. Both analysis methods employ MVA techniques to maximize the signal-to-background separation. The symmetry-based strategy can be found in Appendix C.

In total, three statistical approaches are pursued. The first two independently search for the $H \to e\tau$ and $H \to \mu\tau$ processes, assuming that only one decay mode is present in each analysis (i.e. for the $H \to e\tau$ analysis, the branching ratio for $H \to \mu\tau$ is fixed to zero, and vice versa). The third approach performs a simultaneous fit for both $H \to e\tau$ and $H \to \mu\tau$, relaxing the assumption that one process is absent. In the independent analyses, the $\ell\tau_{\text{had}}$ channel, analyzed using the MC-based method, is combined with either the symmetry-based or MC-based method in the $\ell\tau_{\ell'}$ channel, depending on which offers higher sensitivity. The simultaneous fit, on the other hand, uses only the MC-based approach in both final states. A measurement of the branching ratio difference between $H \to e\tau$ and $H \to \mu\tau$ is also performed using the $\ell\tau_{\ell'}$ channel, both with the symmetry-based and MC-based methods. The results of these searches are published in Ref. [238]. Special emphasis is placed on the MC-based $\ell\tau_{\ell'}$ channel, as it was one of the main contributions of this thesis to the analysis.

7.2 Object definition and trigger selection

The topology of the $H \to e\tau$ and $H \to \mu\tau$ events requires the reconstruction of electrons, muons, τ-leptons, jets, and missing transverse momentum. The number of identified electrons, muons and $\tau_{\text{had-vis}}$ is used in order to categorize the events into the two channels, $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$. The channels are further divided into categories, one targeting VBF production of the Higgs boson and the other targeting the rest of production modes.

This section describes the specific requirements and selections applied to the objects in the analysis, and the trigger algorithms used to select the events.

7.2.1 Object definition

The object reconstruction follows the procedures described in Chapter 5. This section provides a brief summary of the object definitions used in the LFV analyses, with a focus on the additional requirements and selections imposed on the objects.

In the analysis, The primary vertex is identified as the vertex with the highest sum of the squared transverse momenta of the associated tracks. Track reconstruction is detailed in Section 5.1.

The electron reconstruction, identification and isolation are described in Section 5.2. Electrons are required to satisfy the *Loose* likelihood-based identification criteria in the LFV analysis, which achieve an efficiency of roughly 93%. Additionally, they are required to have a $p_{\text{T}} > 15$ GeV and $|\eta|$ values below 1.37 or between 1.52 and 2.47, thus avoiding the region between the barrel and endcap calorimeters. Furthermore, electron candidates must fulfill the *Gradient* isolation requirements, where the efficiency reaches 90% at a p_{T} of 25 GeV and increases to 99% at 60 GeV. Electrons must also pass the *Medium* identification criteria, which offer an 88% efficiency for electroweak processes [204].

Details on the muon reconstruction, identification and isolation can be found in Section 5.3. For their identification, muons must meet the *Loose* selection criteria, which results in an efficiency close to 98%. These muons are required to have $p_{\mathrm{T}} > 10$ GeV and a pseudorapidity range of $|\eta| < 2.5$. In addition, *Medium* identification and *Tight* isolation criteria must be satisfied by the selected muons. The isolation procedure relies on either calorimetric and track-based information (in the case of the MC-based method) or on track information alone (in the Symmetry-based method). The Symmetry-based method option is chosen to allow more activity around the track, enhancing the number of selected events for the data-driven background estimation.

The jet reconstruction and calibration, together with the b-tagging algorithm, are described in Section 5.4. Jets considered in the analysis are reconstructed using a radius parameter of $R = 0.4$. The analysis considers only jets with p_{T} above 20 GeV and pseudorapidity values within $|\eta| < 4.5$. To suppress pile-up contributions, jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.5$ are required to pass the JVT algorithm. In the forward region, the forward JVT method is applied to jets with $|\eta| > 2.5$ to mitigate pile-up effects. Jets containing b-hadrons are identified using the DL1r b-tagging algorithm, operated at an 85% efficiency working point.

Leptonic τ decays are reconstructed as electrons or muons. The reconstruction of hadronic τ-lepton decays is more complex, as described in Section 5.5. The candidates are required to satisfy the *Very Loose* identification criteria, providing 95% efficiency for simulated τ_{had} decays. Additionally, to reduce background contributions from misidentified electrons, the eBDT discriminant is employed. This discriminant is also introduced in Section 5.5.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction follows the procedure described in Section 5.6.

7.2.2 Trigger selection

The trigger selection is based on a logical OR combination of single-electron, single-muon, and di-lepton triggers, both for the $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$ channels. The usage of di-lepton triggers in the $\ell\tau_{\ell'}$ channel increases the acceptance of the analysis by approximately 2% compared to a straightforward OR combination of single-lepton triggers.

The specific triggers used in the analysis vary according to the data-taking period. A matching criterion based on ΔR, is imposed between the trigger object and the reconstructed lepton.

The p_{T} thresholds for single-lepton triggers were set to $p_{\mathrm{T}} > 27\,(25)$ GeV for electrons and $p_{\mathrm{T}} > 27.3\,(21)$ GeV for muons during the 2016–2018 (2015) data-taking period. For the di-lepton triggers, the thresholds were $p_{\mathrm{T}} > 18$ GeV for electrons and $p_{\mathrm{T}} > 14.7$ GeV for muons.

7.3 Event selection and categorization

The event selection process in this analysis is divided into two stages. First, events are sorted into one of the two analysis channels, depending on the decay mode of the τ lepton. The second stage involves further classification of events into categories that are tailored to target the two dominant Higgs boson production mechanisms: ggF and VBF.

Two distinct channels are defined based on the τ decay mode: the $\ell\tau_{\text{had}}$ channel, where one light lepton and one $\tau_{\text{had-vis}}$ object are required, and the $\ell\tau_{\ell'}$ channel, which requires two light leptons of different flavor with no accompanying $\tau_{\text{had-vis}}$.

A summary of the event selection and categorization for the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels, both for the MC-based and Symmetry-based methods, is provided in Table 7.1. These selection criteria are described in more detail in the following sections.

7.3.1 Selection in the MC-based $\ell\tau_{\ell'}$ channel

7.3.1.1 Lepton assignment

The lepton with the higher p_{T} is labeled ℓ_1, while the one with lower p_{T} is designated as ℓ_2. In the LFV Higgs boson decay signal, one lepton originates directly from

Selection	$\ell\tau_{\ell'}$	$\ell\tau_{\text{had}}$
Baseline	exactly $1e$ and 1μ, OS	exactly 1ℓ and $1\tau_{\text{had-vis}}$, OS
	$p_{\text{T}}^{\ell_1} > 45$ GeV	$p_{\text{T}}^{\ell} > 27.3$ GeV
	$p_{\text{T}}^{\ell_2} > 15$ GeV	$p_{\text{T}}^{\tau_{\text{had-vis}}} > 25$ GeV, $\lvert\eta^{\tau_{\text{had-vis}}}\rvert < 2.4$
	30 GeV $< m_{\ell\ell} <$ 150 GeV	$\sum_{i=\ell,\tau_{\text{had-vis}}} \cos\Delta\phi(i, E_{\text{T}}^{\text{miss}}) > -0.35$
	$0.2 < p_{\text{T}}^{e}(\text{track})/p_{\text{T}}^{e}(\text{cluster}) < 1.25$ ($\mu\tau_e$ only)	$\lvert\Delta\eta(\ell, \tau_{\text{had-vis}})\rvert < 2$
	For ℓ_{prompt}: track d_0 significance < 5	$\tau_{\text{had-vis}}$ of 1 or 3 tracks
	If $\ell_{\text{prompt}} = \mu$: electron track d_0 significance < 10	$\tau_{\text{had-vis}}$ passes Tight RNN ID
	b-veto (for jets with $p_{\text{T}} > 25$ GeV and $\lvert\eta\rvert < 2.4$)	
VBF	Baseline	
	≥ 2 jets, $p_{\text{T}}^{\text{j1}} > 40$ GeV, $p_{\text{T}}^{\text{j2}} > 30$ GeV	
	$\lvert\Delta\eta_{jj}\rvert > 3$, $m_{\text{jj}} > 400$ GeV	
	–	$p_{\text{T}}^{\tau_{\text{had-vis}}} > 45$ GeV
non-VBF	Events passing the baseline selection but failing VBF categorization	
	–	veto events if
	–	$90 < m_{\text{vis}}(e, \tau) < 100$ GeV

Table 7.1: Event selection and categorization for the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ SRs for both the MC-based and Symmetry-based methods.

the Higgs boson decay (prompt), and the other from a τ-lepton decay. A lepton-assignment algorithm is used to identify which lepton is prompt and which is the τ-lepton decay product.

Typically, the prompt lepton corresponds to the one with higher p_{T} in the detector frame, and the one with lower p_{T} is the result of the τ-lepton decay. This distinction leads to the definition of the $e\mu$ sample, where the electron is the leading lepton, and the μe sample, where the muon is the leading lepton. These classifications are useful for discussing backgrounds, as certain background processes do not have final states with one prompt and one non-prompt lepton.

However, separating the events into an $e\mu$ and μe sample is not optimal for signal extraction. In the laboratory frame, the leading lepton is not always the prompt one in LFV Higgs boson decays, which dilutes the signal and reduces the overall sensitivity. For example, with the p_{T}-ordering lepton assignment, a $H \to e\tau_\mu$ event can end up in the μe sample, and a $H \to \mu\tau_e$ event can be placed in the $e\mu$ sample, introducing signal cross-contamination in the two samples.

The lepton-assignment algorithm is revised in this analysis to improve the accuracy of the method and be able to obtain a better separation classification of $H \to e\tau_\mu$ and $H \to \mu\tau_e$ events in their respective samples (i.e. the $e\tau_\mu$ and $\mu\tau_e$ samples, respectively). To do so, the leptons are boosted into an approximate Higgs boson rest frame. The reconstruction of the Higgs boson rest frame is not trivial, since the neutrino-momentum information in the z-direction is missing.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ 4-vector is constructed by assuming that the η of the neutrinos is the same as that of the dilepton system. Although this assumption is not always valid, it enhances the accuracy of lepton assignment in events where the Higgs boson is highly boosted. Subsequently, the 4-momentum of the Higgs boson is computed by combining the two reconstructed leptons with the $E_{\mathrm{T}}^{\mathrm{miss}}$ 4-momentum. The invariant mass of the system formed by the two leptons and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is then constrained to the known Higgs mass of 125 GeV. The leptons are boosted into the approximate Higgs boson rest frame, and their 4-momenta are recalculated. The prompt lepton is identified as the one with the highest p_{T} in this frame, and is referred to as ℓ_H, indicating that it is assumed to be the lepton originating from the Higgs boson decay. The other lepton is labeled as ℓ_τ, indicating that it is assumed to be the lepton from the τ-lepton decay.

The samples derived from this lepton assignment are referred to as the $e\tau_\mu$ and $\mu\tau_e$ samples, in contrast to the $e\mu$ and μe samples, defined only by the p_{T} ordering in the laboratory frame. This new classification improves the separation of signal and background processes by more accurately identifying the prompt lepton.

The accuracy of this lepton assignment method has been studied for the signal samples using MC simulation. The results demonstrate a significant improvement in correctly assigning leptons compared to the simple laboratory-frame p_{T} ordering. For the ggF production mode, the accuracy of the laboratory p_{T} ordering is found to be 89.3%, while for the VBF mode, the ordering achieves an accuracy of 81.0%. The approximate Higgs center-of-mass frame lepton assignment yields an accuracy

of 94.9% for ggF events and 93.3% for VBF events. This is the method used in the analysis to perform the lepton assignment. The improvement with respect to the laboratory-frame lepton ordering is significant, especially for the VBF production mode.

7.3.1.2 Preselection

Events that satisfy the baseline selection criteria must include exactly one electron and one oppositely charged muon. No additional leptons are allowed, including $\tau_{\text{had-vis}}$ candidates. The number of vertices in the event is required to be at least one.

The leading lepton is required to have a transverse momentum of $p_{\text{T}}^{\ell_1} > 45$ GeV. To suppress background from $t\bar{t}$ production, events containing b-jets are excluded, and the dilepton invariant mass is constrained to lie between 30 GeV and 150 GeV. The term *baseline* is used to highlight the relatively loose nature of this selection.

To reduce the contamination from events with non-prompt leptons, such as events with misidentified leptons and, to a lesser extent, leptons from $Z \to \tau\tau$ decays, requirements on vertex compatibility are imposed. The significance of the transverse impact parameter, d_0/σ_{d_0}, is required to be less than 5 for the leading lepton (ℓ_1). Both leptons are required to have $|z_0| \sin\theta < 0.5$ mm, where z_0 is the longitudinal impact parameter and θ is the polar angle in the laboratory frame using the ATLAS coordinate system. In the case where the subleading lepton is an electron ($\ell_2 = e$), its d_0 significance must be less than 10 and the ratio between the lepton p_{T} measured in the inner detector ($p_{\text{T}}^{\text{track}}$) and the p_{T} measured in the calorimeter ($p_{\text{T}}^{\text{cluster}}$) must be $0.2 < p_{\text{T}}^{\text{track}}(\ell_2)/p_{\text{T}}^{\text{cluster}}(\ell_2) < 1.25$. This last criterion is designed to suppress the $Z \to \mu\mu$ background, particularly when one of the muons deposits a large fraction of its energy in the calorimeter.

7.3.1.3 Categorization

After the baseline selection, events are categorized into two categories: a VBF category, which targets events produced via the VBF mechanism, and a non-VBF category, which includes all other production modes, dominated by ggF. This categorization is done to be able to use specific selections criteria in each category to enhance the signal-to-background ratio for the LFV Higgs boson signals.

The VBF selection category is designed to improve the signal-to-background ratio by applying criteria that enhance the purity of events originating from VBF production. This set of requirements specifically targets jet kinematic properties and event topology, focusing on features that are characteristic of the VBF production mode. Events are required to have at least two jets with $p_{\text{T}} > 30$ GeV, and the leading jet must have $p_{\text{T}} > 40$ GeV. The two leading jets are required to have a pseudorapidity separation of $\Delta\eta_{\text{jj}} > 3$ and an invariant mass of $m_{\text{jj}} > 400$ GeV.

Events that do not meet the VBF selection criteria are placed into the non-VBF category without any additional requirements. The majority of events passing the baseline selection fall into the non-VBF category, with only approximately 3% entering the VBF category. The event yields in the VBF and non-VBF categories are shown in Table 7.2 for the $H \to e\tau$ search and in Table 7.3 for the $H \to \mu\tau$ search. The predicted signal and background yields are obtained from the 2-POI likelihood fit that will be described in Section 7.7.

7.3.2 Selection in the $\ell\tau_{\text{had}}$ channel

As in the $\ell\tau_{\ell'}$ channel, the event selection process starts with the baseline selection, which is subsequently divided into two categories: VBF and non-VBF. For an event to qualify under the baseline criteria, it must include exactly one reconstructed light lepton (electron or muon, collectively referred to as a lepton) and one reconstructed $\tau_{\text{had-vis}}$, each with opposite charge.

To eliminate background from top-quark processes, any events containing b-tagged jets are excluded. Additionally, angular requirements are applied to reduce contributions from certain backgrounds: $\Sigma \cos \Delta\phi(\ell, E_{\text{T}}^{\text{miss}}) > -0.35$ is imposed to suppress W + jets background, while $|\Delta\eta(\ell, \tau_{\text{had-vis}})| < 2$ helps in mitigating the background. In the $e\tau_{\text{had}}$ channel, an electron veto is applied to the reconstructed $\tau_{\text{had-vis}}$ candidate to suppress contamination from $Z(\to ee)$ + jets events where one electron is misidentified as a $\tau_{\text{had-vis}}$ object.

The VBF category targets the enhancement of the signal originating from the VBF production mode. Specific selection criteria are applied, focusing on jet kinematics. Events that do not fulfill the VBF selection criteria are assigned to the non-VBF category without any additional requirement, except for events in the $e\tau_{\text{had}}$ channel. In the case of $e\tau_{\text{had}}$, further suppression of $Z(\to ee)$ + jets events is necessary, which is achieved by rejecting events if the visible mass $m_{\text{vis}}(e, \tau)$ lies within the $90 < m_{\text{vis}}(e, \tau) < 100$ GeV range. This requirement reduces the $Z(\to ee)$ + jets background by 57%, while lowering the signal acceptance by approximately 20% and reducing the overall data yields by 13%.

7.4 Background estimation

The analysis exploits two primary methods for background estimation: the MC-based method, used in the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels, and the alternative symmetry method, used in the $\ell\tau_{\ell'}$ channel. The MC-based method relies on the simulation of SM processes to estimate the background, correcting the normalization of the main background components using data in control regions. The misidentified leptons are estimated in a data-driven way for both background estimation methods.

The main background of the analysis for both channels is the $Z \to \tau\tau$ process. W + jets and multijet processes are also significant backgrounds in both channels, where one of the jets is misidentified as an electron, a muon or a $\tau_{\text{had-vis}}$. The

	Channel	$e\tau_\mu$					
	Region	Non-VBF			VBF		
	Category	SR	$Z \to \tau\tau$ CR	Top-quark CR	SR	$Z \to \tau\tau$ CR	Top-quark CR
Signal, $H \to e\tau$	ggF	270 ± 160	37 ± 22	14 ± 8	4.3 ± 2.6	0.55 ± 0.33	0.57 ± 0.34
	VH	12 ± 7	1.4 ± 0.9	2.9 ± 1.7	0.22 ± 0.13	0.027 ± 0.023	0.053 ± 0.033
	VBF	15 ± 9	2.0 ± 1.2	1.6 ± 1.0	8 ± 5	0.8 ± 0.5	0.8 ± 0.5
Signal, $H \to \mu\tau$	ggF	19 ± 9	3.6 ± 1.7	1.4 ± 0.7	0.57 ± 0.30	0.045 ± 0.023	0.07 ± 0.05
	VH	2.4 ± 1.2	0.32 ± 0.16	0.51 ± 0.26	0.06 ± 0.04	0.004 ± 0.002	0.011 ± 0.013
	VBF	1.7 ± 0.8	0.18 ± 0.09	0.17 ± 0.09	0.7 ± 0.4	0.045 ± 0.024	0.09 ± 0.05
Background	$Z \to \tau\tau$	23000 ± 400	57500 ± 900	2620 ± 180	450 ± 80	250 ± 40	133 ± 15
	$Z \to \ell\ell$	510 ± 150	340 ± 120	53 ± 19	9 ± 4	5 ± 5	4.3 ± 2.5
	Top-quark	34600 ± 1300	6270 ± 220	298800 ± 1700	1850 ± 70	319 ± 22	15470 ± 120
	SM Higgs ggF	1200 ± 100	1090 ± 60	92 ± 13	37 ± 7	12.1 ± 2.3	5.6 ± 2.1
	SM Higgs VBF	109 ± 4	46.6 ± 2.4	13.1 ± 0.8	74.9 ± 3.2	19.5 ± 1.1	8.0 ± 0.6
	SM Higgs VH	34 ± 5	12.0 ± 1.9	7.6 ± 1.2	0.6 ± 0.4	0.19 ± 0.12	0.19 ± 0.07
	SM Higgs $t\bar{t}H$	8.5 ± 1.4	1.28 ± 0.21	214 ± 34	0.45 ± 0.08	0.080 ± 0.018	8.8 ± 1.5
	VV	31800 ± 1500	11800 ± 600	2020 ± 180	640 ± 70	100 ± 12	106 ± 12
	Misidentified	9200 ± 900	6000 ± 1000	9200 ± 1600	370 ± 50	98 ± 25	590 ± 100
Total Signal		330 ± 180	45 ± 22	22 ± 10	14 ± 6	1.5 ± 0.6	1.5 ± 0.6
Total Background		100800 ± 400	82700 ± 400	313100 ± 600	3450 ± 60	810 ± 31	16340 ± 120
Data		100769	82902	312902	3383	878	16382

Table 7.2: Observed event yields and predictions as computed by the fit in the SR and CRs of the $e\tau_\mu$ channel. The CRs will be described in Section 7.4.1. The prediction for each sample is determined from the 2-POI likelihood fit performed to measure the $H \to e\tau$ and $H \to \mu\tau$ signals using the $\ell\tau_{\ell'}$ data only. The signal event yields are given for $\mathcal{B}(H \to e\tau) = 0.13\%$ and $\mathcal{B}(H \to \mu\tau) = 0.17\%$. This fit is described in Section 7.7. Uncertainties include statistical and systematic components. The background prediction corresponds to the post-fit yields. The SM Higgs boson background includes contributions from $H \to \tau\tau$ and $H \to WW^*$ decays. The uncertainties in the total background prediction include the effect of correlations between individual uncertainty sources as determined by the fit.

Channel		$e\tau_\mu$					
Region		Non-VBF			VBF		
Category		SR	$Z \to \tau\tau$ CR	Top-quark CR	SR	$Z \to \tau\tau$ CR	Top-quark CR
Signal, $H \to e\tau$	ggF	15 ± 9	2.7 ± 1.6	1.1 ± 0.6	0.44 ± 0.27	0.041 ± 0.026	0.08 ± 0.06
	VBF	1.2 ± 0.8	0.14 ± 0.09	0.15 ± 0.09	0.6 ± 0.4	0.042 ± 0.026	0.07 ± 0.04
	VH	1.9 ± 1.1	0.26 ± 0.15	0.41 ± 0.25	0.06 ± 0.04	0.005 ± 0.004	0.012 ± 0.008
Signal, $H \to \mu\tau$	ggF	270 ± 130	38 ± 18	14 ± 7	4.5 ± 2.3	0.55 ± 0.28	0.64 ± 0.33
	VBF	15 ± 7	2.1 ± 1.1	1.6 ± 0.8	8 ± 4	0.9 ± 0.5	0.8 ± 0.4
	VH	12 ± 6	1.5 ± 0.8	3.0 ± 1.5	0.2 ± 0.1	0.022 ± 0.011	0.059 ± 0.031
Background	$Z \to \tau\tau$	18200 ± 400	46400 ± 600	2000 ± 160	360 ± 70	213 ± 32	113 ± 13
	$Z \to \ell\ell$	610 ± 190	640 ± 180	48 ± 22	5.1 ± 2.9	2.1 ± 1.4	1.5 ± 1.1
	Top-quark	30900 ± 1200	5770 ± 200	274500 ± 900	1750 ± 60	306 ± 22	14500 ± 100
	SM Higgs ggF	942 ± 70	910 ± 50	79 ± 11	32 ± 6	10.5 ± 1.9	4.4 ± 1.6
	SM Higgs VBF	92 ± 4	40.9 ± 2.1	11.1 ± 0.7	64.7 ± 2.8	17.4 ± 1.0	7.0 ± 0.5
	SM Higgs VH	27 ± 4	9.7 ± 1.4	6.7 ± 1.3	0.6 ± 0.4	0.17 ± 0.09	0.07 ± 0.10
	SM Higgs $t\bar{t}H$	7.7 ± 1.3	1.25 ± 0.21	200 ± 110	0.43 ± 0.08	0.061 ± 0.023	8.0 ± 1.3
	VV	28000 ± 1300	10600 ± 500	1820 ± 160	610 ± 60	100 ± 12	100 ± 10
	Misidentified	11400 ± 700	8200 ± 600	8900 ± 600	390 ± 40	81 ± 18	460 ± 50
Total Signal		305 ± 160	45 ± 19	19 ± 8	13 ± 5	1.6 ± 0.6	1.6 ± 0.5
Total Background		90500 ± 400	72680 ± 330	287600 ± 500	3230 ± 60	733 ± 28	15210 ± 110
Data		90531	72511	287734	3138	712	15228

Table 7.3: Observed event yields and predictions as computed by the fit in the SR and CRs of the $e\tau_\mu$ channel. The CRs will be described in Section 7.4.1. The prediction for each sample is determined from the 2-POI likelihood fit performed to measure the $H \to e\tau$ and $H \to \mu\tau$ signals using the $\ell\tau_{\ell'}$ data only. The signal event yields are given for $\mathcal{B}(H \to e\tau) = 0.13\%$ and $\mathcal{B}(H \to \mu\tau) = 0.17\%$. This fit is described in Section 7.7. Uncertainties include statistical and systematic components. The background prediction corresponds to the post-fit yields. The SM Higgs boson background includes contributions from $H \to \tau\tau$ and $H \to WW^*$ decays. The uncertainties in the total background prediction include the effect of correlations between individual uncertainty sources as determined by the fit.

top-quark background is important in the $\ell\tau_{\ell'}$ channel. There are other minor background contributions, such as diboson production, $Z \to \ell\ell$, $H \to \tau^+\tau^-$ and $H \to WW^*$, which are considered in the analysis.

The next sections describe the background estimation methods used in each channel, with a focus on the MC-based method in the $\ell\tau_{\ell'}$ channel.

The background estimation in the MC-based analyses relies on several control and validation regions, which are used to constrain the normalization of the background components, validate the modeling of the background processes and estimate the misidentified lepton background. The selection criteria to define these regions are summarized in Table 7.4, and is further described in the following paragraphs.

7.4.1 MC-based $\ell\tau_{\ell'}$ channel

The background estimation for this analysis utilizes both MC simulation and data-driven methods. The main backgrounds include processes with top quarks, diboson events, $Z \to \tau\tau$, misidentified leptons, $Z \to \ell\ell$, and backgrounds from other Higgs boson processes.

The contributions from top-quark backgrounds ($t\bar{t}$ and single-top) and $Z \to \tau\tau$ are determined using MC simulation, with their normalization set through control regions enriched in top quark or $Z \to \tau\tau$ production. These control regions are incorporated into a combined fit that includes both signal and control regions, enabling a direct determination of normalization factors for top quark and $Z \to \tau\tau$

Selection	$\ell\tau_{\ell'}$	$\ell\tau_{\text{had}}$
Top-quark CR VBF and non-VBF	inverted b-veto: ≥ 1 b-tagged jet ($p_{\text{T}} > 25$ GeV and $\lvert\eta\rvert < 2.4$)	
$Z \to \tau\tau$ CR VBF and non-VBF	inverted $p_{\text{T}}^{\ell_1}$ requirement: $35\text{ GeV} < p_{\text{T}}^{\ell_1} < 45\text{ GeV}$	
$Z \to \mu\mu$ CR	$35\,\text{GeV} < p_{\text{T}}^{\ell_1} < 45\,\text{GeV}$ $75\,\text{GeV} < m_{\ell\ell} < 100\,\text{GeV}$ $1.25 < p_{\text{T}}^{\text{track}}(\ell_2)/p_{\text{T}}^{\text{cluster}}(\ell_2) < 3$ $\Delta\phi\left(\ell_2, E_{\text{T}}^{\text{miss}}\right) < 1.5$	
Diboson VR	$p_{\text{T}}^{\ell_2} > 30\,\text{GeV}$ $100\,\text{GeV} < m_{\ell\ell} < 150\,\text{GeV}$ $m_{\text{T}}(\ell_2, E_{\text{T}}^{\text{miss}}) > 20\,\text{GeV}$ $n_{\text{j}} = 0$, jets with $p_{\text{T}} > 30\,\text{GeV}$	

Table 7.4: Event selection and categorization for the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ CRs.

backgrounds in the statistical analysis (this is discussed further in Sections 7.4.1.2 and 7.4.1.1). The simultaneous determination of the normalization factors for these backgrounds in the fit allows the usage of the events in the SRs and CRs to constrain these normalization factors, taking into account also the uncertainties in the background predictions and the correlations between the different sources of systematic uncertainties.

Top quark control regions are selected by inverting the b-tag veto, requiring at least one b-tagged jet. Conversely, $Z \to \tau\tau$ control regions are defined by requiring the leading lepton's transverse momentum to be within 35 GeV $< p_{\mathrm{T}}^{\ell_1} <$ 45 GeV. The events in the top and $Z \to \tau\tau$ control regions are included in the statistical fit, and normalization factors for these backgrounds are determined by fitting across signal and control regions. The determination of this normalization factors relies only on the yield information from the control regions, and not on the shape of any distribution. As in the case of the normalization factors for the top-quark background, the information from the $Z \to \tau\tau$ control regions and the SRs is used to constrain the normalization of the $Z \to \tau\tau$ background.

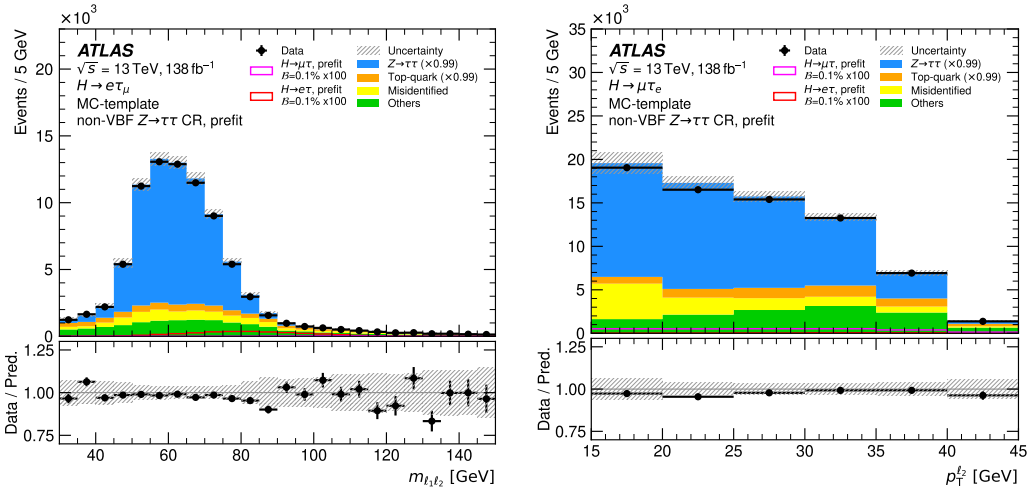
The background from misidentified leptons is estimated with a data-driven method described in Section 7.4.1.5. The diboson background is evaluated using MC simulation, with validation through comparison in a dedicated region, where agreement with data is confirmed (see Section 7.4.1.3). The $Z \to \ell\ell$ background is similarly estimated from MC, and a suitable normalization factor is obtained in a dedicated CR. Backgrounds originating from Higgs boson decays are also assessed through MC simulation.

7.4.1.1 $Z \to \tau\tau$ background

The $Z \to \tau\tau$ process represents a substantial component of the background in this analysis, accounting for approximately 23% of the total background in the non-VBF SR and 11% in the VBF SR. A set of CRs (one for the non-VBF category and one for the VBF category) are defined to validate the $Z \to \tau\tau$ simulation and help in the determination of its normalization within the statistical analysis. These CRs are constructed to be statistically independent of the SRs.

In each analysis category of the $e\tau_\mu$ and $\mu\tau_e$ channels, the $Z \to \tau\tau$ CR is defined by applying a specific transverse momentum criterion on the leading lepton, with 35 GeV $< p_{\mathrm{T}}^{\ell_1} <$ 45 GeV. This setup yields a $Z \to \tau\tau$ purity of around 65% in the non-VBF CR and approximately 32% in the VBF CR, ensuring sufficiently high purity to effectively constrain the $Z \to \tau\tau$ normalization.

For the statistical analysis, the $Z \to \tau\tau$ CRs are incorporated with two normalization factors, each corresponding to one of the categories, to further constrain the $Z \to \tau\tau$ normalization within the SRs. This combined approach enhances the precision of the $Z \to \tau\tau$ background estimate in the final fit.



(a) (b)

Figure 7.2: Pre-fit distributions of key observables for the $Z \to \tau\tau$ non-VBF CRs in the $e\tau_\mu$ (left) and $\mu\tau_e$ (right) final states are shown: (a) $m_{\ell_1\ell_2}$ and (b) $p_{\mathrm{T}}^{\ell_2}$. The hashed band reflects the pre-fit uncertainties, including statistical and systematic contributions. For the $Z \to \tau\tau$ and top-quark backgrounds, normalization factors (shown in the legend) are derived from a likelihood fit performed separately in the VBF and non-VBF categories of the MC-based $\ell\tau_{\ell'}$ channel. Pre-fit signal shapes, assuming $\mathcal{B}(H \to \ell\tau) = 0.1\%$, are scaled by a factor of 100 for improved visibility.

7.4.1.2 Top-quark background

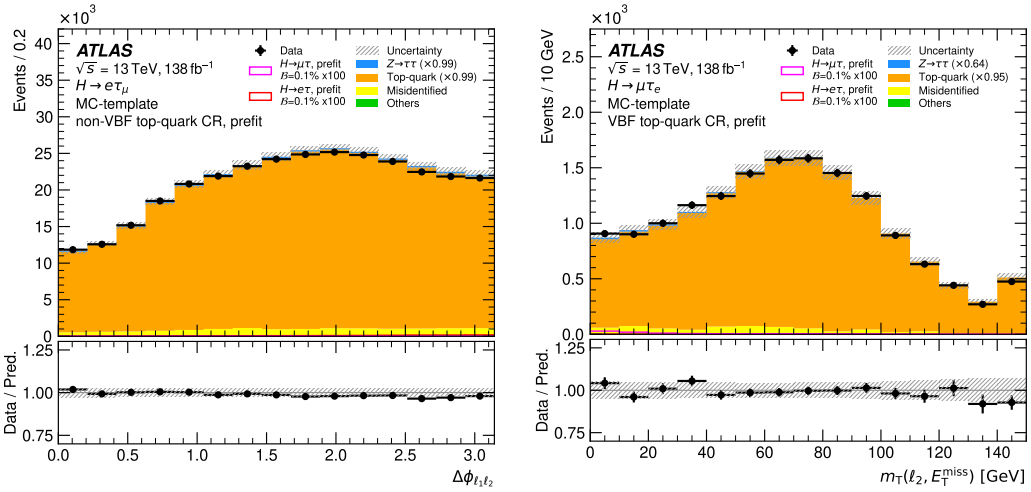
Top quark processes are a primary source of background in this analysis, contributing between 34% and 54% of the total background across different categories. Their modeling is validated through specific CRs that are statistically independent of the SRs. Four CRs are defined in total, one for each category (non-VBF and VBF) and channel ($\mu\tau_e$ and $e\tau_\mu$).

Each top quark CR is constructed by applying the baseline requirements for either non-VBF or VBF classification, except for the b-tagged jet veto, which is inverted to ensure at least one b-tagged jet is present. This configuration results in a high top-quark background purity of approximately 95% in these CRs, providing a high-confidence background sample for comparison with data. The top quark modeling in simulation shows good agreement with data, as shown in Figure 7.3.

For each category, top-quark normalization factors are derived in the statistical analysis simultaneously fitting the data in the SRs and the top-quark CRs.

7.4.1.3 Diboson background

Top-quark and $Z \to \tau\tau$ processes are among the primary backgrounds in this analysis, with dedicated CRs implemented to constrain their normalization in the statistical fit. The diboson background also contributes significantly, ranging from 19% to 32% of the total background depending on the category, but defining a control region for this background that is both orthogonal to the signal regions SRs



(a) (b)

Figure 7.3: Pre-fit distributions of key observables for the top-quark CRs in the $e\tau_\mu$ (left) and $\mu\tau_e$ (right) final states are shown: (a) $\Delta\phi_{\ell_1,\ell_2}$ and (b) $m_{\mathrm{T}}(\ell_2, E_{\mathrm{T}}^{\mathrm{miss}})$. The hashed band reflects the pre-fit uncertainties, including statistical and systematic contributions. For the $Z \to \tau\tau$ and top-quark backgrounds, normalization factors (shown in the legend) are derived from a likelihood fit performed separately in the VBF and non-VBF categories of the MC-based $\ell\tau_{\ell'}$ channel. Pre-fit signal shapes, assuming $\mathcal{B}(H \to \ell\tau) = 0.1\%$, are scaled by a factor of 100 for improved visibility.

and sufficiently pure to effectively constrain it is challenging. Instead, a dedicated validation region (VR) has been developed to assess the accuracy of the diboson modeling.

This diboson VR is derived from the baseline selection with additional criteria to enhance diboson event contributions while minimizing contamination from other backgrounds. Specifically, events in the VR must satisfy the baseline selection and feature a $p_{\mathrm{T}}^{\ell_2}$, greater than 30 GeV, with the visible mass between the two leptons constrained within 100 GeV $< m_{\mathrm{vis}} <$ 150 GeV. Additionally, the transverse mass, m_{T}, calculated between the subleading lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$, must exceed 20 GeV. No jets with p_{T} above 30 GeV are allowed.

These requirements yield a diboson purity of approximately 69.7% for the $\mu\tau_e$ channel and 68.5% for the $e\tau_\mu$ channel, with diboson processes comprising around 67% of the total background in this VR. Good agreement is observed between the data and simulation within the diboson VR, as shown in Figure 7.4. No normalization uncertainty is assigned to the diboson background, as the agreement between data and simulation is satisfactory.

7.4.1.4 $Z \to \ell\ell$ background

The $Z \to \mu\mu$ and $Z \to ee$ processes both feature two leptons of the same flavor in the final state. While the $Z \to \ell\ell$ background is not a dominant component in the analysis, a small fraction of muons may radiate photons with sufficient energy while traversing the electromagnetic calorimeter, leading to their misidentification as

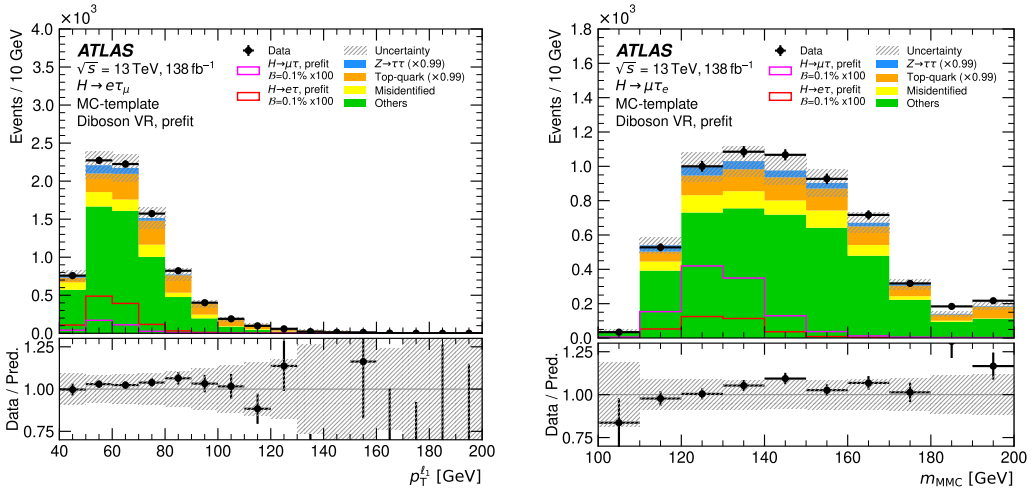
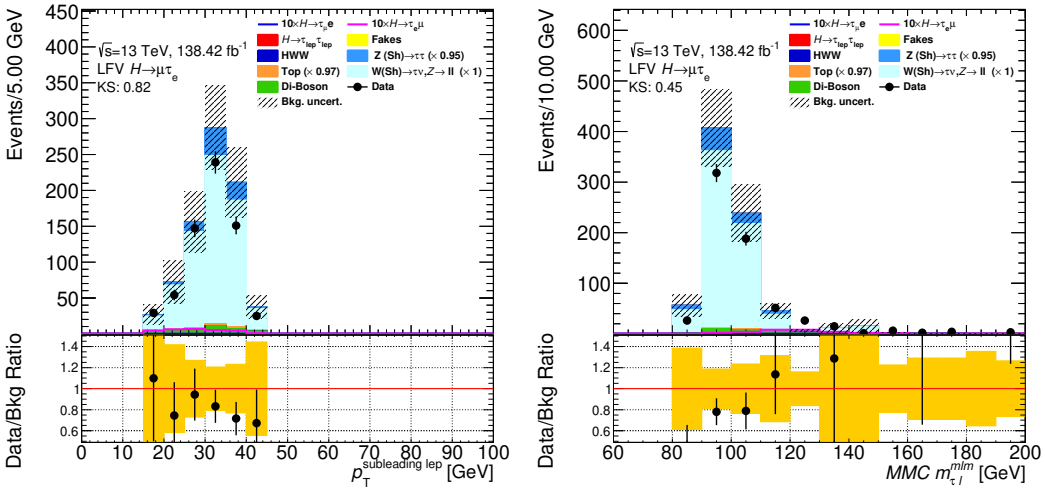


Figure 7.4: Pre-fit distributions of key variables in the diboson validation region are presented: (a) the $p_{\mathrm{T}}^{\ell_1}$ distribution for the $e\tau_\mu$ final state and (b) the m_{MMC} mass distribution for the $\mu\tau_e$ final state. The hashed band represents pre-fit uncertainties, incorporating statistical and systematic components. The $Z \to \tau\tau$ and top-quark backgrounds are scaled by normalization factors (listed in the legend) derived from separate likelihood fits in the VBF and non-VBF categories of the MC-based $\ell\tau_{\ell'}$ channel. Pre-fit signal shapes, assuming $\mathcal{B}(H \to \ell\tau) = 0.1\%$, are amplified by a factor of 100 for improved visibility.

electrons. The probability of misidentifying an electron as a muon is negligible, and thus no corresponding background arises from the $Z \to ee$ process. It is important to distinguish this from the $Z \to \ell\ell$ backgrounds involving jets misidentified as leptons, which are addressed through the data-driven estimation detailed in Section 7.4.1.5.

To investigate potential mismodeling and quantify the normalization uncertainty of the $Z \to \mu\mu$ background, a dedicated CR is defined to enhance the contribution of $Z \to \mu\mu$ processes. The requirements for this control region are designed to be orthogonal to the signal regions and maintain the same selection criteria as the baseline selection, with the following exceptions: the transverse momentum of the leading lepton, $p_{\mathrm{T}}^{\ell_1}$, is constrained within the range of $35\,\mathrm{GeV} < p_{\mathrm{T}}^{\ell_1} < 45\,\mathrm{GeV}$; the visible mass, m_{vis}, is required to be compatible with the Z peak, specifically $75\,\mathrm{GeV} < m_{\ell\ell} < 100\,\mathrm{GeV}$; if the second lepton, ℓ_2, is an electron, the ratio of the track momentum to the cluster energy must satisfy $1.25 < p_{\mathrm{T}}^{\mathrm{track}}(\ell_2)/p_{\mathrm{T}}^{\mathrm{cluster}}(\ell_2) < 3$ to select events where one muon deposits a significant fraction of its energy in the calorimeter while rejecting those lacking cluster information; and the azimuthal angle difference, $\Delta\phi\left(\ell_2, E_{\mathrm{T}}^{\mathrm{miss}}\right)$, must be less than 1.5.

MC simulation studies indicate that misidentified muons are reconstructed as the subleading lepton in nearly 100% of cases. Most $Z \to \mu\mu$ background events are found in the $\mu\tau_e$ channel, although the assignment algorithm could allocatee a fraction of these events to the $e\tau_\mu$ channel. Studies using MC simulations indicate that the $Z \to \mu\mu$ background is negligible in the $e\tau_\mu$ channel. The purity of the



(a) (b)

Figure 7.5: Pre-fit distributions of key variables in the $Z \to \ell\ell$ CR are presented: (a) the $p_{\mathrm{T}}^{\ell_1}$ distribution and (b) the m_{MMC} mass distribution, both for the $\mu\tau_e$ final state. The uncertainty band reflects pre-fit uncertainties, including statistical uncertainties and normalization uncertainties for the backgrounds. The $Z \to \tau\tau$ and top-quark backgrounds are scaled by normalization factors estimated before the fit, scaling the corresponding MC templates to background-subtracted data in the CRs.

$Z \to \mu\mu$ region is approximately 82%. Background-subtracted data exhibit a deviation of 25% from the $Z \to \mu\mu$ MC simulation. A normalization uncertainty of 25% is assigned to $Z \to \mu\mu$ events across all the regions considered in the fit, and the normalization of the $Z \to \mu\mu$ background is scaled by a factor of 0.75 before the fit. Pre-fit distributions in the $Z \to \mu\mu$ CR are shown in Figure 7.5 before multiplying the MC templates by the $Z \to \ell\ell$ normalization factor.

7.4.1.5 Misidentified objects

The background from misidentified leptons primarily arises from events in which reconstructed electrons or muons are actually jets, misidentified as prompt leptons. This background mainly originates from W+jets, multijet events, and hadronically decaying $t\bar{t}$ or single top-quark processes. It accounts for approximately 10 to 12% of the total background in the non-VBF and VBF SRs.

To estimate these backgrounds from data, dedicated CRs, referred to as fake CRs, are defined, as outlined in Figure 7.6. The selection of these regions is guided by two factors: the charge of the leptons (same sign, SS, or opposite sign, OS) and their quality criteria. The method relies on a Transfer Factor (TF), derived from events in fake CRs (SSF and OSF), to estimate the misidentified lepton contribution in the signal region (OSN). The TF is applied to events in a statistically independent region (SSN), which shares characteristics with the signal region but is orthogonal to it.

The procedure involves the following regions:

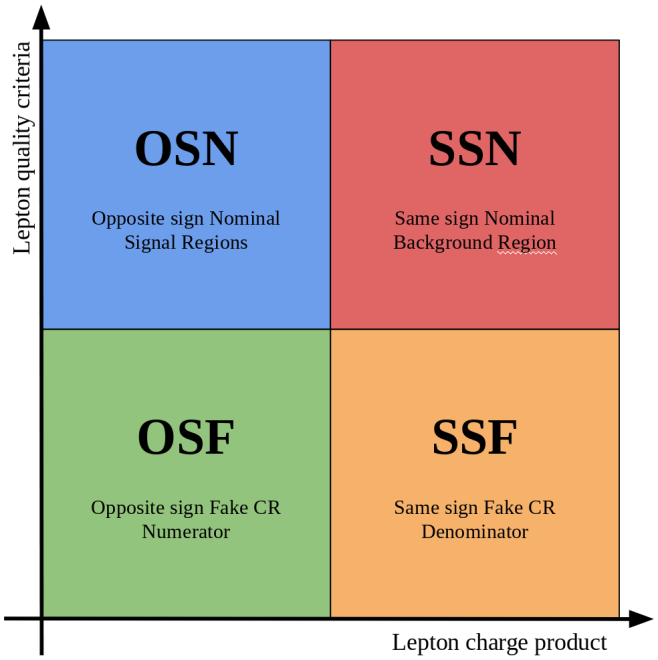


Figure 7.6: Sketch of the event selection for the misidentified lepton background estimation, showing the signal region (OSN), the same-sign region (SSN), and the fake control regions (OSF and SSF). The signal region is defined by the baseline selection, while the fake control regions are defined by inverting the isolation and identification criteria of the subleading lepton.

- **OSN**: The OS nominal region, where both leptons meet isolation criteria; this serves as the signal region.
- **SSN**: The SS nominal region, used as a control to develop templates for modeling the variable distributions.
- **OSF**: The OS fake CR, utilized as the numerator in the TF. This CR is defined by inverting the isolation and identification criteria of the leptons.
- **SSF**: The SS fake CR, providing the denominator in the TF calculation. As for the OSF region, the isolation and identification criteria of the leptons are changed to select events in this region.

These three CRs are statistically independent of each other, ensuring reliability in the misidentified lepton background estimation for the signal region.

Fake CRs are further defined based on the isolation and identification quality criteria of the subleading lepton, as the misidentified lepton is the subleading one in 96% of cases and identified as non-prompt in 85% of cases according to the lepton assignment. Thus, the background estimation proceeds by analyzing $e\mu$ and μe final states independently of $e\tau_\mu$ and $\mu\tau_e$ channel definitions.

Isolation requirements for the fake CRs vary between the $e\mu$ and μe final states: the *Medium* identification and *Gradient* isolation requirements are consistently applied to the leading lepton, while criteria for the subleading lepton differ. In the

μe final state, the electron must fail the *Medium* identification requirement (but still meet the *Loose* identification) or fail the *Gradient* isolation, whereas in the $e\mu$ final state, the muon must fail the *Gradient* isolation but meet the *Medium* identification criteria.

The transfer factors are defined as:

$$f_{\text{trans}} = \frac{N^{\text{OS}}_{\text{fake CR}}}{N^{\text{SS}}_{\text{fake CR}}}, \tag{7.1}$$

where $N^{\text{OS}}_{\text{fake CR}}$ and $N^{\text{SS}}_{\text{fake CR}}$ are the yields of events in the OS and SS fake regions, respectively, with contributions from processes involving two prompt leptons subtracted, as estimated from MC simulation. These backgrounds in the fake CR primarily arise from prompt lepton processes such as $Z \to \tau\tau$, diboson, and leptonically decaying $t\bar{t}$ events. The subtraction uses lepton truth information to ensure accurate removal of events where leptons are classified as prompt electrons or muons. Background electrons from photon conversions are considered part of the misidentified lepton background and are estimated from data. Consequently, events with photon conversions are removed from the MC simulation to avoid double-counting.

To achieve accurate background estimation for misidentified leptons, the TF is calculated independently for $e\mu$ and μe final states and, within each final state, for b-veto and b-tagged events, as well as for events triggered by single-lepton and dilepton triggers.

The transfer factors, calculated using Eq. 7.1, are presented in Table 7.5 as a function of final state, trigger region, and b-tag/veto condition. Values above unity indicate a greater number of events in the OS region than in the SS region. Statistical uncertainties for each TF are shown, representing one of the sources of uncertainty for the misidentified lepton background in this analysis.

region		$e\mu$	μe
b-veto	Single lepton trigger	2.83 ± 0.03	2.18 ± 0.02
	Dilepton trigger	2.01 ± 0.04	2.45 ± 0.15
b-tag	Single lepton trigger	3.65 ± 0.04	2.71 ± 0.02
	Dilepton trigger	2.50 ± 0.50	3.58 ± 0.15

Table 7.5: Transfer factors and their statistical uncertainties across different trigger and b-tag regions, calculated from Fake OS and Fake SS events. Only statistical uncertainties are shown.

The use of OSF and SSF regions to define the TF applied to SSN events provides a more precise modeling of misidentified lepton kinematics. This is because the p_{T} of the misidentified leptons aligns more closely between SS and OS events than between OSN and OSF regions.

Figure 7.7 shows the post-fit distributions of key variables in the $\ell\tau_{\ell'}$ channel, including the mass reconstructed using the MMC algorithm, the collinear mass, the missing transverse energy, and the transverse mass of the subleading lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$. The distributions are shown after the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals, based on data in the MC-based $\ell\tau_{\ell'}$ channel. The fit setup is described in Section 7.7. All distributions show good agreement between data and prediction.

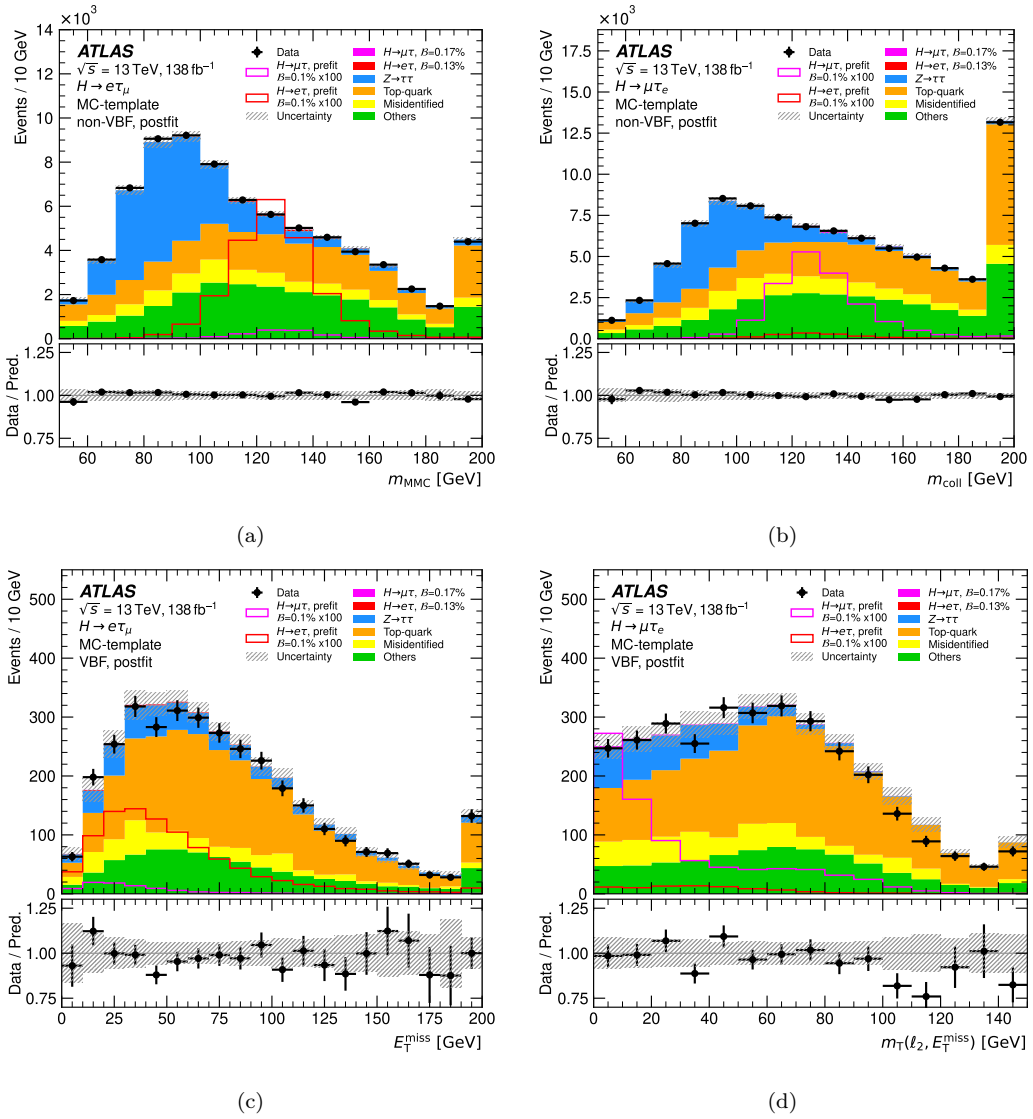


Figure 7.7: Distributions of kinematic variables for the $e\tau_\mu$ (left) and $\mu\tau_e$ (right) final states are shown following a simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals, based on data in the MC-based $\ell\tau_{\ell'}$ channel. The fit setup is described in Section 7.7. For the non-VBF category, (a) the mass m_{MMC} and (b) the collinear mass m_{coll} are displayed, while for the VBF category, (c) the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ and (d) the transverse mass $m_{\mathrm{T}}(\ell_2, E_{\mathrm{T}}^{\mathrm{miss}})$ are presented. The hashed band illustrates the total uncertainties obtained from the likelihood fit. Pre-fit signal shapes assume a branching ratio $\mathcal{B}(H \to \ell\tau) = 0.1\%$, and are scaled by a factor of 100 for visibility. The post-fit signal contributions are included as part of the predicted distributions.

7.4.2 MC-based $\ell\tau_{\text{had}}$ channel

In the MC-based $\ell\tau_{\text{had}}$ channel, the primary background contributions arise from $Z \to \tau\tau$ events, estimated using simulated events. The $Z \to \tau\tau$ events account for 48%–67% of the total background, depending on the category. The second-largest background source consists of events with a jet misidentified as $\tau_{\text{had-vis}}$, contributing 22%–30% of the total background. These misidentified events are evaluated using the Fake-Factor technique. Smaller backgrounds from other processes such as top-quark, $Z \to \ell\ell$, diboson production, $H \to \tau\tau$, and $H \to WW^*$ are also estimated from simulation.

The shape of the $Z \to \tau\tau$ background distribution is modeled using simulation, with its normalization in the VBF and non-VBF categories constrained by data in the BDT score distributions (see Section 7.5) in the SRs. No dedicated CRs are defined for the $Z \to \tau\tau$ background in the $\ell\tau_{\text{had}}$ channel, since the first bins of the BDT score distributions are dominated by $Z \to \tau\tau$ events and the information from the SRs is sufficient to constrain the normalization of this background. The shape of the top-quark background distribution is also modeled using simulation, and its normalization factors in the VBF and non-VBF categories are constrained using the $\ell\tau_{\ell'}$ top-quark CRs in the simultaneous fit of the $\ell\tau_{\text{had}}$ and $\ell\tau_{\ell'}$ channels.

Diboson events are 2%–5% of the total background, depending on the category. In the $\mu\tau_{\text{had}}$ channel, the $Z \to \mu\mu$ process contributes about 5%–6% of the total background in the non-VBF category and 3%–4% in the VBF category. Its modeling is validated in a dedicated validation region, where the $Z \to \mu\mu$ event fraction is ∼65%. The $Z \to \mu\mu$ validation region is defined by applying the baseline selection criteria and two additional requirements: firstly, the τ candidate should be within $|\eta| < 0.1$, and secondly, the collinear mass should be compatible with the Z peak, $90\,\text{GeV} < m_{\tau\tau}^{\text{coll}}(\mu,\tau) < 110\,\text{GeV}$. While no template mismodeling is observed in this region, a global normalization offset of 13% is observed, which is then assigned as an uncertainty in the $Z \to \mu\mu$ contribution in the statistical analysis.

The estimation of the misidentified background in the $\ell\tau_{\text{had}}$ channel follows the same procedure as in the $H \to \tau\tau$ analysis, explained in detail in Section 6.6.2.2. The misidentified background component distribution in the SR is obtained by multiplying the number of events satisfying the SR selection criteria but failing the *Tight* $\tau_{\text{had-vis}}$ identification requirement by a fake factor. The fake factor is defined as the ratio of the number of jets misidentified as $\tau_{\text{had-vis}}$ that meet the *Tight* $\tau_{\text{had-vis}}$ identification criteria to those that do not but still meet the *Very Loose* identification requirement. The fake factor is parameterized based on the p_{T} and track multiplicity of the $\tau_{\text{had-vis}}$. Since the background originates from W+jets and multijet production processes, two independent fake factors, F_W and F_{QCD}, are measured in dedicated W + jets and multijet production CRs, respectively. These regions are defined to be statistically independent of the SRs by inverting specific selection criteria. For the W+jets CR, the selection includes $(\cos\Delta\phi(\ell, E_{\text{T}}^{\text{miss}}) + \cos\Delta\phi(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})) < -0.35$, $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}}) > 60$ GeV, and $m_{\text{T}}(\tau, E_{\text{T}}^{\text{miss}}) > 40$ GeV. For the multijet CR,

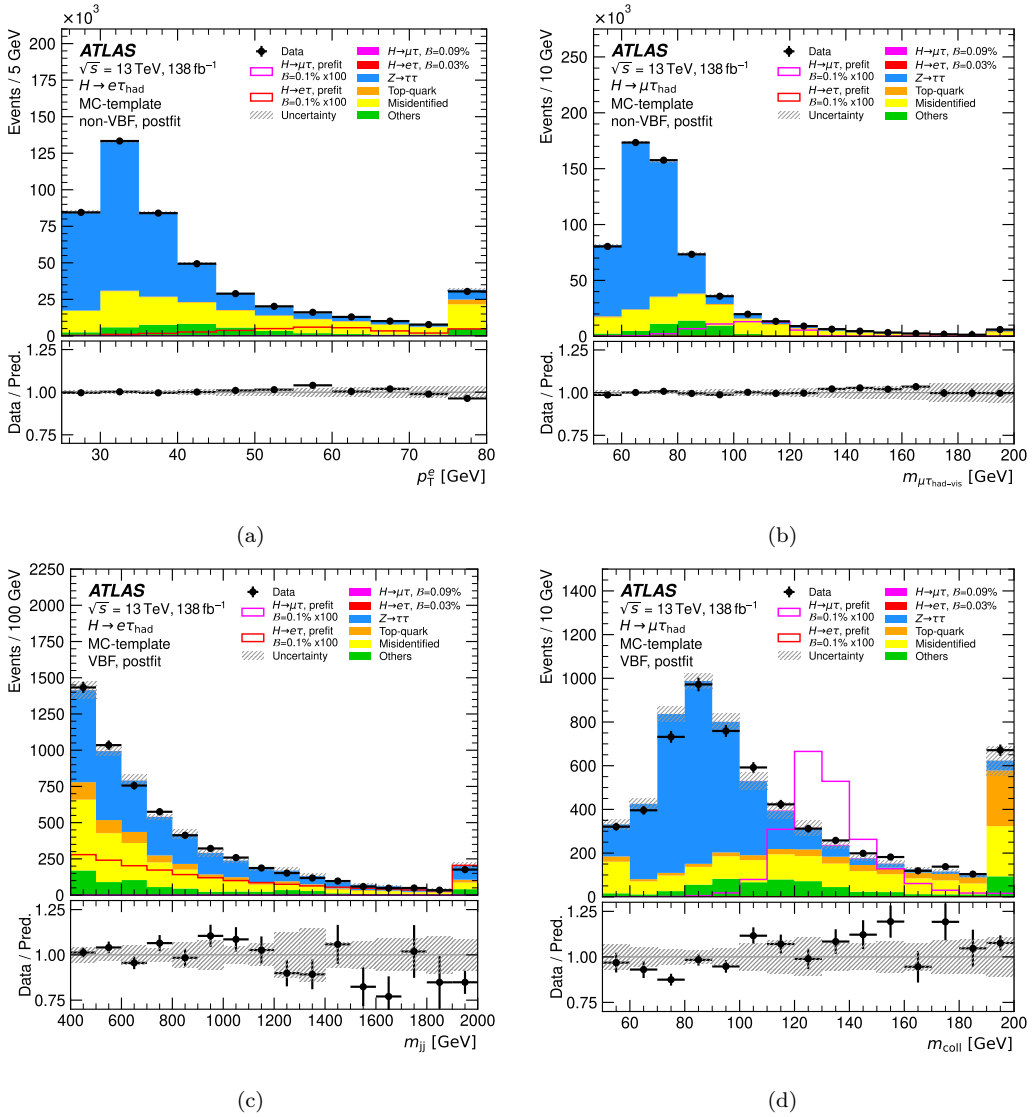


Figure 7.8: Distributions of kinematic variables for the $e\tau_{\text{had}}$ (left) and $\mu\tau_{\text{had}}$ (right) final states are shown following a simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals, based on data in the MC-based $\ell\tau_{\text{had}}$ channel. The fit setup is described in Section 7.7. For the non-VBF category, (a) the electron transverse momentum, p_{T}^{e}, and (b) the visible mass, m_{vis}, are displayed, while for the VBF category, (c) the dijet invariant mass, m_{jj}, and (d) the collinear mass, m_{coll}, are presented. Entries exceeding the horizontal axis range are included in the final bin of each plot. The hashed band represents pre-fit statistical uncertainty, as well as experimental and theoretical uncertainties. Pre-fit signal contributions assume a branching ratio $\mathcal{B}(H \to \ell\tau) = 0.1\%$, scaled by a factor of 100 for visibility. Post-fit signal contributions are incorporated into the predicted distributions.

the selection requires $|\Delta\eta(\ell, \tau_{\text{had-vis}})| \geq 2$ and $m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}}) < 60$ GeV.

The final estimate of the misidentified background contribution in the SRs uses a combined factor: $F = R_{\text{QCD}}F_{\text{QCD}} + (1 - R_{\text{QCD}})F_{W}$. The fraction of multijet events, R_{QCD}, is determined by scaling the multijet CR event count by the ratio of events where the light lepton passes isolation requirements to those where it does not. This ratio is measured in a multijet-enriched region in which the ℓ and τ_{had}

have the same electric charge.

Figure 7.8 illustrates the level of agreement between the data the and the background prediction in the non-VBF and VBF categories for a selection of variables used in the multivariate analysis described in Section 7.5 and that show high discrimination power between signal and background. Signal and background yield normalizations are obtained by performing the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals (discussed in Section 7.8.1) using only data in the $\ell\tau_{\text{had}}$ final state. The distributions show that the data is in good agreement with the background prediction within the total uncertainties.

7.5 Multivariate analysis

To enhance signal separation from multiple background components, two MVA techniques are applied: BDTs, used in the MC-based method, and fully connected NNs, used in the Symmetry method. The non-VBF and VBF event categories employ distinct MVA strategies to exploit the unique characteristics of the VBF topology effectively. For each background estimation approach, the MVA technique demonstrating the highest expected sensitivity is selected, optimizing the signal-to-background discrimination power.

7.5.1 Multivariate techniques in the MC-based $\ell\tau_{\ell'}$ channel

In order to enhance signal-to-background separation, an MVA technique based on BDTs is used. Instead of relying on a single classifier to handle all background types simultaneously, multiple BDTs are trained with each targeting distinct background components. This approach allows each BDT to focus on the unique characteristics of specific background types, enhancing the separation between signal-like and background-like events by focusing on details relevant to each production mode.

Each BDT provides an output score designed to classify events based on their similarity to either signal or background distributions. These scores are then combined into a single metric, used in the statistical fit to evaluate signal presence and strength. The BDT training parameters, listed in Table 7.6, have been selected to balance computational efficiency with model performance. BDTs are trained using the Toolkit for Multivariate Data Analysis (TMVA) [239], and to improve generalization, a k-fold cross-validation process with $k = 5$ is employed, which mitigates overtraining and allows an efficient exploitation of the available statistics.

In k-fold cross-validation, data is split into k subsets, or folds. In each iteration, one fold is held out as a test set, and the remaining $k - 1$ folds are used for training. This procedure is repeated until each fold has served as the test set exactly once, ensuring that the final model evaluations incorporate statistically independent data.

The BDTs employed in this analysis focus on three primary background types. The first BDT ($\text{BDT}_1^{\ell\tau_{\ell'}}$) distinguishes the LFV signal from events with misidentified

BDT Parameter	Value
Number of trees	750
Maximum depth	8
Minimum node size	2.5%
Number of cuts	20
Boost type	Gradient
Use bagged boost	True
Bagged sample fraction	0.5
Shrinkage	0.1

Table 7.6: BDT parameters used for the $\ell\tau_{\ell'}$ analysis.

leptons, estimated through the data-driven method outlined in Section 7.4.1.5. The second BDT ($\text{BDT}_2^{\ell\tau_{\ell'}}$) is trained to classify top-background processes, diboson production, and $H \to WW$ decays, while the third ($\text{BDT}_3^{\ell\tau_{\ell'}}$) targets $Z \to \tau\tau$, $Z \to \ell\ell$, and $H \to \tau\tau$ backgrounds. Each classifier yields a score per event, where scores near zero indicate background-like events, and scores near one suggest signal-like events. These scores are then linearly combined into a single metric, $S_{\text{comb}} = \frac{1}{\sum_{i=1}^{N} c_i} \left(\sum_{i=1}^{N} c_i S_i \right)$, where S_i is the score from each classifier, and c_i are weights optimized for maximizing discovery significance. After normalization, S_{comb} ranges from zero to one, similar to the individual scores.

The optimization of c_i values uses the Asimov significance metric [240], which quantifies the signal-to-background separation for each bin in the final score distribution as $Z_i = \sqrt{2 \left((s_i + b_i) \log \left(1 + s_i/b_i \right) - s_i \right)}$, where s_i and b_i are signal and background events in the i-th bin. The total significance, $Z = \sqrt{\sum_i^{N_{\text{bins}}} Z_i^2}$, is determined by summing over all bins. A systematic scan over coefficient values determines the combination that maximizes Z. Training and optimization are performed together for both $e\tau_\mu$ and $\mu\tau_e$ channels to enhance statistics, though evaluation is performed separately.

Although the same training methodology is used across both VBF and non-VBF categories, different input variables are chosen for each category to exploit jet-related features in the VBF region

7.5.1.1 BDTs in the non-VBF category

Low- and high-level kinematic variables serve as inputs to the BDTs, providing details to enhance signal-to-background separation in the non-VBF category of the LFV $\ell\tau_{\ell'}$ analysis. The variables used as inputs for these BDTs are outlined below:

- The transverse momenta of the Higgs lepton ($p_{\text{T}}^{\ell_H}$) and the tau lepton ($p_{\text{T}}^{\ell_\tau}$) are used, as the $p_{\text{T}}^{\ell_H}$ is generally expected to exceed $p_{\text{T}}^{\ell_\tau}$ in signal-like events,

providing discrimination against backgrounds such as $Z \to \tau\tau$. Both are computed in the Higgs rest frame discussed in Section 7.3.1.1.

- The sum of the subleading lepton p_{T} ($p_{\mathrm{T}}^{\ell_2}$) and $E_{\mathrm{T}}^{\mathrm{miss}}$ and divided by the leading lepton p_{T} ($p_{\mathrm{T}}^{\ell_1}$) is included, as it aids in distinguishing signal-like from background events, particularly those from $Z \to \tau\tau$ processes.

- The angular difference between the Higgs lepton (or tau lepton) and the $E_{\mathrm{T}}^{\mathrm{miss}}$, denoted as $\Delta\phi\left(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}\right)$ ($\Delta\phi\left(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right)$), is expected to be large (small) in signal events. These angles are also computed in the Higgs boson rest frame.

- The transverse mass between the leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$, $m_{\mathrm{T1}} = m_{\mathrm{T}}(\ell_1, E_{\mathrm{T}}^{\mathrm{miss}})$, and $m_{\mathrm{T2}} = m_{\mathrm{T}}(\ell_2, E_{\mathrm{T}}^{\mathrm{miss}})$, provides valuable separation for top-background processes and some separation for $Z \to \tau\tau$ events.

- The angular separation between the leptons, $\Delta R\left(\ell_1, \ell_2\right)$, is another variable expected to differ significantly between signal and background events.

- The $\Delta\alpha$ discriminant is defined by Equation 7.2:

$$\Delta\alpha = \frac{m_H^2 - m_\tau^2}{2p^{\ell_1}p^{\ell_2}} - \frac{p_{\mathrm{T}}^{\ell_1}}{p_{\mathrm{T}}^{\ell_2}}, \tag{7.2}$$

  where m_H and m_τ represent the nominal masses of the Higgs boson and the τ, respectively, and p^{ℓ_1} and p^{ℓ_2} denote the 4-momenta of the leading and subleading leptons. The $p^{\ell_1}p^{\ell_2}$ term represents the product of the two 4-momenta. This discriminant is expected to be close to zero for collinear τ decay products with negligible transverse Higgs boson momentum, which may be the case in signal-like events. In contrast, background events are likely to show nonzero values.

- Two vertex-related variables, $\Delta d_0\left(\ell_1, \ell_2\right)$ (the difference in d_0 between the leptons) and $\sigma_{d_0}^{\ell_\tau}$ (the transverse impact parameter significance of the tau lepton), enhance separation while having reasonable agreement between data and Monte Carlo simulation.

- Three mass reconstructions are utilized to further improve discrimination between signal and background: the visible mass (m_{vis}), the collinear mass ($m_{\tau\tau}^{\mathrm{coll}}$), and the mass from the Missing Mass Calculator technique ($m_{\tau\tau}^{\mathrm{MMC}}$).

A systematic grid search optimizes the combination of BDT coefficients, with the coefficients that yield the highest discovery significance selected. For the non-VBF region, optimal values were found to be $c_1 = 0.2$, $c_2 = 0.9$, and $c_3 = 0.5$.

TMVA's ranking of variable importance helps evaluate BDT performance. An unspecific ranking, based on the Gini index, provides a general measure of separation power, while the method-specific ranking accounts for the frequency and

effectiveness of variable splits in decision tree nodes. Since k-fold cross-validation is used, a ranking is generated per fold, with the final ranking being the average of the variable importance across folds. These rankings are listed in Tables 7.7.

Comparison between the data and the background prediction in the $\mu\tau_e$ non-VBF category is shown in Figures 7.9, 7.10, and 7.11 for the variables used in the non-VBF BDTs. The same distributions for the $e\tau_\mu$ dataset are shown in Figures 7.12, 7.13, and 7.14. The agreement between the data and the background prediction is good within the uncertainties. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes.

The post-fit distributions of the BDT scores for the non-VBF category are shown in Figure 7.19a and 7.19b for the $e\tau_\mu$ and $\mu\tau_e$ datasets, respectively. The signal and background predictions correspond to the best-fit values obtained from the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals described in Section 7.8.1. The agreement between data and the signal and background predictions is good within the total uncertainties, that include statistical and systematic components.

Variable	$\mathrm{BDT}_1^{\ell\tau_{\ell'}}$		$\mathrm{BDT}_2^{\ell\tau_{\ell'}}$		$\mathrm{BDT}_3^{\ell\tau_{\ell'}}$	
	Unspecific	Specific	Unspecific	Specific	Unspecific	Specific
$m_{\tau\tau}^{\mathrm{MMC}}$	0.3390	0.0770	0.3970	0.0808	0.5538	0.0854
$m_{\tau\tau}^{\mathrm{coll}}$	0.2716	0.0649	0.2952	0.0667	0.5121	0.0914
$\Delta\phi\left(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right)$	0.2149	0.0852	0.3736	0.0939	0.0830	0.0650
m_{vis}	0.1762	0.0890	0.1972	0.0963	0.3562	0.1203
m_{T2}	0.1670	0.0616	0.3865	0.0817	0.0426	0.0576
$p_{\mathrm{T}}^{\ell_H}$	0.1434	0.0563	0.1287	0.0441	0.4232	0.0574
$\Delta\alpha$	0.1268	0.0559	0.1075	0.0550	0.3449	0.0596
$\Delta R\left(\ell_1, \ell_2\right)$	0.0944	0.0834	0.1493	0.0858	0.0852	0.0740
$\Delta\phi\left(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}\right)$	0.0769	0.0675	0.1405	0.0695	0.1403	0.0574
$\frac{p_{\mathrm{T}}^{\ell_\tau}}{p_{\mathrm{T}}^{\ell_1}}$	0.0418	0.0640	0.0283	0.0514	0.0477	0.0523
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.0414	0.0484	0.1887	0.0452	0.0426	0.0332
m_{T1}	0.0338	0.0602	0.1137	0.0703	0.2218	0.0670
$\sigma_{d_0}^{\ell_\tau}$	0.0322	0.0563	0.1105	0.0643	0.0017	0.0554
$p_{\mathrm{T}}^{\ell_\tau}$	0.0216	0.0700	0.0382	0.0512	0.0743	0.0734
$\Delta d_0\left(\ell_1, \ell_2\right)$	0.0087	0.0567	0.0736	0.0451	0.0017	0.0515

Table 7.7: Combined unspecific and specific ranking of variables used in the BDTs for the non-VBF category, showing the unspecific separation and specific separation for BDT_1, BDT_2, and BDT_3.

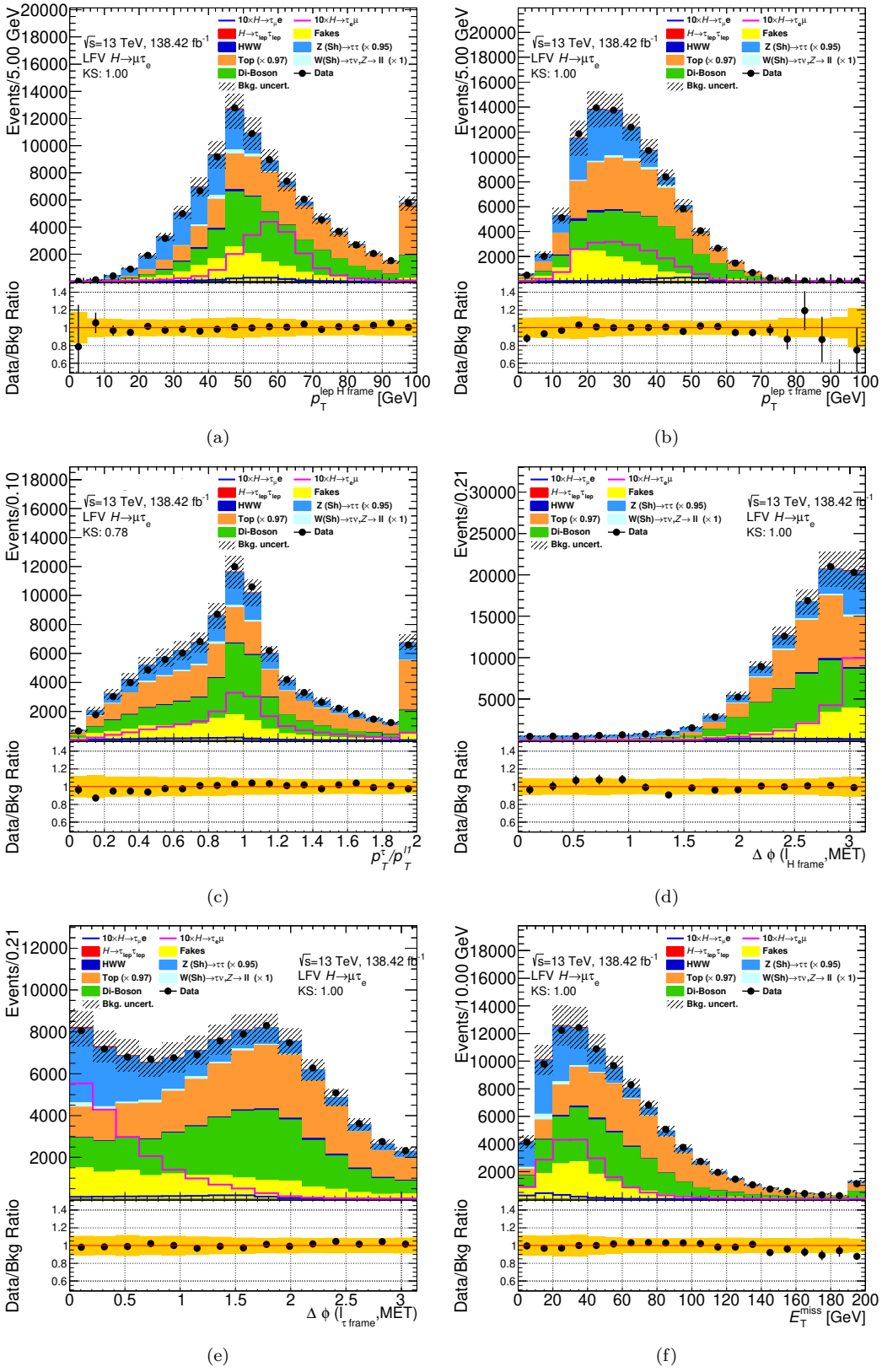


Figure 7.9: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $\mu\tau_e$ channel in the non-VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) $p_{\mathrm{T}}^{\ell_H}$, (b) $p_{\mathrm{T}}^{\ell_\tau}$, (c) $\frac{p_{\mathrm{T}}^{\ell_\tau}}{p_{\mathrm{T}}^{\ell_1}}$, (d) $\Delta\phi\left(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}\right)$, (e) $\Delta\phi\left(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right)$ and (f) $E_{\mathrm{T}}^{\mathrm{miss}}$.

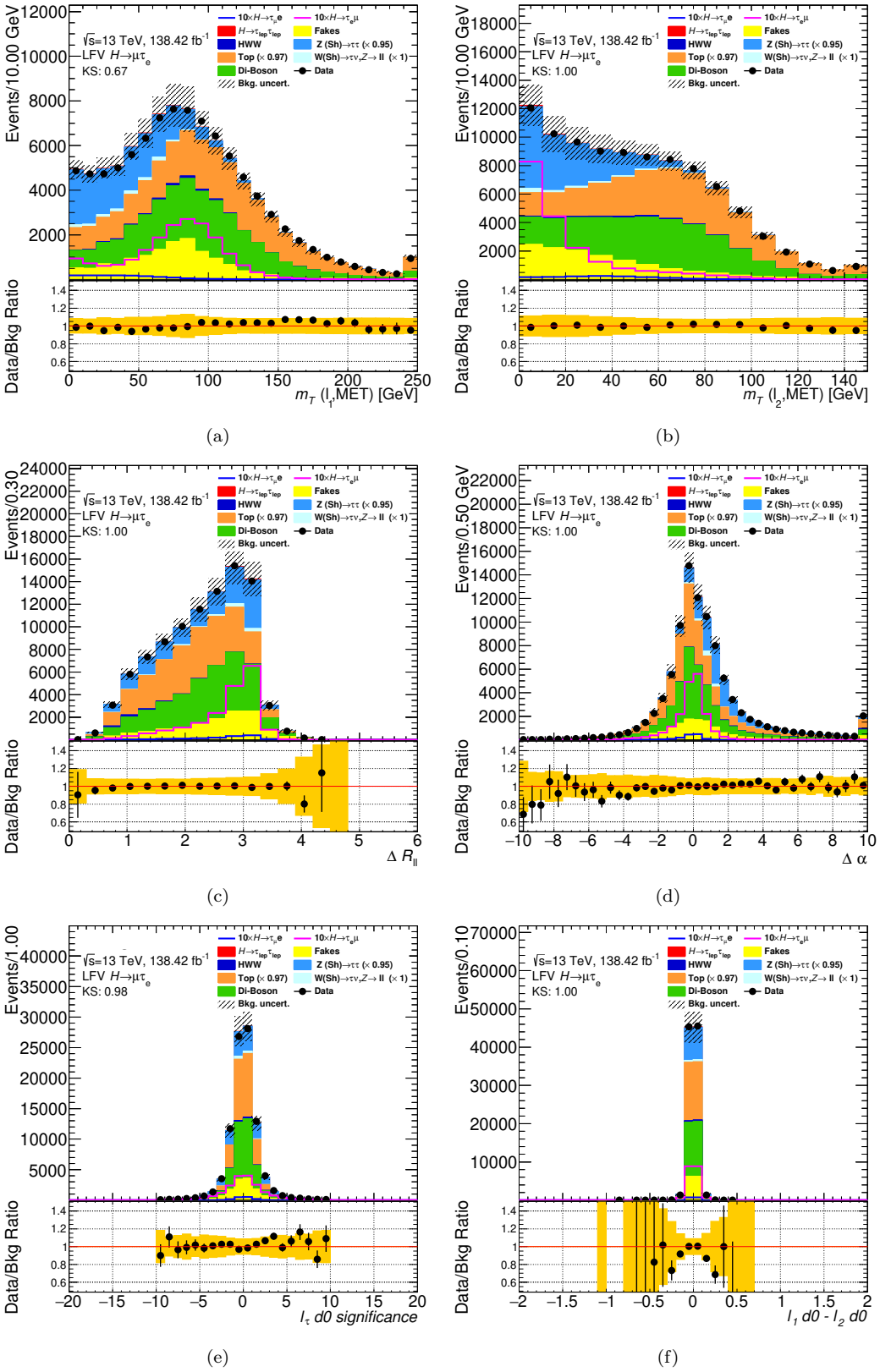
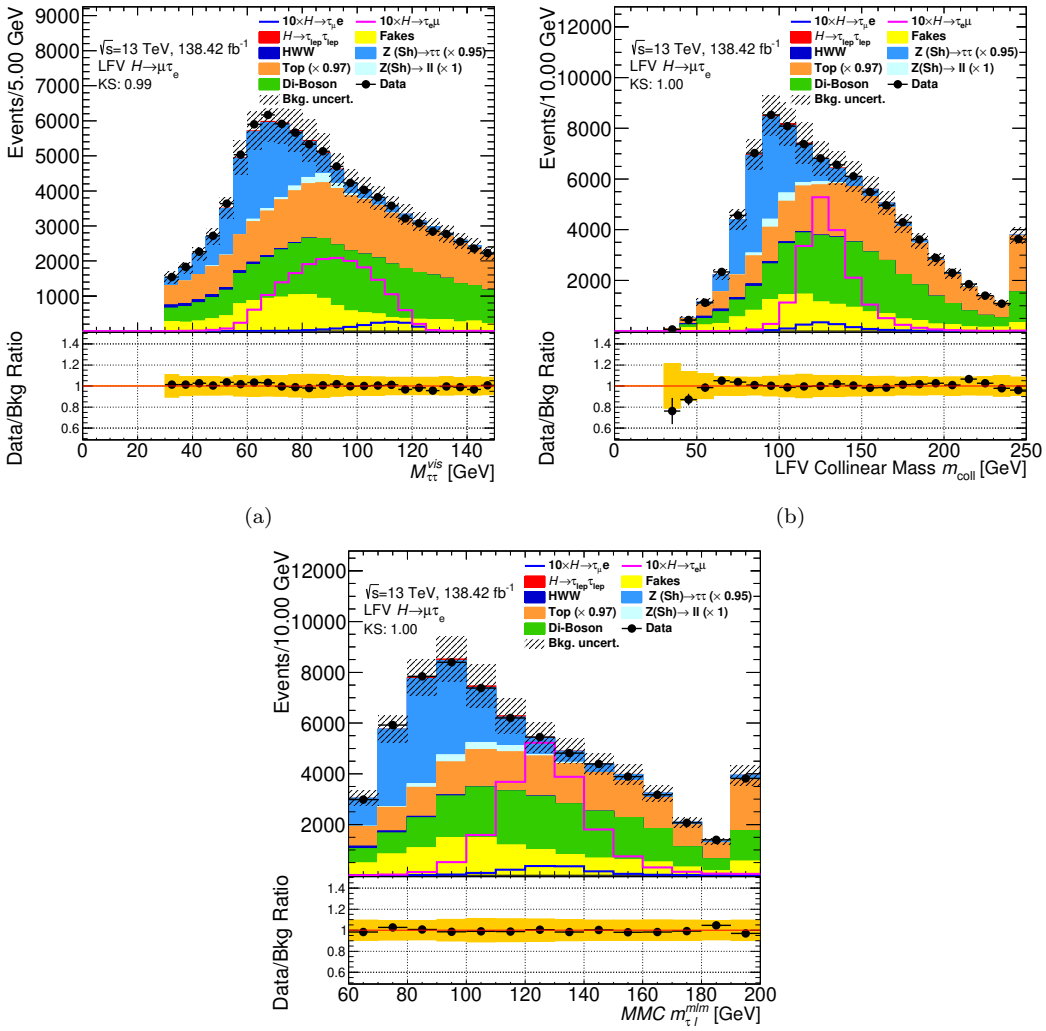


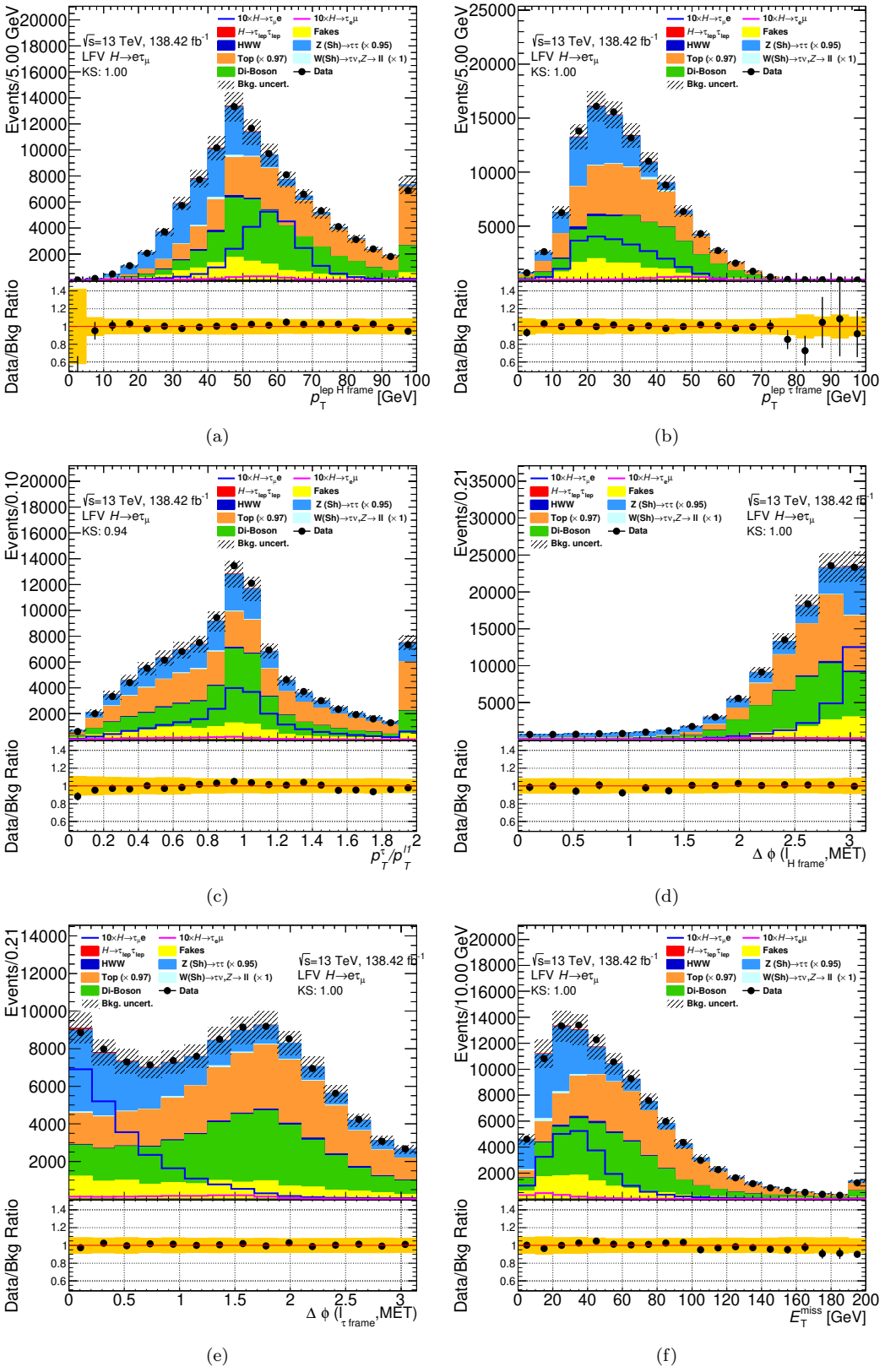
Figure 7.10: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $\mu\tau_e$ channel in the non-VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) m_{T1}, (b) m_{T2}, (c) $\Delta R\,(\ell_1, \ell_2)$, (d) $\Delta\alpha$, (e) $\sigma_{d_0}^{\ell_\tau}$, and (f) $\Delta d_0\,(\ell_1, \ell_2)$.



(a) (b)

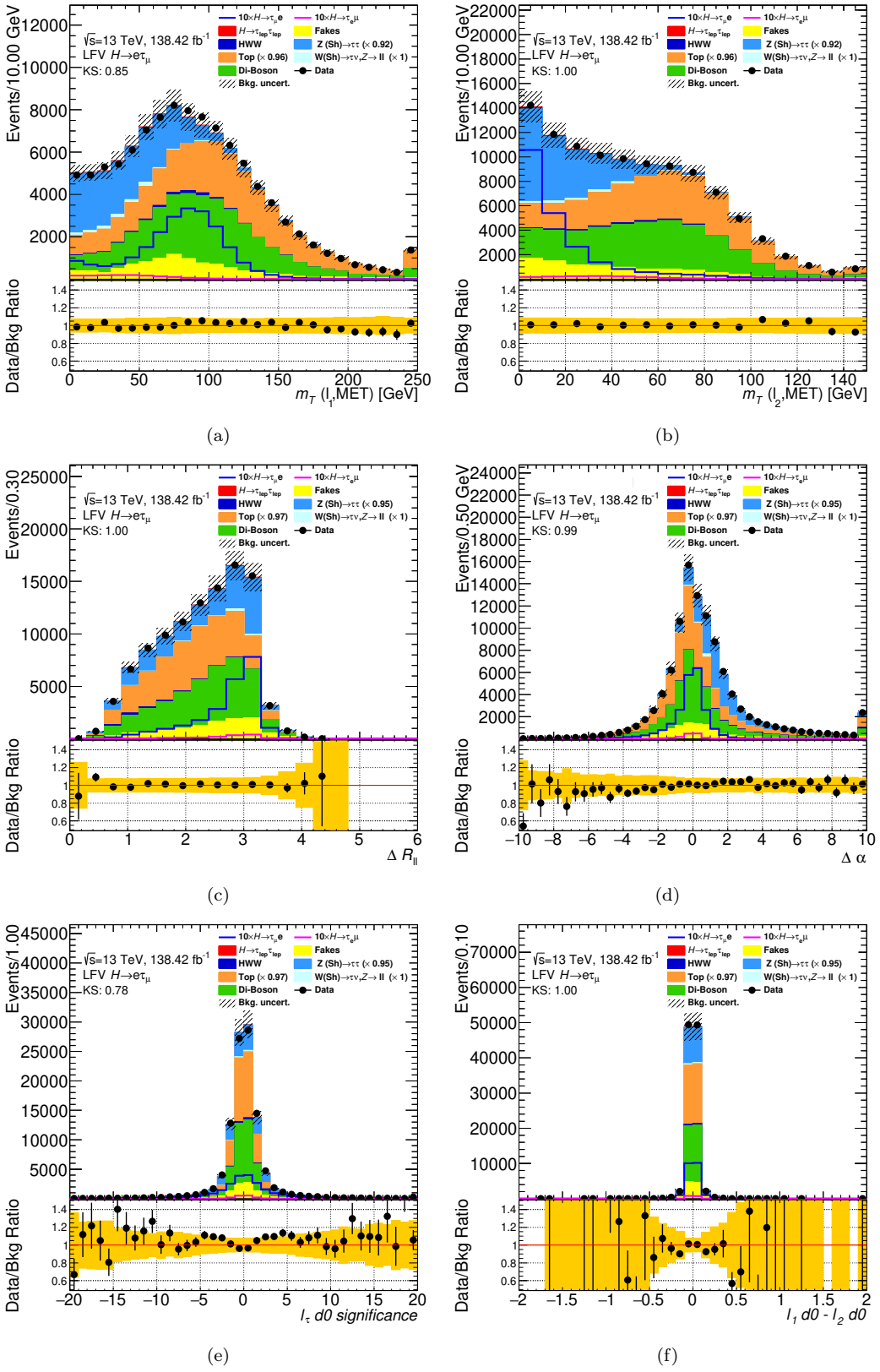
(c)

Figure 7.11: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $\mu\tau_e$ channel in the non-VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) m_{vis}, (b) $m_{\tau\tau}^{\text{coll}}$, and (c) $m_{\tau\tau}^{\text{MMC}}$.



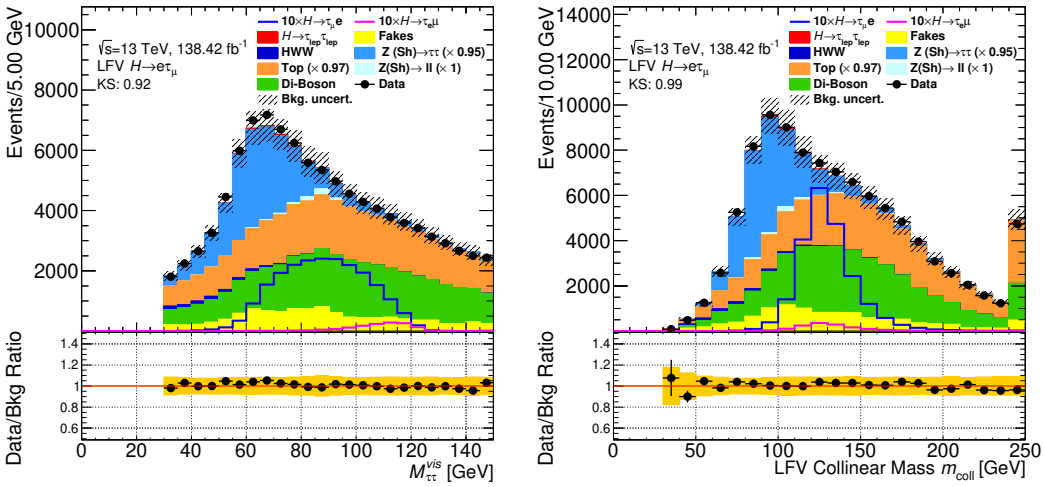
(a) (b) (c) (d) (e) (f)

Figure 7.12: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $e\tau_\mu$ channel in the non-VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) $p_{\mathrm{T}}^{\ell_H}$, (b) $p_{\mathrm{T}}^{\ell_\tau}$, (c) $\frac{p_{\mathrm{T}}^{\ell_\tau}}{p_{\mathrm{T}}^{\ell_l}}$, (d) $\Delta\phi\left(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}\right)$, (e) $\Delta\phi\left(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right)$ and (f) $E_{\mathrm{T}}^{\mathrm{miss}}$.



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

Figure 7.13: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $e\tau_\mu$ channel in the non-VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) m_{T1}, (b) m_{T2}, (c) $\Delta R(\ell_1, \ell_2)$, (d) $\Delta\alpha$, (e) $\sigma_{d_0}^{\ell_\tau}$, and (f) $\Delta d_0(\ell_1, \ell_2)$.



(a) (b)

(c)

Figure 7.14: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $e\tau_\mu$ channel in the non-VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) m_{vis}, (b) $m_{\tau\tau}^{\text{coll}}$, and (c) $m_{\tau\tau}^{\text{MMC}}$.

7.5.1.2 BDTs in the VBF category

For the VBF region, a similar strategy to that used in the non-VBF region is applied. The different jet production mechanisms in VBF signal and background events enable the MVA algorithms to exploit the kinematic properties of the two highest-p_{T} jets for enhanced background suppression. Consequently, these trainings incorporate several jet-related variables in addition to the ones used in the non-VBF trainings. These include:

- m_{jj}: invariant mass of the di-jet system.
- ΔR_{jj}: angular separation between the two leading p_{T} jets.
- $|\eta_{j_1} - \eta_{j_2}|\eta_{j_1}\eta_{j_2}$: separation in η between the two jets.
- η centrality for each lepton (the lepton from the Higgs boson and the lepton from the tau), defined relative to the η of the two jets. This variable is defined as:

$$\eta - \text{centr.} = \exp\left(\frac{-4}{\left(\eta_{j_0} - \eta_{j_1}\right)^2}\left(\eta_{\ell_i} - \frac{\eta_{j_0} + \eta_{j_1}}{2}\right)^2\right). \tag{7.3}$$

  A centrality value close to unity indicates that the lepton is positioned between the two jets, values near $1/e$ indicate alignment with one of the jets, and values below $1/e$ suggest that the lepton lies outside the jet pair

- $\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$: angular separation between the leading jet and the $E_{\mathrm{T}}^{\mathrm{miss}}$.
- $\Delta\phi(j_2, E_{\mathrm{T}}^{\mathrm{miss}})$: angular separation between the subleading jet and the $E_{\mathrm{T}}^{\mathrm{miss}}$.
- $p_{\mathrm{T}}^{\mathrm{tot}}$: magnitude of the vector sum of the transverse momenta for the two leptons, two jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$, effectively acting as a veto for additional jet activity.

The variable rankings for the VBF region BDTs have been computed as explained for the non-VBF BDTs, and are shown in Table 7.8. The optimal coefficients for the VBF BDTs were found to be $c_1 = 0.2$, $c_2 = 0.9$, and $c_3 = 0.3$.

Figures 7.15 and 7.16 show the additional variables used in the VBF BDTs in the $\mu\tau_e$ channel. The same distributions for the $e\tau_\mu$ dataset are shown in Figures 7.17 and 7.18. The agreement between data and the prediction is good within the uncertainties. The post-fit distributions of the BDT scores for the VBF category are shown in Figure 7.19c and 7.19d for the $e\tau_\mu$ and $\mu\tau_e$ datasets, respectively. The signal and background predictions correspond to the best-fit values obtained from the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals described in Section 7.8.1. The agreement between data and the signal and background predictions is good within the total uncertainties, that include statistical and systematic components.

Variable	$\mathrm{BDT}_1^{\ell\tau_{\ell'}}$		$\mathrm{BDT}_2^{\ell\tau_{\ell'}}$		$\mathrm{BDT}_3^{\ell\tau_{\ell'}}$	
	Unspecific	Specific	Unspecific	Specific	Unspecific	Specific
$m_{\tau\tau}^{\mathrm{MMC}}$	0.3685	0.0450	0.4351	0.0552	0.5915	0.0638
$m_{\tau\tau}^{\mathrm{coll}}$	0.2983	0.0399	0.3592	0.0432	0.5140	0.0576
$\Delta\phi\left(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right)$	0.1758	0.0477	0.2744	0.0572	0.0867	0.0390
m_{vis}	0.1703	0.0508	0.1863	0.0551	0.4565	0.0620
$p_{\mathrm{T}}^{\mathrm{tot}}$	0.1510	0.0471	0.1355	0.0481	0.0419	0.0362
m_{T2}	0.1007	0.0395	0.2744	0.0525	0.0702	0.0390
$p_{\mathrm{T}}^{\ell_H}$	0.0953	0.0348	0.1341	0.0268	0.3733	0.0447
$\Delta\alpha$	0.0875	0.0361	0.0695	0.0296	0.3493	0.0385
m_{jj}	0.0554	0.0342	0.0469	0.0340	0.0477	0.0319
$\lvert\eta_{j_1} - \eta_{j_2}\rvert\eta_{j_1}\eta_{j_2}$	0.0478	0.0361	0.0554	0.0391	0.0682	0.0338
$\eta - \mathrm{centr.}(\ell_H)$	0.0429	0.0501	0.0432	0.0524	0.1003	0.0454
m_{T1}	0.0426	0.0381	0.1157	0.0411	0.1502	0.0456
$\eta - \mathrm{centr.}(\ell_\tau)$	0.0423	0.0498	0.0432	0.0540	0.1097	0.0459
ΔR_{jj}	0.0394	0.0415	0.0500	0.0429	0.0570	0.0345
$\Delta R\left(\ell_1, \ell_2\right)$	0.0278	0.0551	0.0167	0.0537	0.0691	0.0671
$\Delta\phi\left(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}\right)$	0.0274	0.0389	0.0492	0.0444	0.1522	0.0362
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.0246	0.0336	0.1341	0.0333	0.0201	0.0270
$p_{\mathrm{T}}^{\ell_\tau}$	0.0205	0.0436	0.0279	0.0385	0.1124	0.0401
$\sigma_{d_0}^{\ell_\tau}$	0.0136	0.0271	0.1076	0.0338	0.0012	0.0273
$\Delta d_0\left(\ell_1, \ell_2\right)$	0.0122	0.0279	0.0624	0.0254	0.0059	0.0283
$\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$	0.0069	0.0514	0.0017	0.0528	0.0335	0.0438
$\frac{p_{\mathrm{T}}^{\ell_\tau}}{p_{\mathrm{T}}^{\ell_1}}$	0.0049	0.0378	0.0293	0.0332	0.0285	0.0309
$\Delta\phi(j_2, E_{\mathrm{T}}^{\mathrm{miss}})$	0.0040	0.0538	0.0022	0.0535	0.0006	0.0460

Table 7.8: Combined unspecific and specific ranking of variables used in the BDTs for the VBF category, showing the unspecific separation and specific separation for $\mathrm{BDT}_1^{\ell\tau_{\ell'}}$, $\mathrm{BDT}_2^{\ell\tau_{\ell'}}$, and $\mathrm{BDT}_3^{\ell\tau_{\ell'}}$.

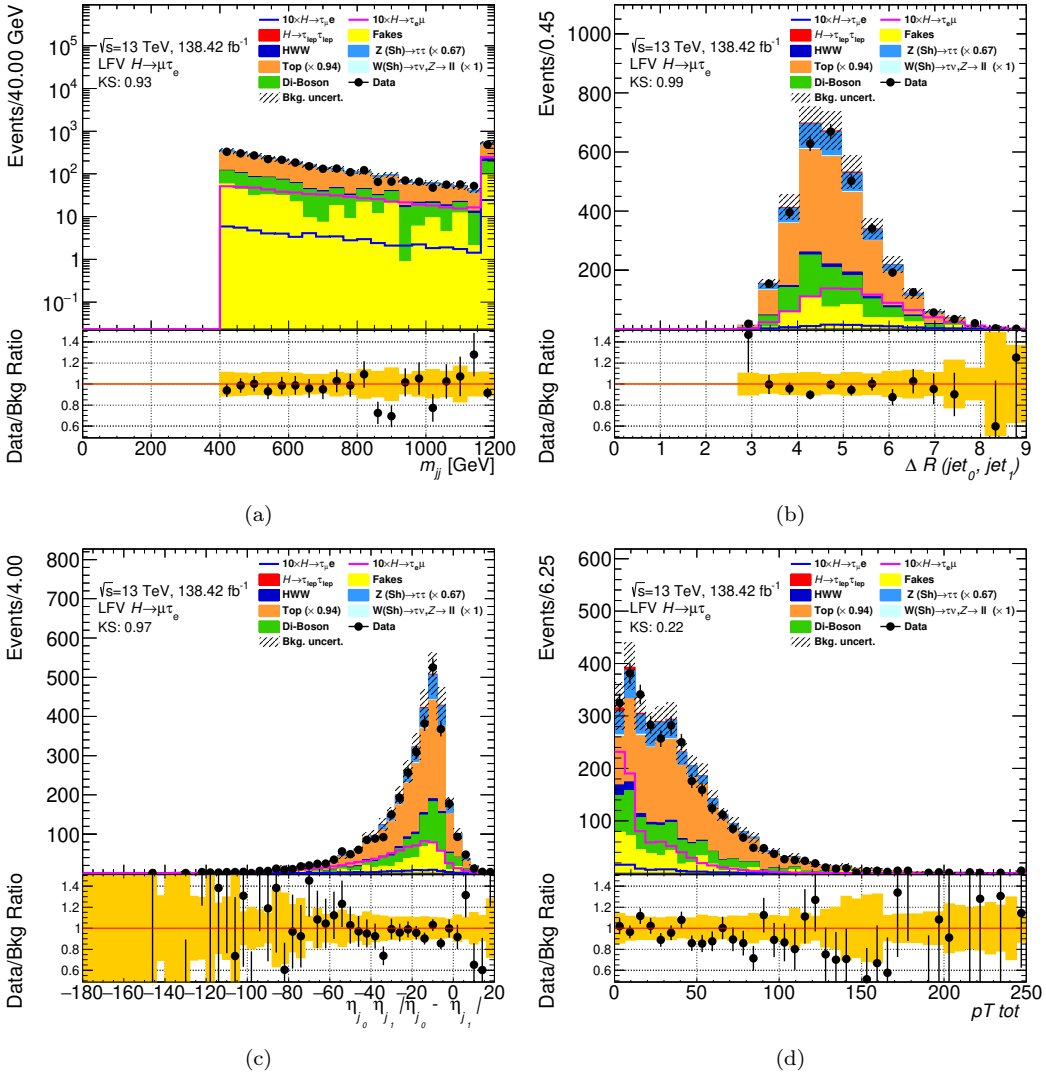


Figure 7.15: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $\mu\tau_e$ channel in the VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) m_{jj}, (b) ΔR_{jj}, (c) $|\eta_{j_1} - \eta_{j_2}|\eta_{j_1}\eta_{j_2}$, and (d) $p_{\mathrm{T}}^{\mathrm{tot}}$.

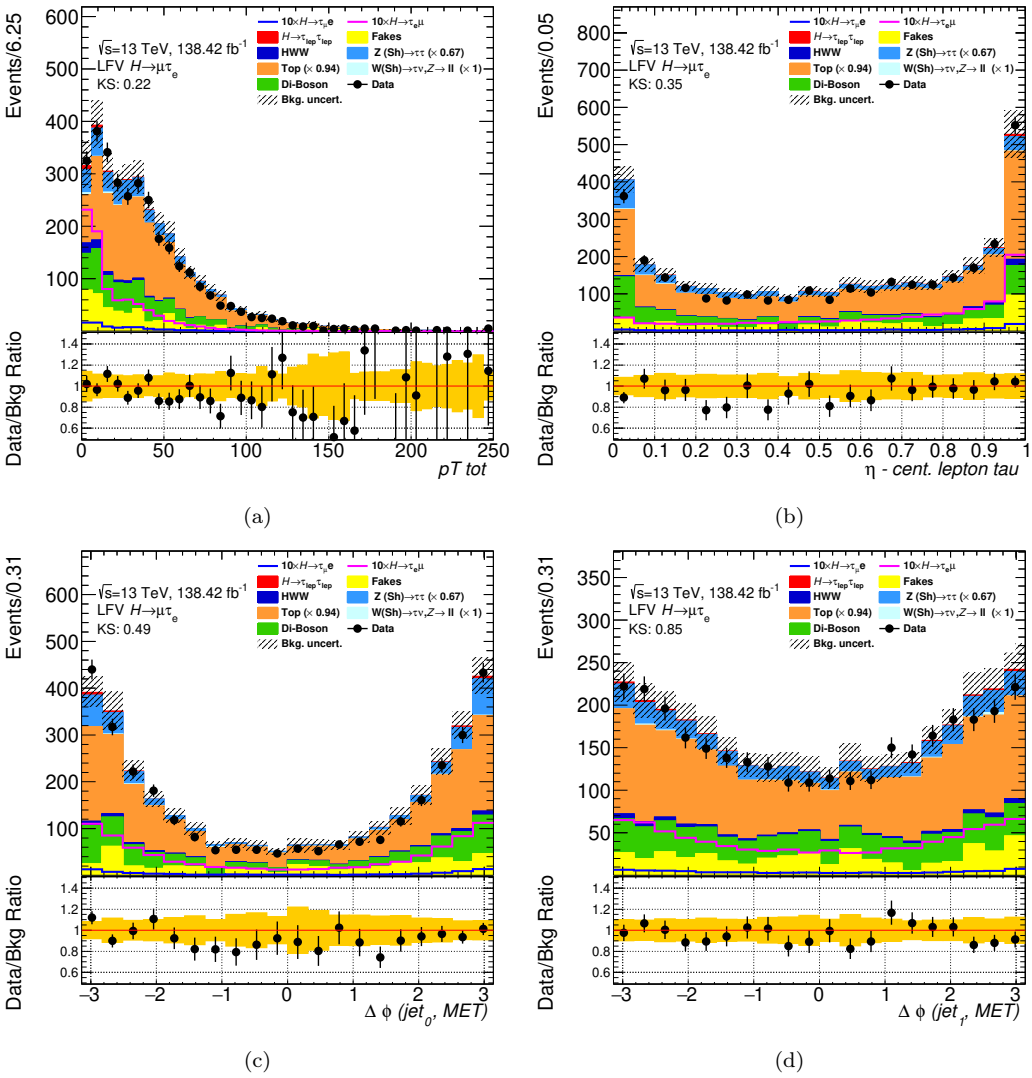


Figure 7.16: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $\mu\tau_e$ channel in the VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) η centrality for the lepton from the Higgs boson, (b) η centrality for the lepton from the tau, (c) $\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$, and (d) $\Delta\phi(j_2, E_{\mathrm{T}}^{\mathrm{miss}})$.

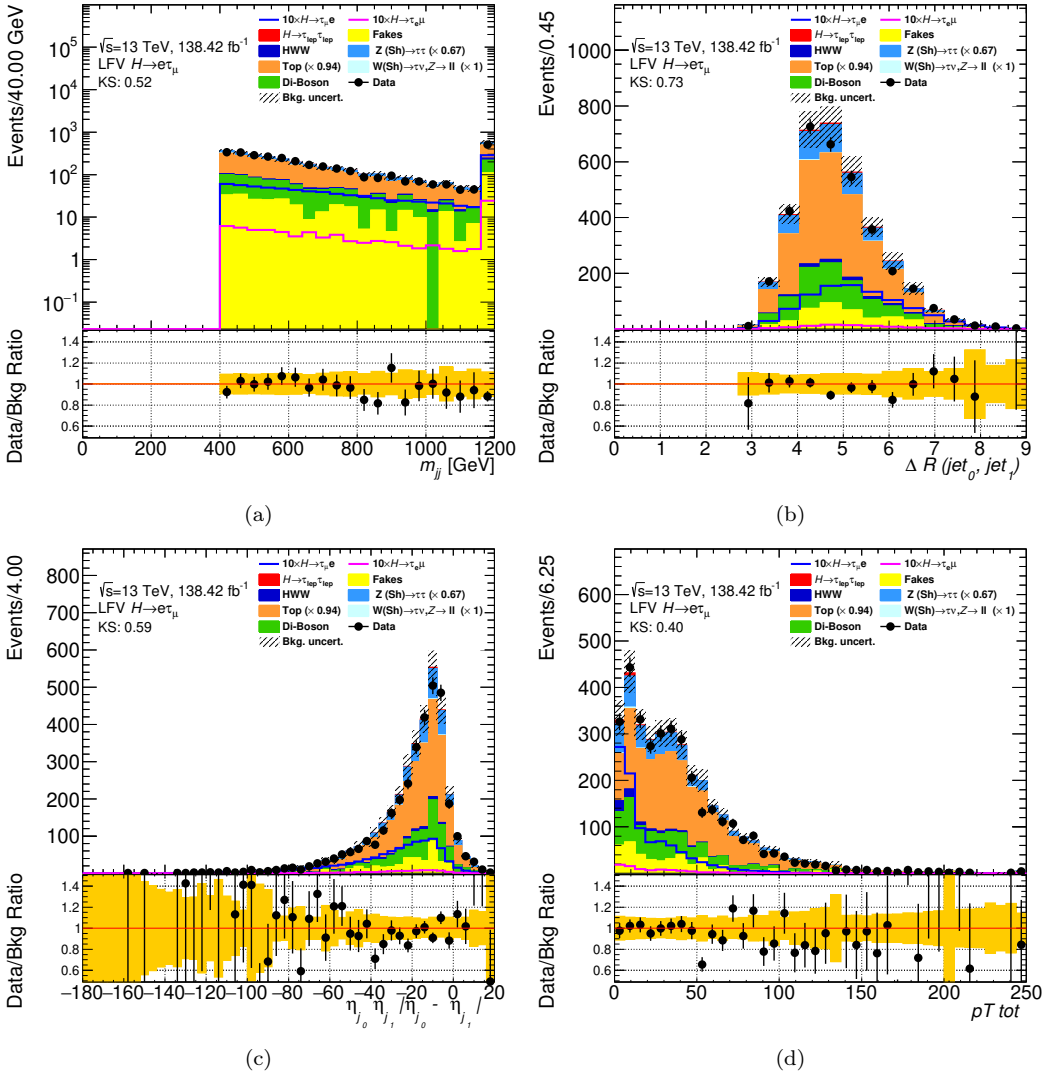


Figure 7.17: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $e\tau_\mu$ channel in the VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) m_{jj}, (b) ΔR_{jj}, (c) $|\eta_{j_1} - \eta_{j_2}|\eta_{j_1}\eta_{j_2}$, and (d) $p_{\mathrm{T}}^{\mathrm{tot}}$.

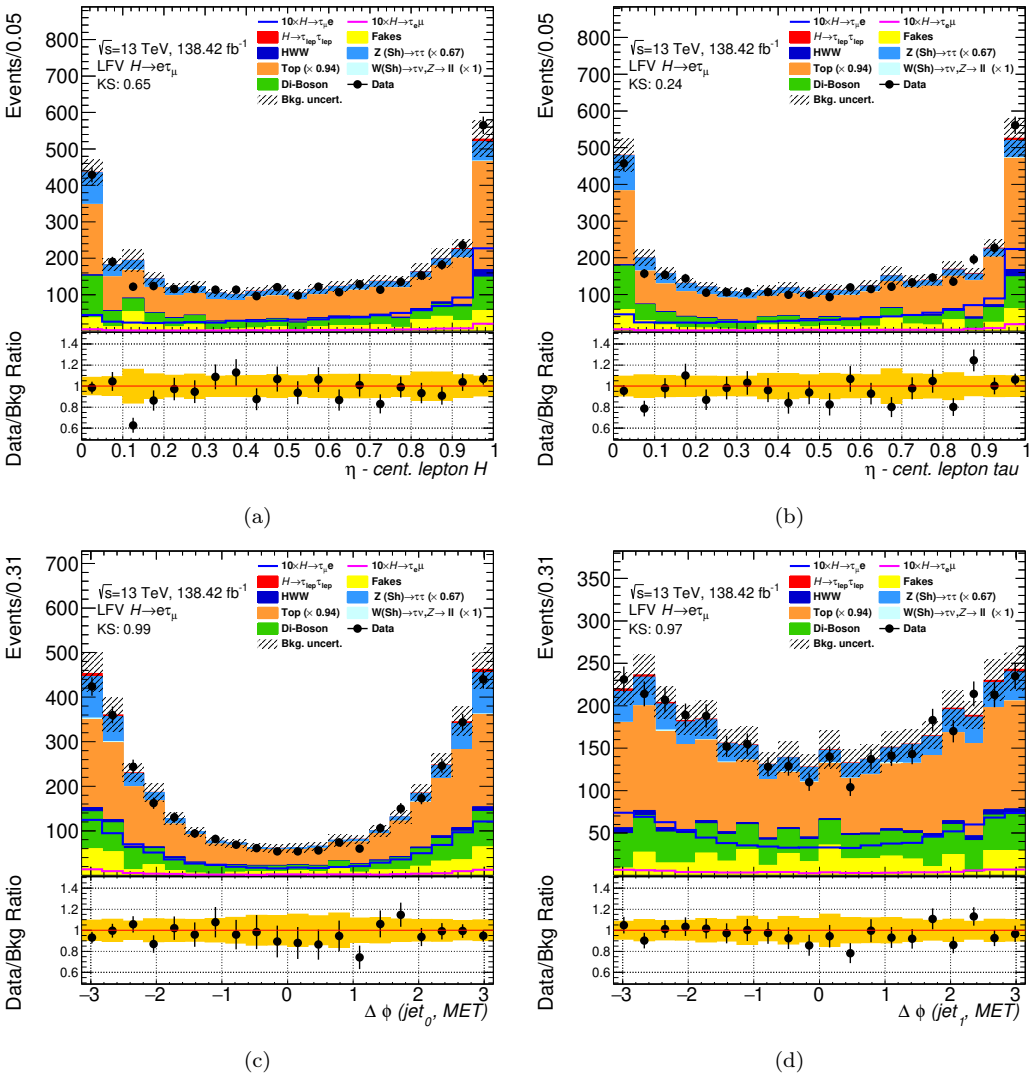


Figure 7.18: Distributions of variables used in the BDTs trained for the MC-based $\ell\tau_{\ell'}$ analysis. The distributions are shown for the $e\tau_\mu$ channel in the VBF category. The uncertainty bands include statistical uncertainties and uncertainties on the normalization of the background processes. The figures show (a) η centrality for the lepton from the Higgs boson, (b) η centrality for the lepton from the tau, (c) $\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$, and (d) $\Delta\phi(j_2, E_{\mathrm{T}}^{\mathrm{miss}})$.

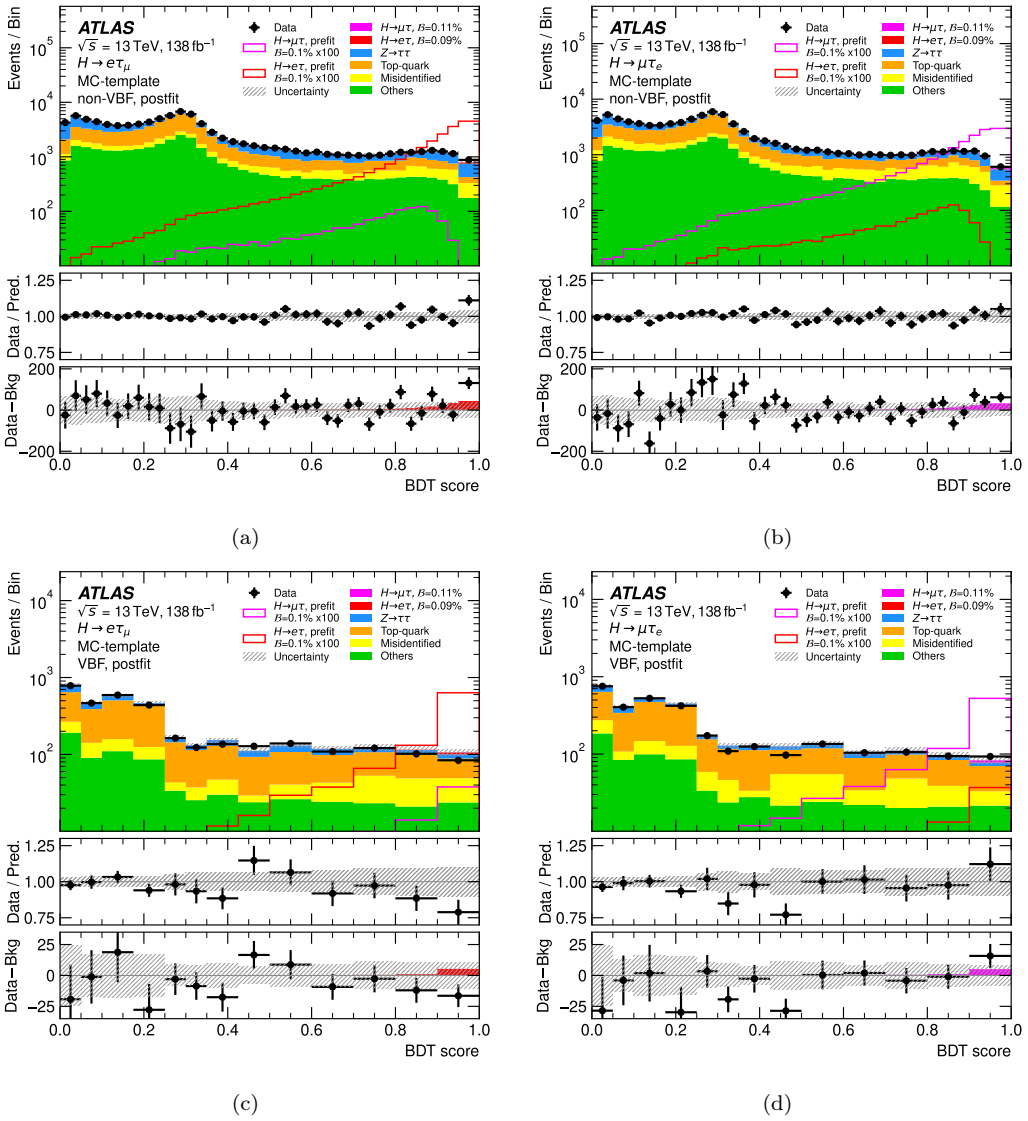


Figure 7.19: The BDT score distributions for the final states $e\tau_\mu$ (left) and $\mu\tau_e$ (right) in the non-VBF (top) and VBF (bottom) categories within the MC-based $\ell\tau_{\ell'}$ channel are presented here. The central panel shows the ratio of data to the predicted yields, which include both background and signal. The lower panel illustrates the residuals between the data and the background after performing the fit. The shaded band represents the total post-fit uncertainty associated with the overall predicted yields. Each sample's prediction is derived from the likelihood fit, applied to concurrently measure $H \to e\tau$ and $H \to \mu\tau$ signals, as detailed in Section 7.8.1. The binning is the one employed in the likelihood fit. Overlayed pre-fit signal shapes, assuming $B(H \to \ell\tau) = 0.1\%$, are scaled by a factor of 100 for visibility. Post-fit signal contributions are added to the background to the background predictions in the upper and middle panels.

7.5.2 BDTs in the MC-based $\ell\tau_{\text{had}}$ channel

In the $\ell\tau_{\text{had}}$ channel of the MC-based method, the method is similar to that used in the $\ell\tau_{\ell'}$ channel. Two BDTs are employed across all cases except for the $e\tau_{\text{had}}$ channel in the non-VBF category, where three BDTs are utilized to enhance the expected signal significance. The use of two or three BDTs, rather than a single classifier, was motivated by this improvement in significance. For instances with three BDTs, $\text{BDT}_1^{\ell\tau_{\text{had}}}$ is trained to distinguish the signal from $Z \to \tau\tau$ events, $\text{BDT}_2^{\ell\tau_{\text{had}}}$ is optimized to separate the signal from misidentified background, and $\text{BDT}_3^{\ell\tau_{\text{had}}}$ differentiates the signal from the remaining background processes. Where two BDTs are applied, $\text{BDT}_{1'}^{\ell\tau_{\text{had}}}$ targets the separation of the signal from $Z \to \tau\tau$ events, while $\text{BDT}_{2'}^{\ell\tau_{\text{had}}}$ separates the signal from all other backgrounds.

The output scores from these BDTs are subsequently combined into a single score by applying a linear weighted sum in the non-VBF categories and a quadratic weighted sum in the VBF categories. For the latter, the quadratic combination yielded a 4% improvement in expected signal significance, whereas in the non-VBF categories, the linear combination yielded the highest expected significance.

The input variables for the BDTs differ between the VBF and non-VBF categories, with jet-related variables added in the VBF category to improve signal-background separation. In the non-VBF category, the BDTs are trained on both low- and high-level kinematic variables, such as the transverse momenta of the lepton from the Higgs boson, $p_{\text{T}}^{\ell_H}$, and the visible decay products of the tau lepton, $p_{\text{T}}(\tau_{\text{had-vis}})$. Additional inputs include the transverse mass between the lepton from the Higgs boson and the $E_{\text{T}}^{\text{miss}}$, $m_{\text{T}}(\ell_H, E_{\text{T}}^{\text{miss}})$, and between the tau lepton and the $E_{\text{T}}^{\text{miss}}$, $m_{\text{T}}(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})$. Angular information—such as the η separation between the lepton and $\tau_{\text{had-vis}}$, $\Delta\eta(\ell_H, \tau_{\text{had-vis}})$; the azimuthal separation, $\Delta\phi(\ell_H, \tau_{\text{had-vis}})$; and the angular separation, $\Delta R(\ell_H, \tau_{\text{had-vis}})$—is also included. Furthermore, mass-related variables like the visible mass, m_{vis}, and collinear mass, m_{coll}, are used. The list is completed with the η of the $\tau_{\text{had-vis}}$ and the difference between the azimuthal separations of the lepton and $\tau_{\text{had-vis}}$ with the $E_{\text{T}}^{\text{miss}}$, $|\Delta\phi(\ell_H, E_{\text{T}}^{\text{miss}})| - |\Delta\phi(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})|$.

In the VBF category, while the variables $\Delta\phi(\ell_H, \tau_{\text{had-vis}})$ and $|\Delta\phi(\ell_H, E_{\text{T}}^{\text{miss}})| - |\Delta\phi(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})|$ are excluded, jet-related variables are included instead. These additional inputs are the invariant mass of the di-jet system, m_{jj}; the number of jets in the event with a p_{T} above 30 GeV, N_{jets}^{30}; and the absolute η separation between the two leading jets, $|\Delta\eta_{\text{jj}}|$. The azimuthal separation between the lepton and the $E_{\text{T}}^{\text{miss}}$, $\Delta\phi(\ell_H, E_{\text{T}}^{\text{miss}})$, is also included.

The BDT distributions for the MC-based $\ell\tau_{\text{had}}$ channel are shown in Figure 7.20. The agreement between data and the prediction is good within the total uncertainties. The predicted signal and background distributions correspond to the post-fit values obtained from the simultaneous fit of the $H \to e\tau$ and $H \to \mu\tau$ signals described in Section 7.8.1.

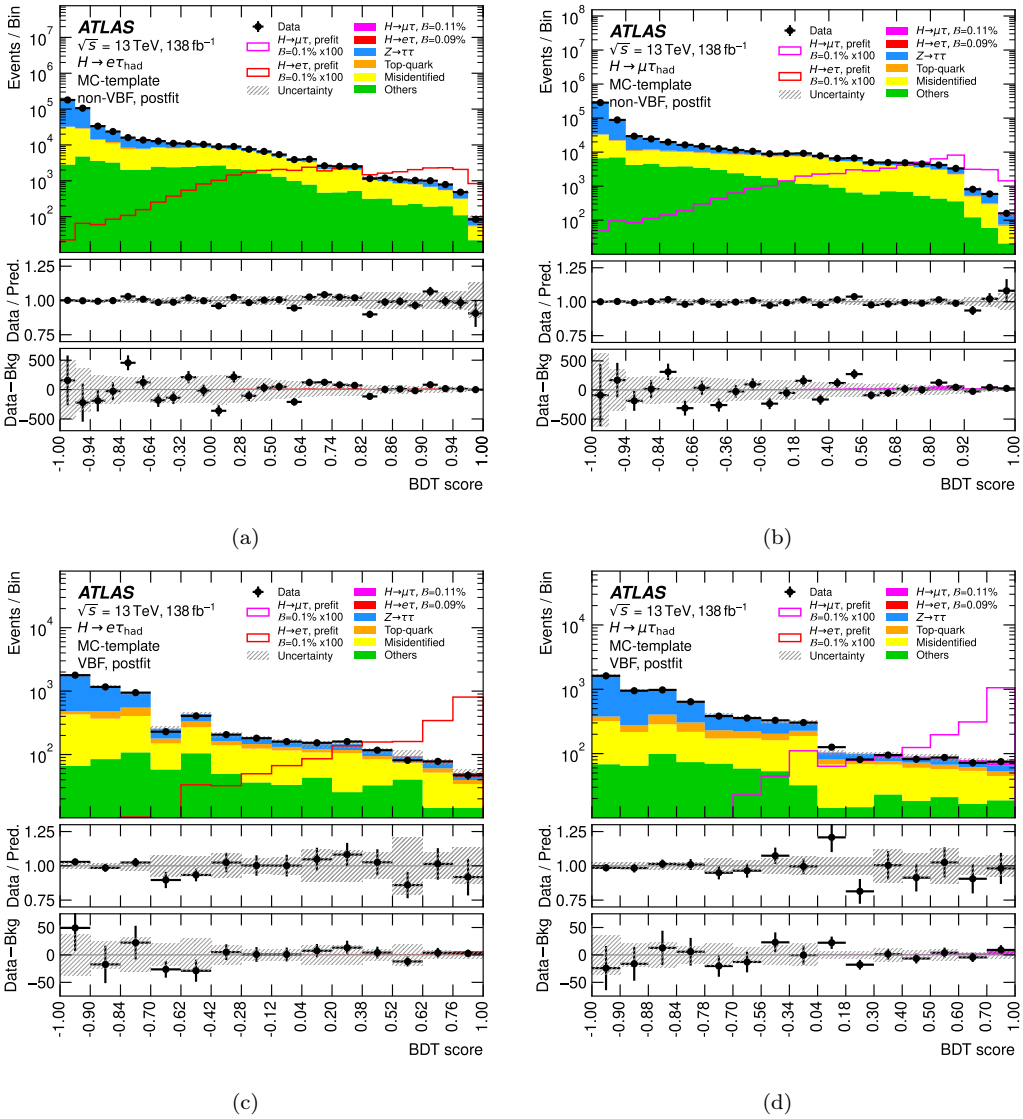


Figure 7.20: The BDT score distributions for the final states $e\tau_{\text{had}}$ (left) and $\mu\tau_{\text{had}}$ (right) in the non-VBF (top) and VBF (bottom) categories within the MC-based $\ell\tau_{\text{had}}$ channel are presented here. The central panel shows the ratio of data to the predicted yields, which include both background and signal. The lower panel illustrates the residuals between the data and the background after performing the fit. The shaded band represents the total post-fit uncertainty associated with the overall predicted yields. Each sample's prediction is derived from the likelihood fit, applied to concurrently measure $H \to e\tau$ and $H \to \mu\tau$ signals, as detailed in Section 7.8.1. The binning is the one employed in the likelihood fit. Overlayed pre-fit signal shapes, assuming $B(H \to \ell\tau) = 0.1\%$, are scaled by a factor of 100 for visibility. Post-fit signal contributions are added to the background predictions in the upper and middle panels.

7.6 Systematic uncertainties

The systematic uncertainties affecting the LFV $H \to e\tau$ and $H \to \mu\tau$ analyses can be categorized into two groups: theoretical uncertainties on signal and background

processes, and experimental uncertainties. The former group includes PDF uncertainties, QCD scale uncertainties, and uncertainties related to the parton showering and hadronization models, as well as uncertainties due to the choice of the MC generator. The latter group includes uncertainties related to the reconstruction and identification of leptons and taus, jet energy scale and resolution, $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, flavor tagging or the prediction of the misidentified background, among others. The systematic uncertainties are treated as nuisance parameters in the likelihood function, as will be described in Section 7.7. The following sections provide a description of the systematic uncertainty sources considered in the LFV $H \to e\tau$ and $H \to \mu\tau$ analyses.

7.6.1 Theoretical uncertainties on the Higgs boson production

The uncertainties on Higgs boson production impact both the signal and the $H \to \tau\tau$ and $H \to WW$ background processes. These uncertainties are treated as uncorrelated across different Higgs production modes. In each mode, the theoretical uncertainties originate from three primary sources: variations in the renormalization and factorization QCD scales, which assess the impact of missing higher-order corrections; non-perturbative effects in the cross-section calculations, such as parton showering, hadronization, and underlying event models; and uncertainties from the PDFs and the value of the strong coupling constant, α_s.

For ggF production, the QCD scale uncertainties are represented by nine sources, which account for variations in the renormalization and factorization scales, resummation scales, and jet multiplicity transitions, specifically for events with 0 to 1 jet and 1 to 2 jets. Shape uncertainties are applied to the cross-section prediction as a function of p_{T}^{H}, covering migrations between the bins $[0, 60, \infty)$ GeV and $[0, 120, \infty)$ GeV. Finally, the ggF cross-section dependence on the chosen top-quark mass is also taken into account.

For the VBF, VH, and $t\bar{t}H$ production modes, the QCD scale uncertainties are estimated by varying the renormalization and factorization scales by factors of 2 and 1/2 around their nominal values, with the constraint $1/2 \leq \mu_R/\mu_F \leq 2$. The envelope of these variations is taken as uncertainty.

Uncertainties in parton showering and hadronization models follow the guidelines outlined in Section 6.7. Specifically, nominal predictions simulated with Pythia8 are compared against those generated with Herwig7. To evaluate uncertainties in matrix element calculations, nominal predictions from Powheg are compared with those from MadGraph5_aMC@NLO. These uncertainties are considered uncorrelated across different Higgs production modes.

The PDF uncertainties are evaluated using the 30 eigenvector variations of the NNPDF3.0 set, allowing for reweighting of events to assess the impact on predictions. Variations in α_s are introduced around its central value to determine

α_s uncertainties. PDF and α_s uncertainties are treated as correlated across all Higgs production modes.

7.6.2 Theoretical uncertainties on the $Z \to \tau\tau$, $Z \to \ell\ell$ and diboson processes

The theoretical uncertainties affecting the $Z \to \tau\tau$ and $Z \to \ell\ell$ background predictions are derived from several key sources. First, the uncertainties on the PDFs are evaluated using 100 variations of the NNPDF3.0NNLO set. The envelope of these variations is taken as the PDF uncertainty. The uncertainties in the renormalization and factorization scales (μ_R/μ_F) are assessed using event weights provided by SHERPA. Additional uncertainties related to jet-to-parton matching, labeled as CKKW uncertainties, are estimated using a truth-level parametrization dependent on jet multiplicity and $p_{\mathrm{T}}(Z)$. Similarly, resummation scale uncertainties, denoted as QSF, are evaluated using a truth-level parametrization based on jet multiplicity and $p_{\mathrm{T}}(Z)$. The α_S uncertainty is also assessed using event weights generated by SHERPA. Lastly, the alternative PDF uncertainty is estimated by comparing predictions using the nominal NNPDF3.0NNLO PDF set against those from the MMHT2014NNLO68CL and CT14NNLO PDF sets.

For the Top background, these theoretical uncertainties are considered only when a normalization factor is not applied to this background, specifically in fits excluding the $\ell\tau_{\ell'}$ channel and its dedicated Top CR. The PDF uncertainties for this background are constructed by combining the available PDF sets with 30 eigenvector variations. Matrix element uncertainties are evaluated by comparing the nominal samples with alternative samples produced using different matrix-element generators while maintaining the same parton shower model. Parton shower uncertainties are similarly constructed by comparing the nominal samples with alternative samples produced using the same matrix-element generator but with a different parton shower model. Additionally, uncertainties associated with final-state and initial-state radiation are estimated by reweighting the event weights of the nominal samples.

7.6.3 Experimental uncertainties

The experimental uncertainties considered in the LFV $H \to e\tau$ and $H \to \mu\tau$ searches closely follow those considered in the $H \to \tau\tau$ analysis and that were described in detail in Section 6.7.3. A brief summary of the experimental uncertainties affecting the LFV $H \to e\tau$ and $H \to \mu\tau$ searches is provided here.

Experimental systematic uncertainties cover contributions from trigger, reconstruction, identification, and isolation efficiencies associated with each object type used in the analysis, as well as uncertainties in luminosity determination and misidentified background estimates.

Muon systematic uncertainties include those affecting the muon momentum scale and resolution, along with identification, reconstruction, isolation, and trigger efficiencies, following the guidelines in Ref. [241].

For electrons, uncertainties include variations in the energy scale and resolution, as well as identification, reconstruction, isolation, and trigger efficiencies, based on recommendations in Ref. [204].

Systematic uncertainties on jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ include contributions from jet energy scale and resolution, jet vertex tagger efficiency, and flavor composition impacts on the jet energy scale. These include also uncertainties on flavor-tagging efficiencies and are propagated into the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. Additionally, uncertainties related to the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term are considered, detailed in Ref. [211, 218]. In alignment with the $H \to \tau\tau$ analysis, the jet energy scale and resolution uncertainties are particularly significant due to their impact on the MMC algorithm.

For τ-leptons, uncertainties include those on the tau energy scale and on tau identification, reconstruction, and isolation efficiencies, as well as the tau trigger efficiency. Uncertainties associated with the eBDT, used to veto electrons misidentified as taus, are also considered, as detailed in Ref. [217].

The integrated luminosity uncertainty is assigned a value of 1.7%, based on the preliminary recommendation in Ref. [242]. Although more recent recommendations exist, this value reflects the available information at the time of analysis.

Two specific uncertainties are applied to the normalization of the $Z \to \ell\ell$ misidentified background. In the $\ell\tau_{\ell'}$ channels, contributions from muons misidentified as electrons are negligible except in the $Z \to \mu\mu$ process. For this, a dedicated CR, described in Section 7.4.1.4, is used to assess discrepancies between data and MC predictions. The systematic uncertainty on the $Z \to \mu\mu$ background is defined by the full difference between data and MC in this CR, with the MC prediction scaled to match data estimates. The electron-to-muon misidentification probability is negligible. As a result, there is no dedicated CR for the $Z \to ee$ process.

In the $\ell\tau_{\mathrm{had}}$ channels, a dedicated validation region is used to derive a data-based normalization uncertainty for the $Z \to \ell\ell$ background in the $\mu\tau_{\mathrm{had}}$ channel. Here, the $Z \to \ell\ell$ misidentification background, predominantly arising from muons misidentified as $\tau_{\text{had-vis}}$, is more accurately represented by data, with the full discrepancy between the validation region and data used as the systematic uncertainty. Electron-to-$\tau_{\text{had-vis}}$ misidentification backgrounds are largely suppressed by the eBDT veto and a cut on $m_{\mathrm{vis}}(e,\tau)$ in the non-VBF region.

7.6.3.1 Uncertainties on the misidentified background estimation

The uncertainties on the misidentified background uncertainties are derived separately for the MC-based $\ell\tau_{\ell'}$, MC-based $\ell\tau_{\mathrm{had}}$, and Symmetry-based $\ell\tau_{\ell'}$ methods. These uncertainties are treated as uncorrelated across the different methods.

In the MC-based $\ell\tau_{\ell'}$ method, the uncertainty evaluation includes five primary sources. The statistical uncertainties are divided into four independent categories, based on whether events are selected by single-lepton or dilepton triggers and whether they satisfy or fail the b-veto criterion. Systematic uncertainties are additionally included, addressing several sources of potential biases. For prompt-lepton subtraction in the same-sign region, an uncertainty of 6–8% is applied. Another systematic source considers jet flavor discrepancies between control regions enriched in misidentified backgrounds and the signal regions. To quantify this, transfer factors are derived in CRs rich in misidentified backgrounds using simulated samples, primarily from W+jets and $V\gamma$ processes. These transfer factors are then applied to events satisfying baseline selection criteria but requiring same-sign leptons, and the resultant BDT distribution is compared with events selected for opposite-sign leptons. Differences observed in these distributions are used to estimate a systematic uncertainty, which ranges between 10% and 30%. Additionally, a non-closure uncertainty addresses residual mismatches between simulated and actual data. This uncertainty is established through alternative CRs where the d_0 requirements are varied. In these CRs, the transfer factors and background predictions are recalculated, and any differences with the nominal misidentified background prediction in taken as an uncertainty, varying from 2% to 12%.

For the MC-based $\ell\tau_{\text{had}}$ channel, the uncertainties include those derived from the Fake-Factor and R_{QCD} methods, which are treated independently within the statistical model. In the VBF category, these uncertainties fall between 2% and 3%, while they are negligible in the non-VBF category, under 0.4%. A systematic uncertainty is added to cover residual discrepancies observed in CRs containing same-sign misidentified backgrounds. Defined by requiring that the τ_{had} and ℓ possess identical charges while meeting SR criteria, the differences observed in this region, between 2% and 10% depending on the category, are symmetrized before being incorporated as a systematic uncertainty.

The Symmetry-based $\ell\tau_{\ell'}$ approach calculates uncertainties using a combination of statistical and systematic components within the Fake-Factor framework. Statistical uncertainties, computed separately for each Fake-Factor bin, include those from prompt-lepton subtractions (mainly from Z+jets CR) and are propagate to the Fake-Factor values. Systematic uncertainties, on the other hand, include uncertainties from subtracting real lepton processes, particularly WZ and ZZ. Additional contributions come from uncertainties in correction factors, due to limited statistics in both the Z+jets and baseline regions. In the statistical model, the statistical uncertainties for Fake-Factor or correction bins are treated as uncorrelated, whereas the systematic uncertainties are correlated across bins. Similarly, uncertainties related to MC-based subtractions are correlated for electron and muon backgrounds.

7.7 Statistical analysis

The statistical analysis employs a likelihood function, $\mathcal{L}(\mu, \theta)$, constructed as a product of Poisson terms for each bin contributing to the search. This includes the MVA score distributions across all SRs and, where applicable, event yields in the CRs of the $\ell\tau_{\ell'}$ MC-based method. The CRs, when included, help constrain the normalization of the major simulated backgrounds, particularly those from top-quark and $Z \to \tau\tau$ processes. The POIs in the likelihood function, $\mu_{e\tau}$ and $\mu_{\mu\tau}$, denote the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, respectively. In addition, nuisance parameters θ account for systematic uncertainties affecting both signal and background contributions. These nuisance terms are modeled using Gaussian constraints. The details on the profile likelihood method used to extract the branching ratios can be found in Appendix A.

To detect any potential signal in the data, the likelihood function is fitted, and the values of the POIs are extracted using the profile-likelihood-ratio test statistic $\tilde{q}_\mu$ [240]. If no significant signal is detected, the method then uses $\tilde{q}_\mu$ and the $\mathrm{CL_s}$ method [243] to place upper limits on the branching ratios.

The analysis is divided into three separate statistical approaches, each defining the POIs differently and incorporating inputs from the different background estimation methods:

- For simultaneous searches of $H \to e\tau$ and $H \to \mu\tau$ signals, both POIs ($\mu_{e\tau}$ and $\mu_{\mu\tau}$) are fitted together across the combined $e\tau_\mu$, $e\tau_{\mathrm{had}}$, $\mu\tau_e$, and $\mu\tau_{\mathrm{had}}$ final states (see Section 7.8.1). In this case, the setup is referred to as the 2-POI fit.

- The $H \to e\tau$ search is performed with a single POI, $\mu_{e\tau}$, estimated by combining the $e\tau_\mu$ and $e\tau_{\mathrm{had}}$ final states (described in Appendix C.4.1), while $\mathcal{B}(H \to \mu\tau)$ is fixed at zero. This setup is referred to as the $e\tau$ 1-POI fit.

- The $H \to \mu\tau$ search, also using a single POI, $\mu_{\mu\tau}$, combines the $\mu\tau_e$ and $\mu\tau_{\mathrm{had}}$ final states (see Appendix C.4.1), with $\mathcal{B}(H \to e\tau)$ set to zero. This configuration is referred to as the $\mu\tau$ 1-POI fit.

Each 1-POI fit incorporates the non-VBF and VBF categories of the $\ell\tau_{\mathrm{had}}$ channel, the non-VBF category of the $\ell\tau_{\ell'}$ channel from the MC-based method, and the VBF category of the $\ell\tau_{\ell'}$ channel from the Symmetry-based method. The results from the 1-POI fits are shown in Appendix C, since they use inputs from the Symmetry-based search. For simultaneous $H \to e\tau$ and $H \to \mu\tau$ measurements, the MC-based method is used to perform a combined fit across the $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$ final states. The specific SRs and CRs applied in each combined fit setup are presented in Figures 7.21. The combination of these channels and categories is designed to maximize the expected sensitivity to the LFV signals. In any fit that includes an SR from the $\ell\tau_{\ell'}$ MC-based channel, corresponding CRs for top-quark and $Z \to \tau\tau$ backgrounds are also included.

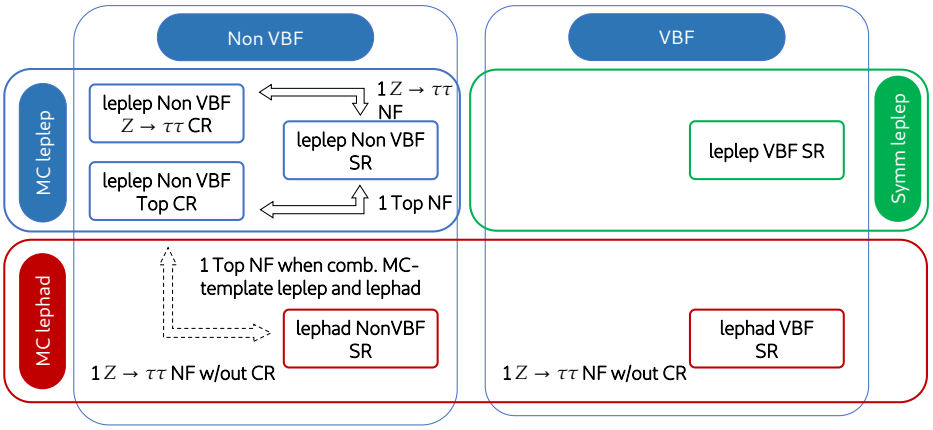


Figure 7.21: Schema showing the analysis categories used in the 1-POI fits, including the different normalization factors that are included in the fit.

To complement these analyses, standalone fits of individual channels and categories are performed.

The main result from the analysis is the combined measurement of the $H \to e\tau$ and $H \to \mu\tau$ signals, as described in Section 7.8.1. This fit incorporates all the signal and control regions defined for the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels of the MC-based method. The setup is shown in Figure 7.22.

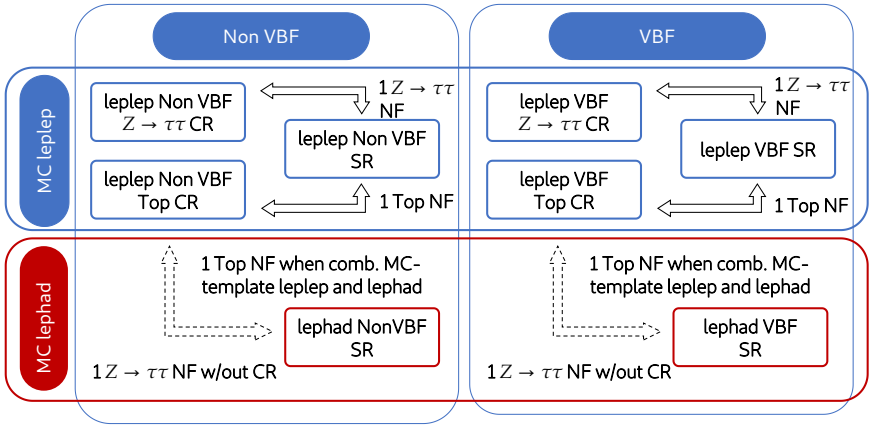


Figure 7.22: Schema showing the analysis categories used in the 2-POI fit, including the different normalization factors that are included in the fit.

To mitigate fluctuations in systematic uncertainties arising from limited MC sample sizes, several modifications are made. For electron, muon, tau, and jet systematic uncertainties where both upward and downward variations are available, a symmetrization technique is implemented. In this approach, the bin-by-bin

average variation, calculated as (up − down)/2, is taken as the upward variation and subsequently mirrored to define the downward variation. For remaining systematic uncertainties, corrections are only applied to bins where both variations go in the direction relative to the nominal value. In such cases, the larger of the two variations is mirrored to maintain consistency. When only one variation is provided, the downward variation is generated as the mirrored counterpart. Once these adjustments are made, smoothing is performed.

After smoothing, pruning is applied to the nuisance parameters in each analysis region. Nuisance parameters are pruned when normalization effects from the up or down variations differ by less than 0.1% relative to the nominal value, or when the difference between the up and down shape-only histograms and the nominal histogram remains below 0.1% across all bins. Furthermore, bin-by-bin statistical uncertainties are neglected when they are below 1% of the bin content.

The binning of the BDT or NN distributions, which serve as the final discriminants in the fit, is initially based on finely binned histograms with equidistant bins. Bins are merged iteratively to improve stability and reduce statistical uncertainties. Specifically, any bin is combined with its preceding bin if the statistical uncertainty of any background template (including misidentified events, $Z \to \tau\tau$, $\bar{t}t$, SM Higgs and diboson) is greater than 100% or if the total background yield within that bin falls below 5 events.

7.8 Results

7.8.1 Combined measurement of $H \to e\tau$ and $H \to \mu\tau$

The MC-based method enables a simultaneous measurement of the $H \to e\tau$ and $H \to \mu\tau$ signal branching ratios without requiring the assumption of an absent $H \to \mu\tau$ signal when fitting $\mathcal{B}(H \to e\tau)$ and vice versa. This approach estimates the two POIs, $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, by performing a combined fit across the $e\tau_\mu$, $e\tau_{\text{had}}$, $\mu\tau_e$, and $\mu\tau_{\text{had}}$ final states. The MC-based method is used to perform the combined fit, incorporating the SRs and CRs defined for the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels, both for the non-VBF and VBF categories. The fit not only allows simultaneous determination of the two signals but also constrains the nuisance parameters by utilizing the $e\tau$ and $\mu\tau$ datasets. In addition, the fit is able to take into account any cross-contamination of the $H \to e\tau$ signal in the $\mu\tau$ dataset, and vice versa.

The analysis framework includes eight SRs and eight CRs, with each $\ell\tau_{\ell'}$ SR paired with a top-quark and a $Z \to \tau\tau$ CR. The normalization factors for the $Z \to \tau\tau$ and top-quark processes are shared across SRs and CRs but remain uncorrelated between the VBF and non-VBF categories. Additionally, the top-quark normalization factors are unified across the $e\tau_\mu$, $\mu\tau_e$, $e\tau_{\text{had}}$, and $\mu\tau_{\text{had}}$ channels, while the $Z \to \tau\tau$ normalization factors are independently assigned to $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ final states. As explained before, there is no dedicated CR for the top-quark

background in the $\ell\tau_{\text{had}}$ channel, and the normalization factor is shared with the $\ell\tau_{\ell'}$ channel. There are also no dedicated CRs for the $Z \to \tau\tau$ background in the $\ell\tau_{\text{had}}$ channel, and in this case the normalization factors are determined solely with data from the SRs. Overall, the fit utilizes two normalization factors for the top-quark processes and four normalization factors for the $Z \to \tau\tau$ processes.

7.8.1.1 Normalization factors for the background processes and nuisance parameters

The signal strengths for the $H \to e\tau$ and $H \to \mu\tau$ processes and the normalization factors for the top-quark and $Z \to \tau\tau$ backgrounds are determined simultaneously in the 2-POI fit. Their best-fit values are presented in Figure 7.23, indicating their total uncertainties, including both statistical and systematic components. The normalization factors for the main backgrounds are compatible with the unity with the exception of the $Z \to \tau\tau$ background in the $\ell\tau_{\ell'}$ VBF category. This shows that the MC simulation is able to reproduce the data well, and that the normalization factors are consistent with the MC normalization. The signal strengths are discussed later in Section 7.8.1.2.

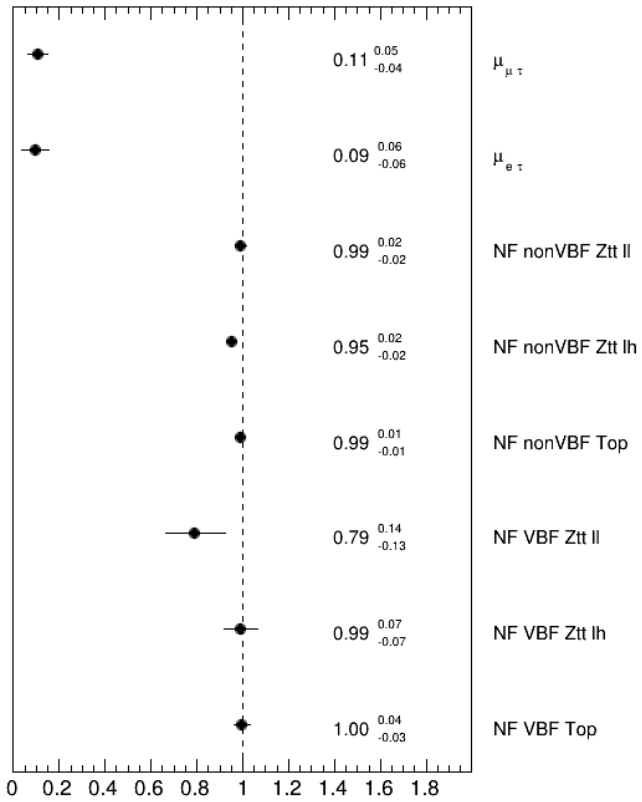


Figure 7.23: Normalization factors for the top-quark and $Z \to \tau\tau$ backgrounds and signal strengths for the $H \to e\tau$ and $H \to \mu\tau$ processes, as determined in the 2-POI fit. The error bars represent the total uncertainties, including both statistical and systematic components.

The best-fit values of the nuisance parameters in the 2-POI fit setup are shown in Figures 7.24 and 7.25. The nuisance parameters are organized by the categories of systematic uncertainties. Large constraints are visible for the flavor-tagging efficiency uncertainties, due to the constraining power provided by the top-quark CRs. NPs related to the misidentified background estimation are also constrained, since these uncertainties are very conservative at pre-fit level. Several NPs related to the JER are pulled below the 1σ level. As in the $H \to \tau\tau$ analysis, the JER impacts the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, affecting the MMC reconstruction. The $m_{\tau\tau}^{\mathrm{MMC}}$ is a relevant variable in the MVA algorithms used in the analysis, and the analysis is sensitive to the JER.

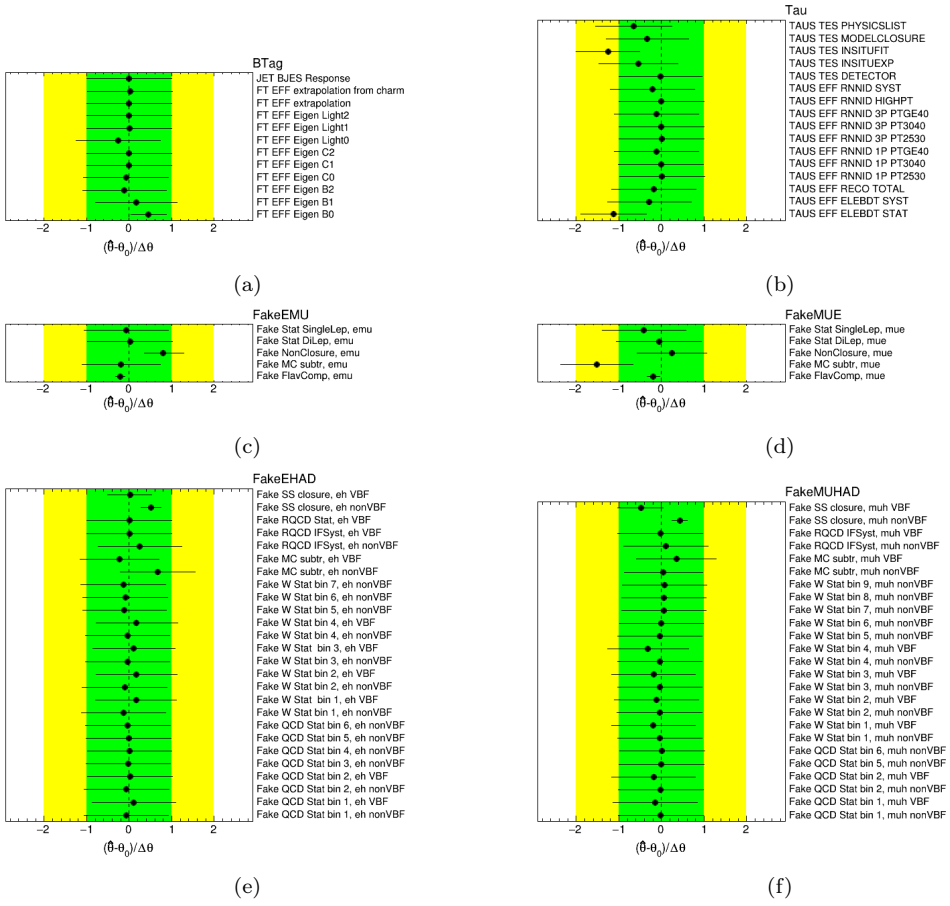


Figure 7.24: Best-fit values of the nuisance parameters in the 2-POI fit, organized by the different sources of systematic uncertainties: (a) flavor-tagging, (b) τ-leptons, (c) misidentified backgrounds in the $e\tau_\mu$ channel, (d) misidentified backgrounds in the $\mu\tau_e$ channel, (e) misidentified backgrounds in the $e\tau_{\mathrm{had}}$ channel, and (f) misidentified backgrounds in the $\mu\tau_{\mathrm{had}}$ channel.

7.8.1.2 Best-fit results for the branching ratios and exclusion limits

Figure 7.26 presents the derived 95% CL upper limits for the $H \to e\tau$ and $H \to \mu\tau$ signals, along with the contribution of each category. Observed excesses of 2.4σ and 1.6σ are found for the $H \to \mu\tau$ and $H \to e\tau$ signals, respectively. For the $H \to e\tau$

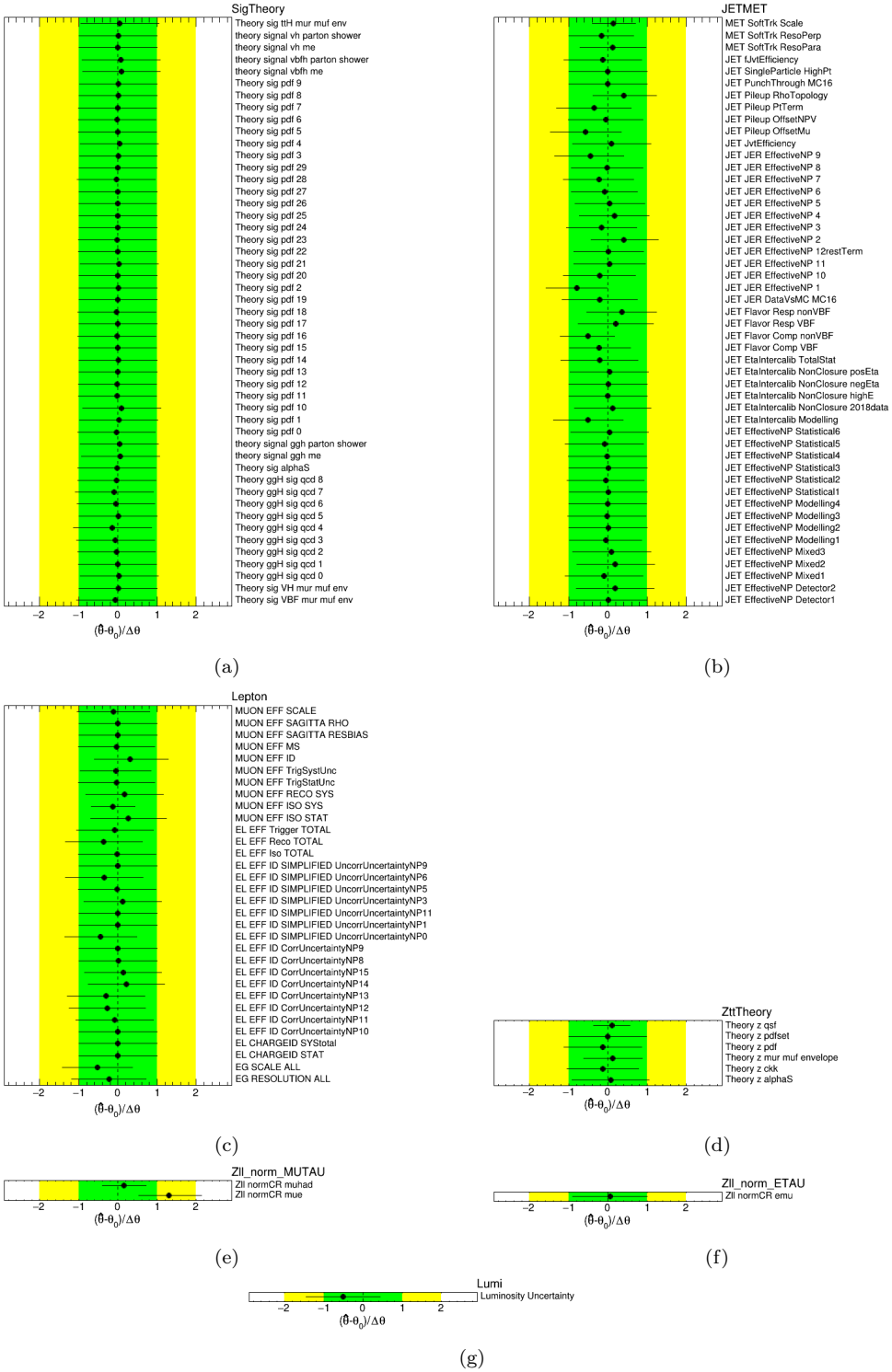
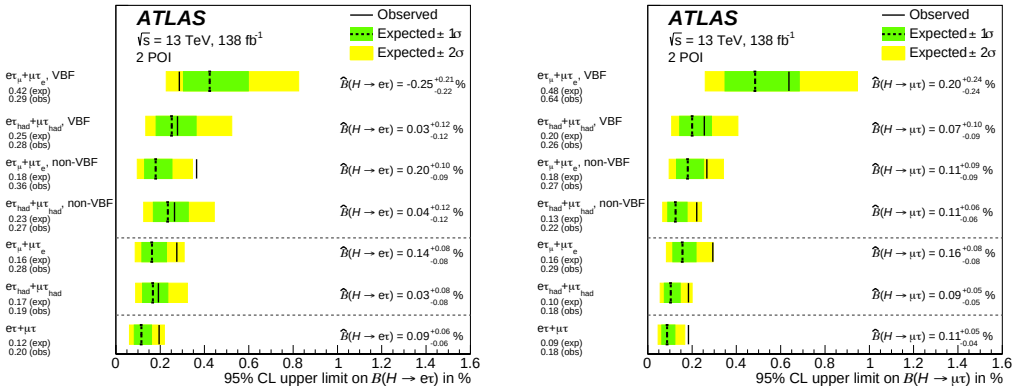


Figure 7.25: Best-fit values of the nuisance parameters in the 2-POI fit, organized by the different sources of systematic uncertainties: (a) theoretical uncertainties on the signal, (b) jet and $E_{\mathrm{T}}^{\mathrm{miss}}$, (c) electrons and muons, (d) theoretical uncertainties on the $Z \to \tau\tau$ background, (e) normalization of the $Z \to \ell\ell$ background in the $\mu\tau$ sample, (f) normalization of the $Z \to \ell\ell$ background in the $e\tau$ sample, and (g) luminosity.

signal, the observed excess is primarily attributed to the non-VBF category within the $\ell\tau_{\ell'}$ final state, using the inputs from the MC-based method. In contrast, the

excess in the $H \to \mu\tau$ signal is mainly driven by the non-VBF category of the $\ell\tau_{\text{had}}$ final state.

The 95% CL upper limits on the branching ratios of lepton-flavor-violating Higgs boson decays, including the $H \to e\mu$ search described in Ref. [82], are presented in Figure 7.27.



(a) (b)

Figure 7.26: Fit results for the simultaneous 2-POI measurement of the $H \to e\tau$ and $H \to \mu\tau$ signals. 95% CL upper limits on the LFV branching ratios of the Higgs boson are shown. These limits, shown for (a) $H \to e\tau$ and (b) $H \to \mu\tau$, are represented by solid lines for observed results and dashed lines for expected results. The best-fit branching ratio values ($\hat{B}$) are also included in percentage. Each likelihood fit incorporates the channels and categories listed along the y-axis, with signal and control regions from all other channels and categories excluded. Results from individual fits for standalone channels or categories are compared with the combined fit shown in the last row.

The results of the simultaneous fit to the $H \to e\tau$ and $H \to \mu\tau$ signals, including the 68% and 95% CL contours, are displayed in Figure 7.28. Since the two branching ratios are included in the same likelihood function, it is possible to scan the likelihood as a function of the two POIs to derive the 68% and 95% CL contours in the $\mu_{e\tau}$-$\mu_{\mu\tau}$ plane. The result is compatible with the SM prediction within 2.1σ. The figure also shows that the excess is driven by the $\mu\tau$ signal.

Figure 7.29 shows the upper limits on the individual coupling matrix elements $Y_{\tau\ell}$ and $Y_{\ell\tau}$, as obtained from the simultaneous fit of the two signals. The branching ratios are translated into constraints on the coupling matrix elements using equation 1.37. The 95% CL observed upper limits on the branching ratios translate into constraints on the coupling matrix elements, namely $\sqrt{|Y_{\tau e}|^2 + |Y_{e\tau}|^2} < 0.0013$ and $\sqrt{|Y_{\tau\mu}|^2 + |Y_{\mu\tau}|^2} < 0.0012$. Additionally, the indirect constraints from $\tau \to \ell\gamma$ searches [76, 78] are also shown for comparison.

Compared with these indirect limits, the direct limits from this search offer approximately an order of magnitude improvement for $Y_{\tau e}$ and $Y_{e\tau}$, and a factor of ten improvement for $Y_{\tau\mu}$ and $Y_{\mu\tau}$. These results show the effectiveness of direct LHC searches, that can probe the LFV Higgs couplings with higher sensitivity than indirect constraints.

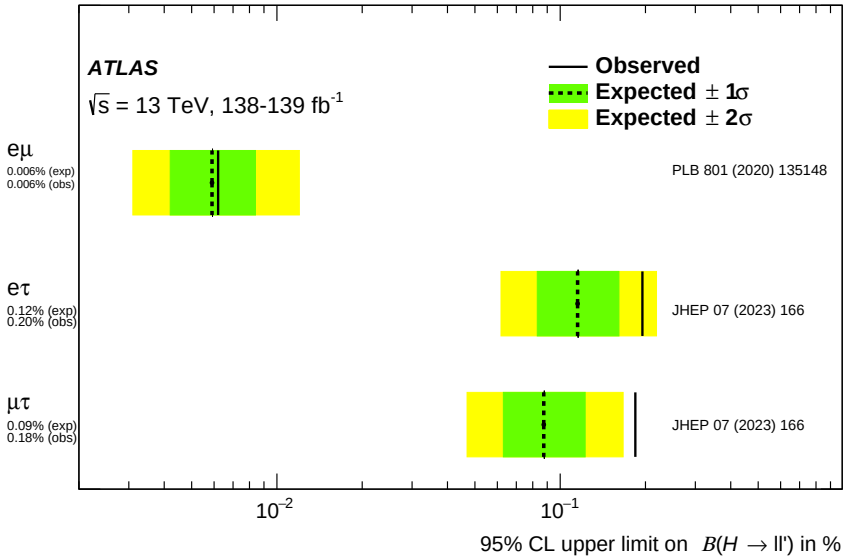


Figure 7.27: Upper limits at 95% CL on the branching ratios of the LFV decays, $H \to e\mu$, $H \to e\tau$, and $H \to \mu\tau$. The two last limits are obtained from the analysis presented in this thesis. The expected limits are shown as dashed lines, and the green and yellow bands represent the 1σ and 2σ uncertainty bands. The observed limits are shown as solid lines.

For $H \to \mu\tau$, the derived constraints are stricter than the naturalness limits, $|Y_{\tau\ell}Y_{\ell\tau}| \lesssim m_\tau m_\ell / v^2$, where v denotes the Higgs field vacuum expectation value This limits prevent large off-diagonal Yukawa matrix terms from causing a non-hierarchical mass spectrum [76]. This naturalness threshold remains unachieved in the $H \to e\tau$ case, due to the small mass value of the electron.

The impact of the different sources of systematic uncertainties on the observed branching ratios is shown in Table 7.9. The largest contributions come from the background sample size, the data sample size and the systematic uncertainties on the misidentified backgrounds. The searches are dominated by systematic uncertainties, but these are of statistical nature: the uncertainties on the background sample size are caused by the limited size of the MC samples, and include also the bin-by-bin statistical uncertainties of the misidentified background templates; the systematic uncertainties on the misidentified backgrounds are large in part due to the limited size of the datasets used to derive the systematic uncertainties.

The ranking plots of the different nuisance parameters and background normalization factors in the 2-POI fit are shown in Figure 7.30 for the $H \to e\tau$ and $H \to \mu\tau$ signals. The ranking plots show the high impact of the nuisance parameters controlling the background statistical uncertainties in the bins of the MVA distributions where the signal is expected to be present. The nuisance parameters related to the misidentified backgrounds are also ranked high. The misidentified backgrounds are not the leading background in the signal regions, but due to their complex and diverse topology, the MVA algorithms are not able to separate them from the signal as efficiently as the other backgrounds. The bins where the signal

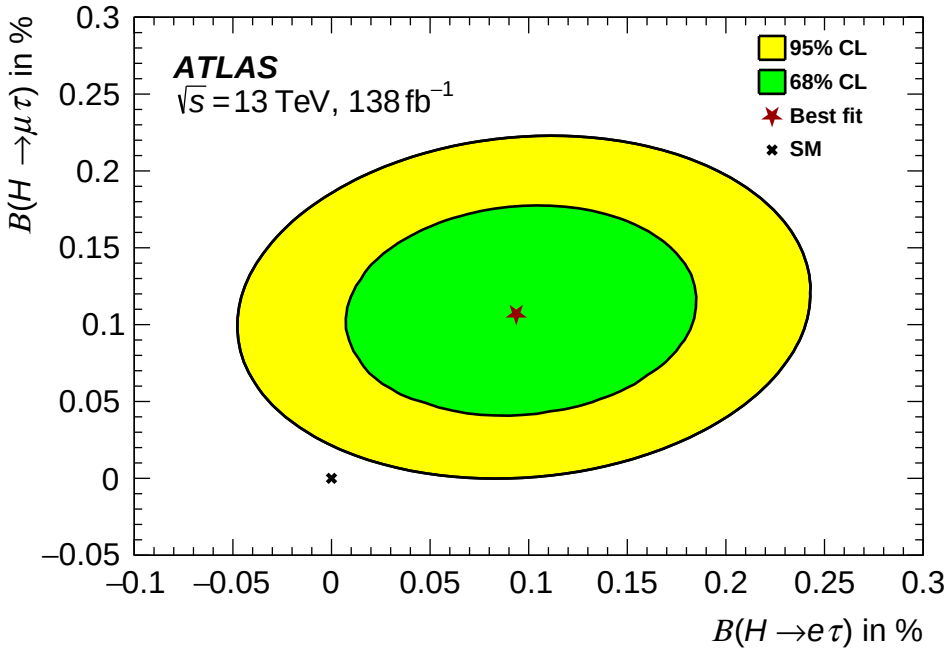


Figure 7.28: Best-fit value (red star) of the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, given in %, showing also the SM expectation for the branching ratios (black cross). The 68% and 95% CL contours are shown.

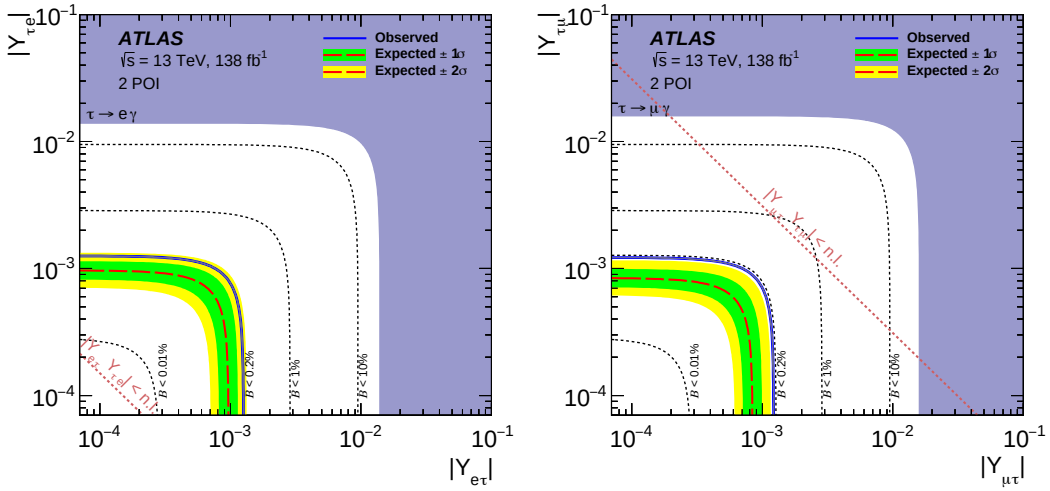
is expected to be present suffer from larger background contributions from the misidentified backgrounds than other background-populated bins. Combined with the conservative uncertainties on the misidentified backgrounds, this leads to a high impact of these nuisance parameters on the signal extraction.

7.8.1.3 Compatibility among categories in the MC-based analysis

The MC-based analysis estimates the parameters of interest using data from various channels and categories. In the nominal configuration, the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ are fitted simultaneously. The compatibility among categories is tested by comparing the results of the combined fit where the branching ratios are shared among all categories, to the results of fits where the POIs are different in each of the categories entering the fit.

The nominal fit setup includes 2 POIs, $\mu_{e\tau}$ and $\mu_{\mu\tau}$, while in the alternative setup (if both $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ datasets are included), 8 POIs are defined, one for each category (non-VBF and VBF in the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels). When the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ datasets are analyzed separately, 4 POIs are defined, corresponding to one for each category in each dataset

Each fit minimizes the negative log-likelihood of its corresponding model, denoted as $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively. These minimum likelihood values form the basis of a test statistic, calculated as $-2\log\left(\mathcal{L}_1/\mathcal{L}_2\right)$, which asymptotically follows a χ^2 distribution with degrees of freedom corresponding to the difference in dimensionality between Θ, representing the full parameter space, and Θ_0, the subset



(a) (b)

Figure 7.29: The expected (red long-dashed line) and observed (solid blue line) 95% CL upper limits from the simultaneous 2-POI fit of the two searches on the absolute values of the couplings $Y_{\tau\ell}$ and $Y_{\ell\tau}$ are shown, alongside the most stringent indirect limits from $\tau \to \ell\gamma$ searches (dark purple region) for (a) $\ell = e$ or (b) $\ell = \mu$. The short-dashed lines indicate limits corresponding to different branching ratios (0.01%, 0.2%, 1%, and 10%), while the dotted line marks the naturalness limit (denoted by n.l.).

under the null hypothesis, H_0, which assumes compatibility among categories for the LFV branching ratio measurement [244].

Thus, the likelihood functions used to derive the test statistic incorporate both 2-POI and 8-POI (4 POI) configurations if the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ datasets are combined (separated), resulting in a χ^2 distribution with 6 (2) degrees of freedom.

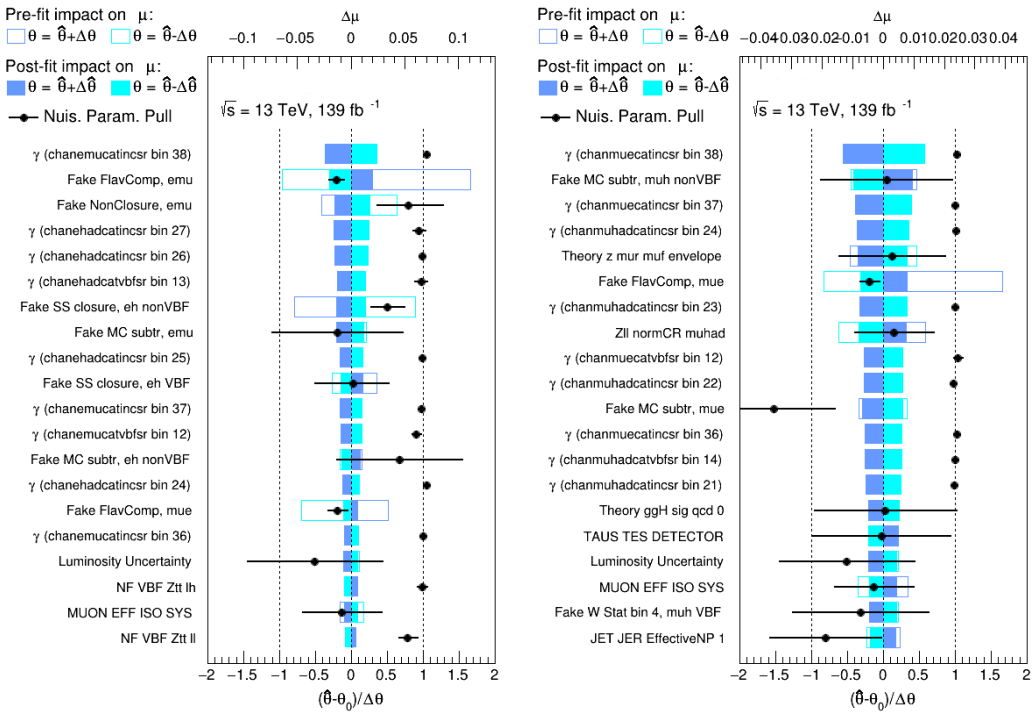
Standalone test results, shown in Table 7.10, indicate no significant deviations. The table shows the best-fit values and uncertainties for the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, both for the 2-POI and 4 POI/8 POI fits. The p-value for the compatibility between categories is calculated and converted into a significance. For the combined fit using the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ datasets, a compatibility p-value of 36.5% (or 0.91σ) is observed, indicating that the branching ratios are compatible among categories.

2-POI Source of uncertainty	Impact on observed $[10^{-4}]$ $\hat{\mathcal{B}}(H \to e\tau)$	 $\hat{\mathcal{B}}(H \to \mu\tau)$
Flavour tagging	0.7	0.2
Misidentified background ($e\tau_{\text{had}}$)	2.1	0.3
Misidentified background ($e\tau_{\mu}$)	2.7	0.3
Misidentified background ($\mu\tau_{\text{had}}$)	0.6	1.4
Misidentified background ($\mu\tau_{e}$)	0.9	1.0
Jet and $E_{\text{T}}^{\text{miss}}$	1.2	0.9
Electrons and muons	1.4	0.5
Luminosity	0.6	0.4
Hadronic τ decays	0.9	0.9
Theory (signal)	0.8	0.8
Theory (Z + jets processes)	0.8	1.0
$Z \to \ell\ell$ normalization ($e\tau$)	$<$0.1	$<$0.1
$Z \to \ell\ell$ normalization ($\mu\tau$)	0.2	0.9
Background sample size	3.7	2.3
Total systematic uncertainty	5.1	3.6
Data sample size	3.0	2.7
Total	5.9	4.5

Table 7.9: Impact of the different sources of systematic uncertainties on the observed branching ratios for the $H \to e\tau$ and $H \to \mu\tau$ signals using the 2-POI fit.

	2-POI	4-POI / 8-POI				p-value	Z/σ
	Combined	$\ell\tau_{\ell'}$ non-VBF	$\ell\tau_{\ell'}$ VBF	$\ell\tau_{\text{had}}$ non-VBF	$\ell\tau_{\text{had}}$ VBF		
$\mu_{e\tau}$	$0.14^{+0.08}_{-0.08}$	$0.21^{+0.09}_{-0.09}$	$-0.36^{+0.21}_{-0.21}$	-	-	0.198	1.29
$\mu_{\mu\tau}$	$0.16^{+0.08}_{-0.08}$	$0.15^{+0.09}_{-0.08}$	$0.17^{+0.23}_{-0.23}$	-	-		
$\mu_{e\tau}$	$0.03^{+0.08}_{-0.08}$	-	-	$0.04^{+0.12}_{-0.12}$	$0.03^{+0.12}_{-0.12}$	0.945	0.07
$\mu_{\mu\tau}$	$0.09^{+0.05}_{-0.05}$	-	-	$0.10^{+0.06}_{-0.06}$	$0.03^{+0.12}_{-0.12}$		
$\mu_{e\tau}$	$0.09^{+0.06}_{-0.06}$	$0.20^{+0.08}_{-0.08}$	$-0.35^{+0.21}_{-0.22}$	$0.04^{+0.12}_{-0.12}$	$0.03^{+0.12}_{-0.12}$	0.365	0.91
$\mu_{\mu\tau}$	$0.11^{+0.05}_{-0.04}$	$0.15^{+0.08}_{-0.08}$	$0.18^{+0.24}_{-0.23}$	$0.09^{+0.06}_{-0.06}$	$0.06^{+0.10}_{-0.09}$		

Table 7.10: Statistical test results evaluating the compatibility among categories for the 2-POI and 4-POI/8-POI fits.



(a) (b)

Figure 7.30: Ranking plots of the different nuisance parameters and background normalization factors in the 2-POI fit for the (a) $H \to e\tau$ and (b) $H \to \mu\tau$ signals.

Chapter 8

Conclusions and outlook

This thesis presents three main areas of research within the ATLAS experiment at the LHC. The first part covers the development of simulation software of the Tile Calorimeter readout electronics for the HL-LHC upgrades. The next two parts, focused on the experimental exploration of the Higgs sector, include measurements of Higgs boson production differential cross-sections in the $H \to \tau^+\tau^-$ decay channel and direct searches for lepton-flavor-violating decays into $e\tau$ and $\mu\tau$ final states, using the complete Run 2 dataset of 140 fb^{-1} of pp collisions recorded at $\sqrt{s} = 13$ TeV.

The TileCal Pulse Simulator for the HL-LHC upgrades

The High-Luminosity LHC represents a significant upgrade in the LHC's luminosity, with an expected peak luminosity of 7.5×10^{34} cm^{-2}s^{-1}, five times higher than the LHC's design luminosity. This upgrade will allow the LHC experiments to collect a larger dataset, enabling high-precision measurements and extending the reach for new physics searches. However, the increased luminosity also poses new challenges for the detectors, which must be upgraded to cope with the higher radiation levels and increased pile-up conditions.

The Tile Calorimeter, the central hadronic calorimeter of the ATLAS experiment, will undergo a significant upgrade. The readout electronics will be replaced to handle the increased trigger rates and cope with the increased radiation levels. Furthermore, the trigger and readout architecture will be completely redesigned. The data from the calorimeter cells will be reconstructed in real-time at the rate of 40 MHz, allowing the trigger system to make decisions based on the reconstructed information with the full granularity of the calorimeter. The development of online signal reconstruction algorithms is essential to ensure the robustness of the calorimeter's performance under high pile-up conditions.

The TileCal Pulse Simulator is a software package capable of injecting signals into the TileCal digitization chain, simulating the response of the TileCal channels.

The Pulse Simulator has been extended to model the readout electronics for the HL-LHC upgrades, including important features such as the pulse shape produced by the upgraded front-end electronics, the pile-up effects at the HL-LHC or the noise conditions. A continuous read-out mode has been implemented, allowing the simulation of continuous chains of ADC counts from the TileCal channels, mimicking the input that the online reconstruction algorithms implemented in the off-detector electronics will receive.

This work on the TileCal Pulse Simulator provides an essential tool for validating and optimizing the signal reconstruction algorithms in preparation for the HL-LHC data-taking. The Pulse Simulator's flexibility allows for extensive testing of the signal reconstruction algorithms.

Differential cross-section measurements of Higgs boson production in the $H \to \tau^+\tau^-$ decay channel

The discovery of the Higgs boson in 2012 marked a significant milestone in particle physics. It not only confirm the existence of the last missing piece of the Standard Model, but also opened a new era where the properties of the Higgs boson could be studied in detail in order to confirm its consistency with the Standard Model predictions and to search for deviations and new interactions that could hint at new physics beyond the Standard Model.

The cross-section measurements of the Higgs boson presented in this thesis are based on the full Run 2 dataset of 140 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV, recorded by the ATLAS experiment. The analysis did not only focus on the total Higgs boson production cross-section but also on differential cross-section measurements in the $H \to \tau^+\tau^-$ decay channel.

The cross-sections were measured for the four main Higgs production processes. For the VBF production mode, the cross-section ratio to the SM prediction was measured to be $0.93^{+0.17}_{-0.15}$, being the highest precision measurement achieved for this channel across all Higgs boson measurements. Differential cross-section measurements were performed within the Simplified Template Cross-Sections framework, providing a detailed view of the Higgs boson production as a function of kinematic variables such as p_{T}^{H}, m_{jj} or the number of jets in the event. Eight distinct bins were defined in the VBF phase space, based on p_{T}^{H} and m_{jj}. Eighteen STXS bins in total were measured, yielding a p-value of 6% for compatibility with the SM, indicating reasonable agreement. Additionally, for VBF with $p_{\mathrm{T}}^{H} \geq 200$ ($p_{\mathrm{T}}^{H} < 200$) GeV and $m_{\mathrm{jj}} > 1.5$ TeV, the cross-section was found to be $1.29^{+0.39}_{-0.34}$ ($0.12^{+0.34}_{-0.33}$) times the SM prediction, being this the first measurement in this high-p_{T}^{H} range and the most precise in the lower-p_{T}^{H} range.

Unfolded differential cross-section measurements were performed in a phase space enhanced in VBF production, specifically in the distributions of p_{T}^{H}, $p_{\mathrm{T}}(j_0)$, $\Delta\phi_{jj}^{\mathrm{signed}}$, and $\Delta\phi_{jj}^{\mathrm{signed}}$ vs p_{T}^{H}. These unfolded distributions allow model-independent

comparisons with theoretical predictions and serve as a basis for Effective Field Theory interpretations. The analysis of these distributions revealed good agreement with SM predictions within typical uncertainties of 30–50%, which remain predominantly limited by the statistical uncertainties of the data sample.

In the context of the Standard Model Effective Field Theory framework, these unfolded measurements provide sensitivity to potential beyond-the-Standard-Model interactions. The analysis placed constraints on six dimension-six operators, including c_{HW}, c_{HB}, c_{HWB}, and their CP-odd counterparts, all of which affect the VBF Higgs boson production vertex. For each coefficient, 95% confidence intervals were derived considering both the interference term alone and the combined interference and BSM contributions. Notably, the constraints on the CP-odd Wilson coefficient $c_{H\tilde{W}}$, derived as $[-0.31, +0.88]$, represent the strongest limits available to date from any Higgs boson decay channel.

These results complement measurements of Higgs boson properties obtained from other decay channels. The latest ATLAS measurements of the Higgs boson VBF production cross-section in the $H \to WW^*$ decay channel, using the full Run 2 dataset, reported a cross-section times branching ratio relative to the SM prediction of $0.93^{+0.23}_{-0.20}$ [245]. In the $H \to ZZ^*$ channel, this ratio was measured as 1.21 ± 0.44 [246], while in the $H \to \gamma\gamma$ and $H \to b\bar{b}$ channels, the values were $1.20^{+0.28}_{-0.25}$ [247] and $0.99^{+0.36}_{-0.34}$ [248], respectively.

Compared to the result of $0.93^{+0.17}_{-0.15}$ in the $H \to \tau^+\tau^-$ channel, these measurements demonstrate the consistency of Higgs boson production cross-sections across various decay modes. The $H \to \tau^+\tau^-$ channel offers a unique advantage in probing Higgs boson production in the VBF mode, providing the most precise ATLAS standalone measurement of the VBF production cross-section. Moreover, it underscores the value of differential cross-section measurements in the $\tau^+\tau^-$ final state for enhancing our understanding of Higgs boson production and properties.

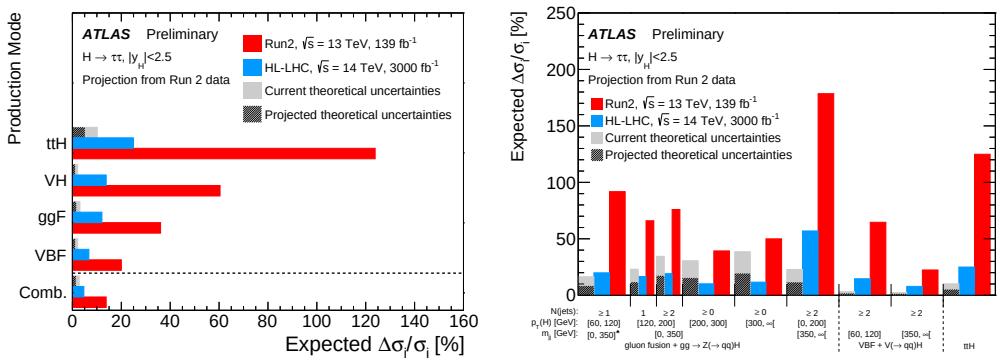
Similarly, the CMS Collaboration has reported measurements of Higgs boson production in the $H \to \tau^+\tau^-$ decay channel using pp collision data at $\sqrt{s} = 13$ TeV, corresponding to a total integrated luminosity of 138 fb^{-1} [223]. These studies cover the ggF, VBF, and VH production mechanisms, using a combination of event categorization techniques and machine learning methods. The results are provided as signal strengths normalized to SM predictions and as cross-section times branching ratio values across up to 16 kinematic regions defined by the STXS framework.

Two analyses target the simultaneous determination of ggF and VBF production contributions. The cut-based (CB) approach classifies events using categories optimized for the signal-to-background ratio and employs one- and two-dimensional discriminants to exploit kinematic features of the $\tau^+\tau^-$ and jet final states. In contrast, the NN approach relies on multiclassifier models, using their output as the primary observable for distinguishing signal and background. The NN analysis demonstrates greater sensitivity, achieving 30% tighter constraints on the inclusive VBF signal strength and up to 30% (40%) stronger constraints on the ggF (VBF)

Stage-1.2 STXS bins compared to the CB analysis. A separate analysis, focused on VH production, further complements these measurements.

The inclusive signal strength, derived from the combination of NN and VH analyses, is determined to be 0.82 ± 0.11, with ggF and VBF signal strengths measured at 0.67 ± 0.19 and 0.81 ± 0.17, respectively, and VH production yielding 1.79 ± 0.45.

The sensitivity of Higgs boson production cross-section measurements in the $H \to \tau^+\tau^-$ decay channel has been projected for the HL-LHC using results from Ref. [222], the first ATLAS analysis targeting this decay channel using the full Run 2 dataset. The extrapolation assumes a dataset of 3000 fb^{-1} at $\sqrt{s} = 14$ TeV and includes scaling factors for the increased integrated luminosity, higher center-of-mass energy, and improved detector performance. This study anticipates significant improvements in precision for both the total cross-section measurement and the STXS measurements. The STXS granularity used for the extrapolation only includes 9 bins, compared to the 18 bins used in the measurement presented in this thesis.



(a) (b)

Figure 8.1: Expected precision of the Higgs boson production cross-section measurements in the $H \to \tau^+\tau^-$ decay channel at the HL-LHC, compared to the Run 2 results. The projections are based on the Run 2 analysis and assume a dataset of 3000 fb^{-1} at $\sqrt{s} = 14$ TeV. Figure (a) shows the expected precision for the inclusive $pp \to H \to \tau^+\tau^-$ as well as for the production-specific cross-sections. Figure (b) shows the expected precision for the STXS measurements presented in Ref. [222]. The plots also show the current and projected theoretical uncertainties for the cross-sections.

For the inclusive $pp \to H \to \tau^+\tau^-$ cross-section, a total uncertainty of approximately 5% is expected. For specific Higgs boson production processes, the uncertainties are projected to be 11% for ggF, 7% for VBF, 14% for VH, and 24% for $t\bar{t}H$. These projections reflect both the increased dataset size and the expected reduction in experimental and theoretical uncertainties at the HL-LHC. These projections are shown in Figure 8.1a.

The scaling procedure accounts for the increased center-of-mass energy expected at the HL-LHC, and the larger dataset size, which reduces statistical uncertainties. Improvements in detector calibration, triggering, reconstruction, and

object identification are also included, with specific reductions in uncertainties for b-tagging, τ lepton identification, and $E_{\mathrm{T}}^{\mathrm{miss}}$resolution. Theoretical uncertainties, including those on cross-section predictions, parton distribution functions, and modeling of initial and final state radiation, are expected to decrease by a factor of two.

Within the STXS framework, measurements are expected to achieve diverse levels of precision across kinematic regions. For VBF-dominated topologies, uncertainties as low as 7% are projected. In contrast, ggF production in complex kinematic regimes, such as events with high jet multiplicity, large invariant mass of the dijet system, and $p_{\mathrm{T}}^{H} < 200$ GeV, may still experience uncertainties up to 50%. However, for ggF events with high Higgs transverse momentum, precision improves to approximately 10%, demonstrating the potential of these measurements in the HL-LHC era. The expected precision for the STXS measurements is shown in Figure 8.1b.

Searches for lepton-flavor-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$

Complementing the Higgs cross-section measurements, the thesis also presents two direct searches for lepton-flavor-violating decays of the Higgs boson, $H \to e\tau$ and $H \to \mu\tau$. These searches exploit an integrated luminosity of 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV and apply distinct background estimation techniques. A simultaneous fit of the potential $H \to e\tau$ and $H \to \mu\tau$ signals was performed, incorporating the MC-based $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$ channels. While slight excesses were observed, they fell below the threshold for confirming the presence of new physics. The observed (expected) 95% confidence level upper limits on the branching ratios were 0.20% (0.12%) for $H \to e\tau$ and 0.18% (0.09%) for $H \to \mu\tau$, reflecting a sensitivity enhancement of factors up to 2.5 and 1.6, respectively, compared to prior analyses at 13 TeV using 36 fb^{-1} of data. These limits on the LFV branching ratios are also translated into constraints on the off-diagonal elements of the Yukawa matrix, namely $\sqrt{|Y_{\tau e}|^2 + |Y_{e\tau}|^2} < 0.0013$ and $\sqrt{|Y_{\tau\mu}|^2 + |Y_{\mu\tau}|^2} < 0.0012$.

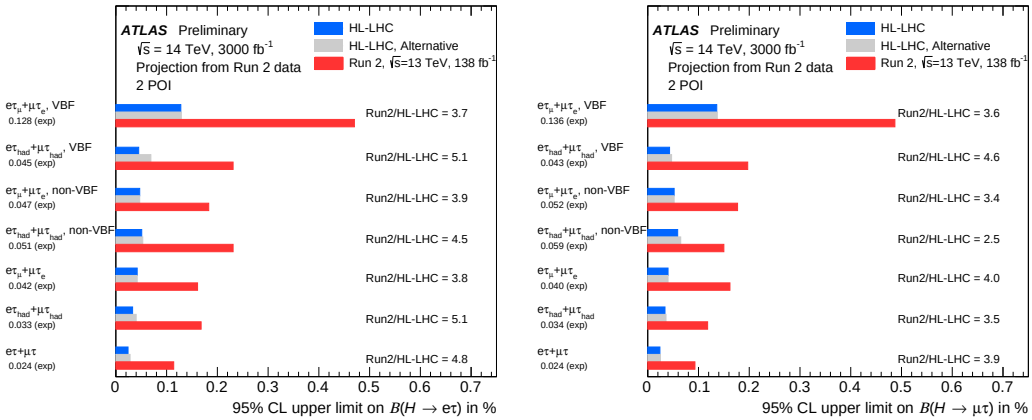
These improvements were achieved not only through the increased dataset but also through the optimization of the analysis strategy, including the implementation of new multivariate techniques, the refinement of the lepton assignment algorithm and the implementation of the simultaneous measurement of the $H \to e\tau$ and $H \to \mu\tau$ branching ratios in the statistical analysis.

A similar search for $H \to e\tau$ and $H \to \mu\tau$ decays was conducted by the CMS collaboration, utilizing 137 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV [85]. In this analysis, the results were obtained through a likelihood fit applied to the output of a BDT classifier, which was specifically trained to distinguish the LFV signal from major background processes in the $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$ channels. The BDT was trained using a combination of ggF and VBF signal events, and was designed

to separate the LFV signal from the main backgrounds, including misidentified leptons, $Z \to \ell\ell$, and $Z \to \tau\tau$ processes in the $\ell\tau_{\text{had}}$ channel, as well as $Z \to \tau\tau$, $t\bar{t}$, and misidentified lepton events in the $\ell\tau_{\ell'}$ channel.

The $Z \to \tau\tau$ background was derived from data using an embedding technique [249], while the misidentified lepton background was estimated through data-driven methods. All remaining backgrounds were modeled using MC simulation. Events were categorized into signal regions designed to target the VBF and ggF production mechanisms, with the ggF signal further divided into subcategories based on jet multiplicity. The analysis assumed the presence of only one possible LFV decay mode at a time and the final results were extracted via a likelihood fit to the BDT classifier output.

No significant excess over the background expectation was observed in the data. The resulting 95% CL upper limits on the branching ratios were measured to be 0.22% (expected 0.16%) for $H \to e\tau$ and 0.15% (expected 0.15%) for $H \to \mu\tau$.



(a) (b)

Figure 8.2: Expected 95% CL upper limits on the branching ratios for (a) $H \to e\tau$ and (b) $H \to \mu\tau$ decays at the HL-LHC. The limits are provided for the combination of the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels, as well as for the individual channels. They are compared to the expected limits derived from the Run 2 dataset and are based on the 2-POI likelihood fit methodology used in the Run 2 analysis. Two scenarios are evaluated for the HL-LHC dataset: one assuming no statistical uncertainties in the background sample sizes and another assuming that background sample sizes scale with the integrated luminosity.

The results of the $H \to e\tau$ and $H \to \mu\tau$ ATLAS searches have been extrapolated to the HL-LHC dataset, assuming pp collisions at $\sqrt{s} = 14$ TeV with a total integrated luminosity of 3000 fb^{-1} [250]. The extrapolation is based on the 2-POI likelihood fit methodology used in the Run 2 analysis, incorporating more precise theoretical predictions for signal and background processes, reduced experimental systematic uncertainties due to detector upgrades, modifications in production cross-sections, and the anticipated increase in background sample sizes. Two scenarios are evaluated for the background sample sizes: one where they scale

with the integrated luminosity, and another where statistical uncertainties in the background sample sizes are neglected.

The expected 95% CL upper limits on the branching ratios for $H \to e\tau$ and $H \to \mu\tau$ decays at the HL-LHC are presented in Figure 8.2. These results are shown for the combination of the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ channels, as well as for the individual categories. The expected 95% CL upper limits on the branching ratios are 0.024% for both $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$. This represents an improvement by factors of 4.8 and 3.9, respectively, compared to the limits achieved in the Run 2 analysis.

In the $\ell\tau_{\text{had}}$ channel, the category targeting the VBF production of the Higgs boson demonstrates the highest sensitivity, with expected improvements at the HL-LHC ranging from 4.6 to 5.1 factors. This enhanced sensitivity is attributed to the larger dataset, as the VBF signal region benefits more significantly from the increased luminosity compared to the non-VBF region, which exhibits higher yields. The VBF region, being more statistically limited due to lower event rates, profits more from the extended dataset. Conversely, in the $\ell\tau_{\ell'}$ channel, the category targeting gluon-gluon fusion production provides the best sensitivity.

At the HL-LHC, the analysis remains dominated by statistical uncertainties, even with a dataset of 3000 fb^{-1}, owing to the expected reductions in systematic uncertainties due to the improved theoretical predictions and the detector upgrades. However, the relative impact of systematic uncertainties on the total uncertainty is expected to increase as statistical uncertainties diminish.

The analyses presented in this thesis extend the understanding of the Higgs boson and its interactions, enhancing the sensitivity of the ATLAS experiment to the measurement of the Higgs boson properties in the $H \to \tau^+\tau^-$ decay channel, and exploring lepton-flavor-violating decays that would hint at new physics beyond the Standard Model. Future analyses with the Run 3 dataset, and with the anticipated High-Luminosity LHC dataset, will allow for higher precision in cross-section measurements and further constraints on BSM interactions, in the quest to understand the Higgs boson, its properties, and its role in the Standard Model and beyond.

Appendix A

Statistical methods: the profile likelihood

The probability of observing n_i events in the ith bin of a final discriminating variable follows a Poisson distribution. The expected event count in bin i, which could include both background (b_i) and signal (s_i) contributions, is expressed as $\mu s_i + b_i$, if just one signal process is considered. The background and signal predictions, b_i and s_i, respectively, are derived from the background estimation and Monte Carlo simulations. The parameter μ represents the signal strength modifier. A value of $\mu = 0$ means only background events are expected, while $\mu = 1$ means the signal events match the predicted yield. This signal strength can have different meanings depending on the context. For example, it is possible to interpret μ as the ratio of the measured cross-section to the Standard Model cross-section prediction (this is the case in the differential cross-section measurements presented in this thesis) or as the branching ratio of a BSM process (as in the searches for the lepton-flavor violating Higgs boson decays that are also presented in this thesis). In any case, μ is the primary parameter of interest (POI) in the analysis.

The parameter $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ denotes a set of nuisance parameters (NPs), where p is the number of nuisance parameters. These parameters influence the predicted signal and background but are not the central focus of the measurement. The NPs are usually constrained by auxiliary measurements, and account for systematic uncertainties, such as those related to the detector response, background estimation, and theoretical uncertainties. The treatment of these parameters will be discussed in detail later in this appendix.

In the expression above, b_i represent the total background prediction, which can be decomposed into several components. Some of these components can be also modified by a set of parameters $\vec{\lambda}$, which are called normalization factors. These factors are different from the NPs, as they are not constrained by auxiliary measurements, but are treated as free parameters in the fit, as the signal strength modifiers $\vec{\mu}$.

Under these conditions, the Poisson likelihood for bin i is:

$$\text{Pois}(n_i|\mu,\vec{\theta},\vec{\lambda}) = \frac{(\mu s_i(\vec{\theta}) + b_i(\vec{\theta},\vec{\lambda}))^{n_i}}{n_i!} e^{-(\mu s_i(\vec{\theta})+b_i(\vec{\theta},\vec{\lambda}))}. \tag{A.1}$$

This expression can be generalized to include multiple signal processes, each with its own signal strength $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$, where k iterates over the number of signal processes. In this case, the expected event count in bin i is given by $\sum_k \mu_k s_{i,k}(\vec{\theta}) + b_i(\vec{\theta})$. The likelihood function for bin i is then:

$$\text{Pois}(n_i|\vec{\mu},\vec{\theta},\vec{\lambda}) = \frac{\left(\sum_k \mu_k s_{i,k}(\vec{\theta}) + b_i(\vec{\theta},\vec{\lambda})\right)^{n_i}}{n_i!} e^{-\left(\sum_k \mu_k s_{i,k}(\vec{\theta})+b_i(\vec{\theta},\vec{\lambda})\right)}. \tag{A.2}$$

Once the data is collected, the observed event counts in all bins are fixed. If the likelihood function is maximized with respect to $\vec{\mu}$ and $\vec{\theta}$, the best estimates of these parameters are obtained. In a typical analysis, the discriminant contains several statistically independent bins, and different discriminants from different signal regions are combined. The overall likelihood combines the likelihoods of all bins and signal regions multiplying the individual bin likelihoods:

$$\mathcal{L}(\vec{n}|\vec{\mu},\vec{\theta},\vec{\lambda}) = \prod_{r\in\text{regions}} \prod_{i\in\text{bins}} \text{Pois}(n_{r,i}|\vec{\mu},\vec{\theta},\vec{\lambda}). \tag{A.3}$$

The parameters $\hat{\vec{\mu}}$, $\hat{\vec{\theta}}$ and $\hat{\vec{\lambda}}$ that maximize $\mathcal{L}(\vec{n}|\vec{\mu},\vec{\theta},\vec{\lambda})$ represent the best estimates, commonly referred to as the maximum likelihood estimators. This maximization is typically performed using the MINUIT [251] tool by minimizing $-\ln\mathcal{L}$. When multiple signal regions are combined, additional terms for each signal region are included in the overall likelihood product.

A.1 Treatment of Systematic Uncertainties

Systematic uncertainties, discussed in detail earlier in the thesis, are incorporated into the likelihood model through nuisance parameters, $\vec{\theta}$. Each nuisance parameter represents a systematic effect that can modify the signal and background predictions.

Statistical uncertainties on the background prediction related to the size of the sample used to estimate the background are treated separately and will be discussed in the next section.

Systematic uncertainties are often determined through auxiliary measurements, providing additional constraints. To incorporate this into the model, the likelihood function is modified by multiplying a Gaussian constraint term for each nuisance parameter θ_p, assuming a Gaussian probability density for the auxiliary measurement θ_0:

$$\text{Gauss}(\theta_p^0|\theta_p,\sigma_p^0) = \frac{1}{\sqrt{2\pi}\,\sigma_p^0} \exp\left(-\frac{(\theta_p-\theta_p^0)^2}{2\sigma_p^{0\,2}}\right). \tag{A.4}$$

Here, θ_p^0 represents the nominal value of the systematic uncertainty, and σ_p is the size of the $\pm 1\sigma$ variation obtained from auxiliary measurements. These variations represent how much the systematic uncertainty can shift the prediction of the signal or background on the final discriminating variable when the systematic effect is varied within its uncertainty. As the prediction moves further from the nominal value, the likelihood term decreases, controlled by σ_p. However, such shifts may still enhance the overall likelihood by improving the fit to the data. Since these constraint terms are independent of individual bins, the systematic uncertainties are treated as correlated across all bins. With these constraints, the likelihood function can be expressed as:

$$\mathcal{L}(\vec{n}|\vec{\mu},\vec{\theta},\vec{\lambda}) = \prod_{r\in\text{regions}} \prod_{i\in\text{bins}} \text{Pois}(n_{r,i}|\vec{\mu},\vec{\theta},\vec{\lambda}) \prod_{p} \text{Gauss}(\theta_p^0|\theta_p,\sigma_p^0). \tag{A.5}$$

Including terms in the likelihood covering different scales can lead to numerical instabilities and practical complications. To mitigate this, nuisance parameters are often rescaled so that the probability density functions used to constrain them have a mean of zero and a standard deviation of one. This normalization simplifies the handling of systematic uncertainties, particularly when a given systematic affects multiple categories, as it allows the systematic to be encoded in each category through the impact of the nuisance parameter on the category yields. The effect of moving the nuisance parameter θ_p away from its nominal value is computed for two fixed values of θ_p, $\theta_p^0 \pm \sigma_p$. However, the likelihood is continuous, and its maximization can lead to values of θ_p that are not exactly $\theta_p^0 \pm \sigma_p$. To account for this, the effect of the systematic uncertainty is modeled as a continuous interpolation between the $\pm 1\sigma$ limits. Systematic effects can be decomposed into acceptance (or normalization) components, η, and shape components, σ_i, in bin i. For each parameter, the nominal predictions are $\eta_0 = 1$ and σ_i^0, while the $\pm 1\sigma$ variations are represented by $\eta_\pm$ and $\sigma_i^\pm$. These two components are interpolated in different ways. For normalization effects, an exponential interpolation is used to ensure positivity, while a linear interpolation is employed for shape variations:

$$\begin{aligned} \eta(\theta_p) &= \begin{cases} \left(\eta^+\right)^{\theta_p} & \theta_p \geq 0 \\ \left(\eta^-\right)^{-\theta_p} & \theta_p < 0 \end{cases} \\ \sigma_i(\theta_p) &= \begin{cases} \sigma_i' + \theta_p\left(\sigma_i^+ - \sigma_i'\right) & \theta_p \geq 0 \\ \sigma_i' - \theta_p\left(\sigma_i' - \sigma_i^-\right) & \theta_p < 0 \end{cases}. \end{aligned} \tag{A.6}$$

For $\theta_p = 0$, the nominal value is recovered, while $\theta_p = \pm 1$ gives the $\pm 1\sigma$ variations. This results in the constraint:

$$\text{Gauss}(\theta_p^0) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta_p^2}{2}\right). \tag{A.7}$$

Each nuisance parameter is constrained by a Gaussian term like the one above. It is important to note that even if the size of the systematic uncertainty is an input from auxiliary measurements to the likelihood model, the fit to the data can constrain the uncertainty to a different value if the observable distribution is sensitive to the systematic effect.

A.2 Statistical uncertainties on the background sample size

Incorporating statistical uncertainties on the background prediction involves introducing nuisance parameters, reflecting potential deviations of the predicted background rates from their true values. While statistical uncertainties on the signal prediction are typically assumed to be negligible (analysis are designed to have large acceptances to the targeted signals and the signal predictions are based on Monte Carlo simulations with large statistics), uncertainties on the background are parametrized by one nuisance parameter, denoted as γ_i, for each statistically independent bin i. The approach explained here, known as light Beeston-Barlow method [252], captures the potential variation between the predicted background rate, b_i, and the true rate, which is determined by fitting the predictions to the observed data.

The true rates for each bin could, in principle, be derived by assigning individual nuisance parameters per process and bin for statistically independent processes. However, this introduces a significant number of nuisance parameters. To simplify the analysis, the total background is treated as a single combined process.

The background prediction b_i in each bin is computed as the sum of the weights w_j of the N_i events in the bin, expressed as:

$$b_i = \sum_{j=1}^{N_i} w_j. \tag{A.8}$$

The associated statistical uncertainty on b_i, denoted by σ_{b_i}, is given by:

$$\sigma_{b_i} = \sqrt{\sum_{j=1}^{N_i} w_j^2}\,. \tag{A.9}$$

This uncertainty is incorporated into the likelihood function through an additional term $\mathcal{L}_\gamma$. This term reflects the fact that the observed background rate could differ from the predicted one. Assuming an auxiliary measurement is made with the same relative statistical uncertainty $\sigma_{\text{rel},i}$ as that of the predicted background b_i, the result of this auxiliary measurement is expressed by m_i unweighted events (i.e., where all $w_j = 1$). The Poissonian constraint term added to the likelihood function takes the form:

$$\mathcal{L}_\gamma = \frac{(\gamma_i m_i)^{n_i} e^{-\gamma_i m_i}}{n_i!}. \tag{A.10}$$

In this expression, n_i represents the observed number of events, γ_i is the nuisance parameter for bin i, and m_i represents the mean expected background events in that bin. The true rate equals the predicted one for $\gamma_i = 1$. The relative statistical uncertainty $\sigma_{\text{rel},i}$ is related to the number of unweighted events m_i by the following relation:

$$\sigma_{\text{rel},i} = \frac{\sigma_{b_i}}{b_i} = \frac{1}{\sqrt{m_i}}. \tag{A.11}$$

Since m_i is not necessarily an integer, the gamma distribution is used in place of a discrete Poisson distribution, and the following additional term is included in the likelihood function:

$$\mathcal{L}_\gamma(\gamma_i|m_i) = \frac{1}{\Gamma(m_i)}(\gamma_i m_i)^{n_i} e^{-\gamma_i m_i}, \tag{A.12}$$

where Γ denotes the gamma function. Furthermore, to account for these uncertainties, the background prediction b_i must be scaled by the nuisance parameter γ_i, modifying the prediction term in the likelihood function:

$$\mathcal{L}(\vec{n}|\vec{\mu},\vec{\theta},\vec{\lambda},\vec{\gamma}) = \prod_{r\in\text{regions}} \prod_{i\in\text{bins}} \text{Pois}(n_{r,i}|\vec{\mu},\vec{\theta},\vec{\lambda},\vec{\gamma})\mathcal{L}_\gamma(\gamma_{r,i}|m_{r,i}) \prod_p \text{Gauss}(\theta_p^0|\theta_p,\sigma_p^0), \tag{A.13}$$

where $\vec{\gamma}$ represents the set of nuisance parameters γ_i for each bin i that account for the statistical uncertainties on the background prediction and the Poisson terms includes the background prediction b_i scaled by γ_i:

$$\text{Pois}(n_i|\vec{\mu},\vec{\theta},\vec{\lambda},\vec{\gamma}) = \frac{\left(\sum_k \mu_k s_{r,i,k}(\vec{\theta}) + \gamma_{r,i} b_{r,i}(\vec{\theta},\vec{\lambda})\right)^{n_i}}{n_i!} e^{-\left(\sum_k \mu_k s_{r,i,k}(\vec{\theta}) + \gamma_{r,i} b_{r,i}(\vec{\theta},\vec{\lambda})\right)}. \tag{A.14}$$

A.3 Profile-likelihood unfolding

The profile-likelihood can be modified to unfold distributions from the detector level to the particle level. This process is known as profile likelihood unfolding (PLU). This approach models the detector-level distribution by treating each truth-level bin of the particle-level distribution as a subcomponent of the signal at the detector level. The key elements of this procedure are the response matrix, acceptance, and efficiency corrections, which account for how events migrate between bins due to detector resolution, and how many events pass the detector-level and truth-level selections.

The likelihood function for the unfolding process compares the observed number of events in each bin at the detector level to the predicted number of

events. This prediction is formed by folding the truth-level cross sections through the response matrix and adding the expected background. The likelihood can be written as:

$$\mathcal{L}\left(\vec{n}|\vec{\mu},\vec{\theta},\vec{\lambda},\vec{\gamma}\right) = \prod_{r\in\text{regions}}\prod_{i\in\text{detector bins}} \text{Pois}\Bigg(n_{r,i}\Bigg| L_{\text{int}}\sum_{j\in\text{truth bins}} \mathcal{R}_{r,ij}(\vec{\theta})\mu_j\sigma_j^{\text{MC}} + \gamma_{r,i}b_{r,i}(\vec{\theta},\vec{\lambda})\Bigg)\mathcal{L}_\gamma(\gamma_{r,i}|m_{r,i})\prod_p \text{Gauss}(\theta_p^0|\theta_p,\sigma_p^0), \tag{A.15}$$

where $n_{r,i}$ is the observed number of events in region r and detector-level bin i, L_{int} is the integrated luminosity, and $\mathcal{R}_{ij,r}$ is the response matrix. The signal prediction associated to a truth bin j is expressed here as the product of the integrated luminosity and the cross-sections in the MC simulation, $L_{\text{int}}\sigma_j^{\text{MC}}$. In this expression, the signal strength modifiers $\vec{\mu}$, which are the free parameters in the fit, scale the cross-sections in each truth-level bin.

The response matrix $\mathcal{R}_{r,ij}$ encodes the probability that an event generated in truth-level bin j is reconstructed in detector-level bin i. This matrix accounts for the smearing of the true signal due to the detector's finite resolution, leading to migrations between bins. The response matrix is composed of the migration matrix $\mathcal{M}_{ij}$, the efficiency ϵ_j, and out-of-acceptance correction $f_{r,i}$:

$$\mathcal{R}_{r,ij} = \frac{\mathcal{M}_{r,ij}\epsilon_j}{f_{r,i}}. \tag{A.16}$$

The migration matrix $\mathcal{M}_{r,ij}$ is calculated as:

$$\mathcal{M}_{ij} = \frac{N_{r,ij}^{\text{det.}\cap\text{fid.}}}{N_j^{\text{det.}\cap\text{fid.}}}, \tag{A.17}$$

where $N_{ij}^{\text{det.}\cap\text{fid.}}$ is the expected number of events that are both reconstructed in detector-level bin i and generated in truth-level bin j. This captures the migration between different truth-level and detector-level bins.

Efficiency corrections ϵ_j account for events that pass the fiducial selection at truth level but are not successfully reconstructed at detector level. These are defined as:

$$\epsilon_j = \frac{N_j^{\text{det.}\cap\text{fid.}}}{N_j^{\text{fid.}}}, \tag{A.18}$$

where $N_j^{\text{fid.}}$ is the number of events that satisfy the fiducial selection at the truth level. This ratio ensures that the unfolding procedure corrects for detector inefficiencies.

Acceptance corrections α_i are introduced to account for events that satisfy the detector-level selection but come from configurations outside the fiducial region at the truth level. These are given by:

$$f_{r,i} = \frac{N_{r,i}^{\text{det.} \cap \text{fid.}}}{N_{r,i}^{\text{det.}}}, \tag{A.19}$$

where $N_{r,i}^{\text{det.}}$ is the number of events in detector-level bin i.

Systematic uncertainties, represented by nuisance parameters $\vec{\theta}$, enter the likelihood through both the signal term (via the response matrix $\mathcal{R}_{r,ij}$) and the background term $b_{r,i}$. These uncertainties reflect potential variations in the detector response and background predictions.

In the unfolding process, the differential cross-sections modifiers at truth level are treated as free parameters in the fit. The profile likelihood fit adjusts these parameters such that the predicted detector-level distributions best match the observed data. Unlike regular unfolding methods, which rely on direct inversion of the response matrix, PLU incorporates systematic uncertainties and allows the statistical treatment of both the signal and background in a single framework.

Regularization is sometimes used to smooth out statistical fluctuations in the unfolded result, though it is not required in cases where bin-to-bin migrations are small. The analysis discussed in this thesis does not use regularization, and therefore the topic is not covered in this appendix.

A.4 Likelihood minimization

Once the profile likelihood function is constructed, the next step is to minimize it to obtain the best estimates of the parameters. This task involves navigating a high-dimensional parameter space, which can include hundreds of variables (POIs, NFs, NPs). The procedure is not only responsible for finding the global minimum but also for assessing the correlations between the various input parameters. To achieve this, the minimization algorithm utilizes the Davidon-Fletcher-Powell (DFP) method.

The process begins by initializing the parameters, $\vec{X}$ at certain starting values. At this stage, the gradient of the negative log-likelihood function, denoted by $\vec{G}$, is calculated. Initially, it is assumed that the parameters are uncorrelated, which simplifies the Hessian matrix (the matrix of second derivatives) by setting it to the identity matrix. This assumption serves as a reasonable starting point in many physical models, allowing the minimization to proceed effectively.

Once the gradient is known, a line search is performed in its direction, aiming to locate the parameter α that minimizes the negative log-likelihood function $F(\vec{X} - \alpha V\vec{G})$, where V represents the covariance matrix of the parameters. This covariance matrix, initially taken as the inverse of the Hessian, is refined as the minimization progresses. The line search is an iterative process that adjusts the parameter values in the direction of the gradient, effectively moving the function toward the minimum.

After each adjustment, the covariance matrix V is updated using DFP formulas, which progressively account for the correlations between the variables.

As the procedure advances, the global minimum is approached, and the distance to the minimum is estimated by calculating the quantity known as the estimated distance to minimum, which is given by $\vec{G}^T V \vec{G}$. When this value falls below a predefined threshold, the minimization is considered to have converged.

The minimization process yields a correlation matrix, derived from the covariance matrix, that captures the relationships between the NPs and the POIs. The diagonal elements of this matrix provide the post-fit uncertainties of the parameters. Although these uncertainties are typically symmetric, they are valid only in the local region around the minimum, where the likelihood function can be approximated by a parabola in each direction.

This quadratic approximation can be insufficient in several cases, for example, when the number of events in the bins is small. In these cases, the Minos algorithm is used to provide more accurate uncertainty estimates. This algorithm does not rely on the assumption of a parabolic shape of the likelihood function, but scans the likelihood around the minimum in order to obtain the correct uncertainties, taking also into account the correlations between the parameters.

A.5 Hypothesis testing

Hypothesis testing is a statistical procedure used to evaluate the presence of a potential signal or to establish upper bounds on its production rate, characterized by the product of the cross section and branching ratio, denoted as μ. The method relies on defining two competing hypotheses: the null hypothesis H_0, which represents the absence of the signal ($\mu = 0$), and the alternative hypothesis H_1, which assumes a signal contribution ($\mu > 0$). A *test statistic*, represented as t, serves as the basis for assessing the degree of compatibility between the observed data and each hypothesis. Determining the behavior of t requires knowledge of its probability density functions (pdfs) under both H_0 and H_1.

To decide between H_0 and H_1, a threshold value for the test statistic, t_a, is chosen such that the probability of obtaining $t \geq t_a$ under H_0 matches a predefined significance level α:

$$\alpha = \int_{t_a}^{\infty} f(t|H_0)\,\mathrm{d}t. \tag{A.20}$$

If the observed test statistic, t_{obs}, exceeds t_a, H_0 is rejected in favor of H_1. Conversely, the probability of failing to reject H_0 when H_1 is true, known as β, is defined as:

$$\beta = \int_{-\infty}^{t_a} f(t|H_1)\,\mathrm{d}t. \tag{A.21}$$

The complementary quantity, $1 - \beta$, is referred to as the power of the test and measures the likelihood of correctly identifying H_1 as true. A well-designed test statistic maximizes this power for a fixed α. According to the Neyman-Pearson

lemma [253], the likelihood ratio is the optimal test statistic for distinguishing between simple hypotheses.

For practical applications, a common choice of test statistic for a fixed signal strength μ is:

$$t_\mu = -2 \ln \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \tag{A.22}$$

where $\mathcal{L}$ represents the likelihood function, $\hat{\mu}$ and $\hat{\theta}$ are the unconditional maximum likelihood estimators (MLEs) of the signal strength and nuisance parameters, respectively, and $\hat{\hat{\theta}}$ is the conditional MLE for a specified value of μ. In the asymptotic limit, Wilks' theorem [244] establishes that t_μ follows a χ^2 distribution with one degree of freedom.

The observed value of the test statistic, $t_{\mu,\text{obs}}$, can be used to calculate a p-value, which measures the probability of obtaining a result at least as extreme as the observed one under the null hypothesis:

$$p_\mu = \int_{t_{\mu,\text{obs}}}^{\infty} f(t_\mu | H_0) \, \mathrm{d}t. \tag{A.23}$$

This p-value can be expressed as a corresponding significance Z in terms of a standard normal distribution:

$$Z = \Phi^{-1}(1 - p_\mu), \tag{A.24}$$

where Φ^{-1} is the inverse cumulative distribution function of the standard Gaussian.

A.5.1 Signal Discovery

Discovering a new signal involves rejecting the background-only hypothesis, corresponding to $\mu = 0$. For this purpose, a specialized test statistic q_0 is employed:

$$q_0 = \begin{cases} -2 \ln \frac{\mathcal{L}(0, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, & \text{if } \hat{\mu} \geq 0, \\ 0, & \text{if } \hat{\mu} < 0. \end{cases} \tag{A.25}$$

By construction, q_0 is non-negative, ensuring stability against statistical fluctuations that may produce negative estimates of $\hat{\mu}$. The significance of the discovery can then be expressed as $Z = \sqrt{q_0}$. The particle physics community conventionally requires a significance of 5σ ($p \leq 2.87 \times 10^{-7}$) for a discovery and 3σ for evidence.

A.5.2 Setting Upper Limits

To establish upper limits on μ, the hypothesis incorporating both signal and background contributions is treated as H_0, while the background-only case is H_1. The test statistic $\tilde{q}_\mu$ used in this context is defined as:

$$\tilde{q}_\mu = \begin{cases} -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(0,\hat{\hat{\theta}}(0))}, & \text{if } \hat{\mu} < 0, \\ -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\theta})}, & \text{if } 0 \leq \hat{\mu} \leq \mu, \\ 0, & \text{if } \hat{\mu} > \mu. \end{cases} \tag{A.26}$$

The confidence level for excluding a signal strength μ, denoted as CL_s, is calculated using:

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_{\mu=0}}, \tag{A.27}$$

where p_μ corresponds to the p-value of the signal-plus-background hypothesis, and $p_{\mu=0}$ represents the p-value of the background-only hypothesis. The exclusion of H_0 at a confidence level of $1-\alpha$ occurs if $\mathrm{CL}_s < \alpha$. This approach provides a conservative limit by penalizing hypotheses with limited experimental sensitivity [243], ensuring that poorly sensitive tests do not lead to unjustified exclusions.

Appendix B

The VBF-like gluon-gluon fusion contribution in the $H \to \tau\tau$ STXS measurement

As illustrated in Figure 6.42, a strong correlation is observed between the low-p_{T}^{H} VBF parameters of interest and the scaling parameter for the VBF-like ggF contribution. This latter contribution corresponds to ggF production at low p_{T}^{H} with two jets and invariant mass m_{jj} above 350 GeV. The analysis was designed to separate these contributions across distinct regions—specifically, VBF_0, which is enriched in ggF and $Z \to \tau\tau$ events, and VBF_1, which is enriched in VBF events. However, sensitivity to the VBF-like ggF component within this analysis remains limited.

Among the four VBF low-p_{T}^{H} parameters, three show values below the SM expectation, while the VBF-like ggF parameter has a best-fit value significantly above the SM prediction. Additional studies were performed to explore the interrelationship of these effects. Due to its purpose of reconstructing the Higgs boson mass with precision, the $m_{\tau\tau}^{\mathrm{MMC}}$ provides limited separation between these two signals, which can exhibit comparable contributions across certain signal regions, as shown in Figure 6.7. This correlation between the VBF and VBF-like ggF POIs is thus attributable to the considerable ggF contamination in VBF-enriched regions and the similar shape of the respective templates.

To assess the influence of the VBF-like ggF contribution on the fit, a modified fit was conducted. In this approach, normalization factors for the low-p_{T}^{H} VBF contributions were merged to those of their corresponding VBF-like ggF contributions, grouped by m_{jj} bins. Here, the relative scaling between the VBF-like ggF and VBF contributions was fixed to the SM theory predictions. Results of this test are displayed in Figure B.1, where the best-fit values align more closely with the SM expectations. The p-value for the compatibility of these results with the SM is 12%.

This test indicates that the current fit configuration does not effectively separate the two signal contributions. Additionally, the fit can exploit normalization

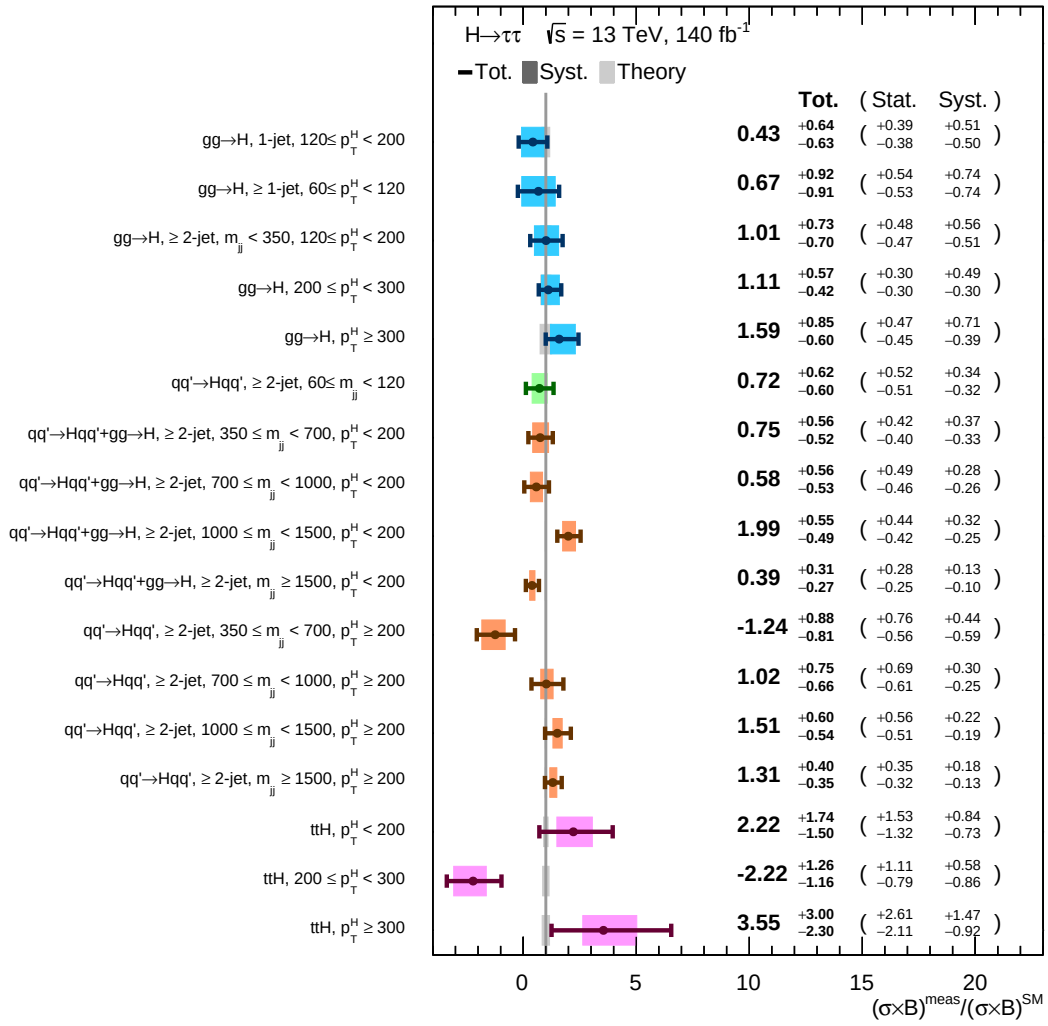


Figure B.1: Measured values for $\sigma_H \times B(H \to \tau\tau)$ relative to the SM expectations in the simplified template cross-section measurement when the VBF-like ggF contribution is merged with the low-p_{T}^H VBF contributions. The error bars and shaded areas show the total uncertainty and systematic uncertainty in the measurements, respectively.

factors associated with specific signal templates to account for data fluctuations (e.g., underfluctuations in VBF_1 regions). This adjustment can be made while still accommodating the signal contributions in other regions, such as VBF_0, which experiences fewer statistical fluctuations and is more sensitive to the VBF-like ggF component. The test demonstrates that the anti-correlation between the parameters of interest explains the low values of the low-p_{T}^H VBF parameters through the high values of the VBF-like ggF contribution.

Appendix C

Searches for lepton-flavor-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ final states using the Symmetry method

The symmetry method is a data-driven approach that relies on the symmetry between electrons and muons in SM processes to estimate the background in the $\ell\tau_{\ell'}$ channel. The data is divided into two samples, $e\tau_\mu$ and $\mu\tau_e$, according to the lepton assignment, and the background in each sample is estimated from the other sample, taking into account the different experimental effects that affect electrons and muons.

C.1 Selection in the Symmetry-based $\ell\tau_{\ell'}$ channel

The event selection in the Symmetry-based method closely mirrors that of the MC-based method, with a few differences. These differences are necessary to ensure that the symmetry between electrons and muons is maintained after the event selection. The assignment of leptons follows the same procedure as in the MC-based approach.

In the Symmetry-based method, the cut on the leading lepton transverse momentum is set to $p_{\mathrm{T}}^{\ell_1} > 35$ GeV. This adjustment increases the number of selected events, which is beneficial for the data-driven background estimation and the training of the multivariate algorithms, as will be detailed in Section 7.5.

The symmetry method is based on the inherent symmetry of SM backgrounds under the exchange of electrons and muons. To maintain this symmetry, the selection criteria are adapted to ensure equal treatment of the electron and muon. The $|\eta|$ requirement, typically applied to electrons, is also imposed on muons, requiring $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$. Additionally, a $d_0/\sigma_{d_0} < 10$ criterion is applied to both electrons and muons. However, the ratio of the track p_{T} to

the cluster p_{T}, which is used in the MC-based method, is not applied here. This criterion was originally applied only when $\ell_2 = e$, and extending it to muons would violate the symmetry between the two leptons.

C.2 Background estimation in the Symmetry $\ell\tau_{\ell'}$ channel

The e/μ-symmetry method provides a data-driven approach for distinguishing between $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$. This method relies on two primary assumptions. The first assumption is that SM processes exhibit a symmetry when swapping prompt electrons and prompt muons. As a result, the kinematic distributions of these prompt leptons are expected to be nearly identical. The second assumption is that the flavour-violating Higgs decays break this symmetry, leading to differences in the distributions for such decays.

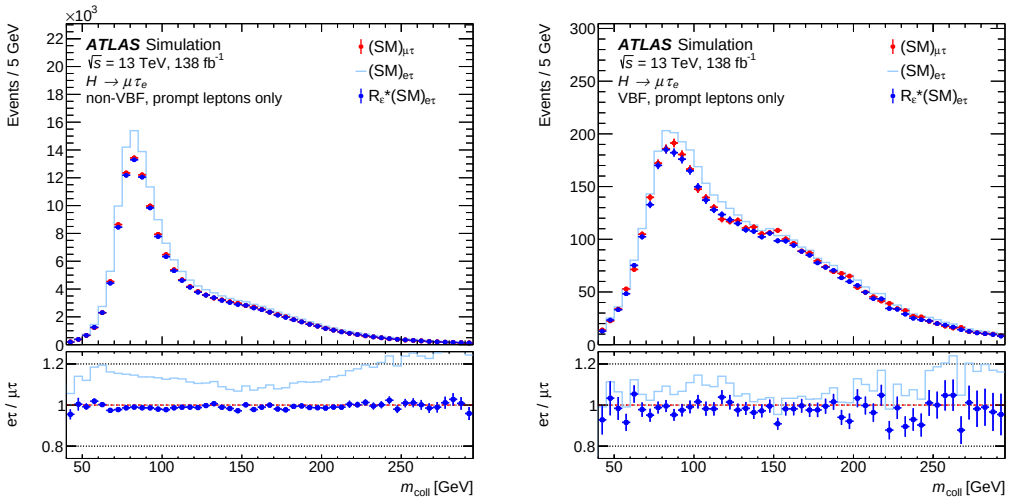
Under the e/μ-symmetry hypothesis, the SM background is assumed to be symmetric between the $\mu\tau_e$ and $e\tau_\mu$ samples. However, the $H \to \mu\tau$ signal predominantly populates the $\mu\tau_e$ dataset, given that the p_{T} of the muon from the Higgs decay tends to be larger than the p_{T} of the electron resulting from the τ-lepton decay. Conversely, the $H \to e\tau$ signal is more likely to appear in the $e\tau_\mu$ dataset. Therefore, when searching for $H \to \mu\tau$ decays, the SM background can be estimated from the $e\tau_\mu$ dataset, and the same holds for the reverse case when searching for $H \to e\tau$ decays.

Nevertheless, detector-related effects introduce deviations from the expected e/μ-symmetry, and these effects need to be considered. One source of asymmetry comes from events that contain misidentified or non-prompt leptons. These leptons arise from a variety of sources, including misidentified jets, misclassified light leptons, hadronic decays within jets, and electrons from photon conversions. The contributions from these particles are not identical between the $\mu\tau_e$ and $e\tau_\mu$ datasets. Another source of asymmetry comes from differences in trigger, reconstruction, identification, and isolation efficiencies for muons and electrons, which vary depending on kinematic factors such as p_{T}, $|\eta|$, and ϕ.

The misidentified leptons are classified into two groups based on their estimation methods: MC-fakes, estimated from simulations, and FF-fakes, derived from data using the fake-factor method.

For MC-fakes, events arise from misidentifications such as τ_{had} as light leptons, muons as electrons, or photons as electrons. The main contributors are processes like $Z \to \tau\tau$, top-quark, diboson, $Z \to \mu\mu$, $Z \to ee$, and $V\gamma$, with $Z \to \mu\mu$ being the largest source, particularly when low-p_{T} muons are misidentified as electrons. The simulation of $Z \to \mu\mu$ is validated in a region dominated by such events. A 16% uncertainty is applied to all MC-fakes contributions.

The FF-fakes category includes jets misidentified as leptons and leptons from hadronic decays in jets, primarily from W + jets and to a lesser extent from multijet



(a) (b)

Figure C.1: The collinear mass distributions, m_{coll}, are shown for simulated events meeting the (a) non-VBF and (b) VBF selection criteria. These distributions compare events with a leading muon and a subleading electron originating from τ-lepton decay (denoted as $(\text{SM})_{\mu\tau}$) against events with a leading electron and subleading muon without efficiency correction ($(\text{SM})_{e\tau}$) and with the efficiency correction applied ($R_\epsilon(\text{SM})_{e\tau}$). In the lower panel, the ratio of the $e\tau$ to $\mu\tau$ predictions is displayed. Dark blue points represent the ratio for the $e\tau$ prediction with efficiency correction applied, $R_\epsilon(\text{SM})_{e\tau}/(\text{SM})_{\mu\tau}$, while the light blue line indicates the $e\tau$ prediction without efficiency correction, $(\text{SM})_{e\tau}/(\text{SM})_{\mu\tau}$. Only statistical uncertainties are included.

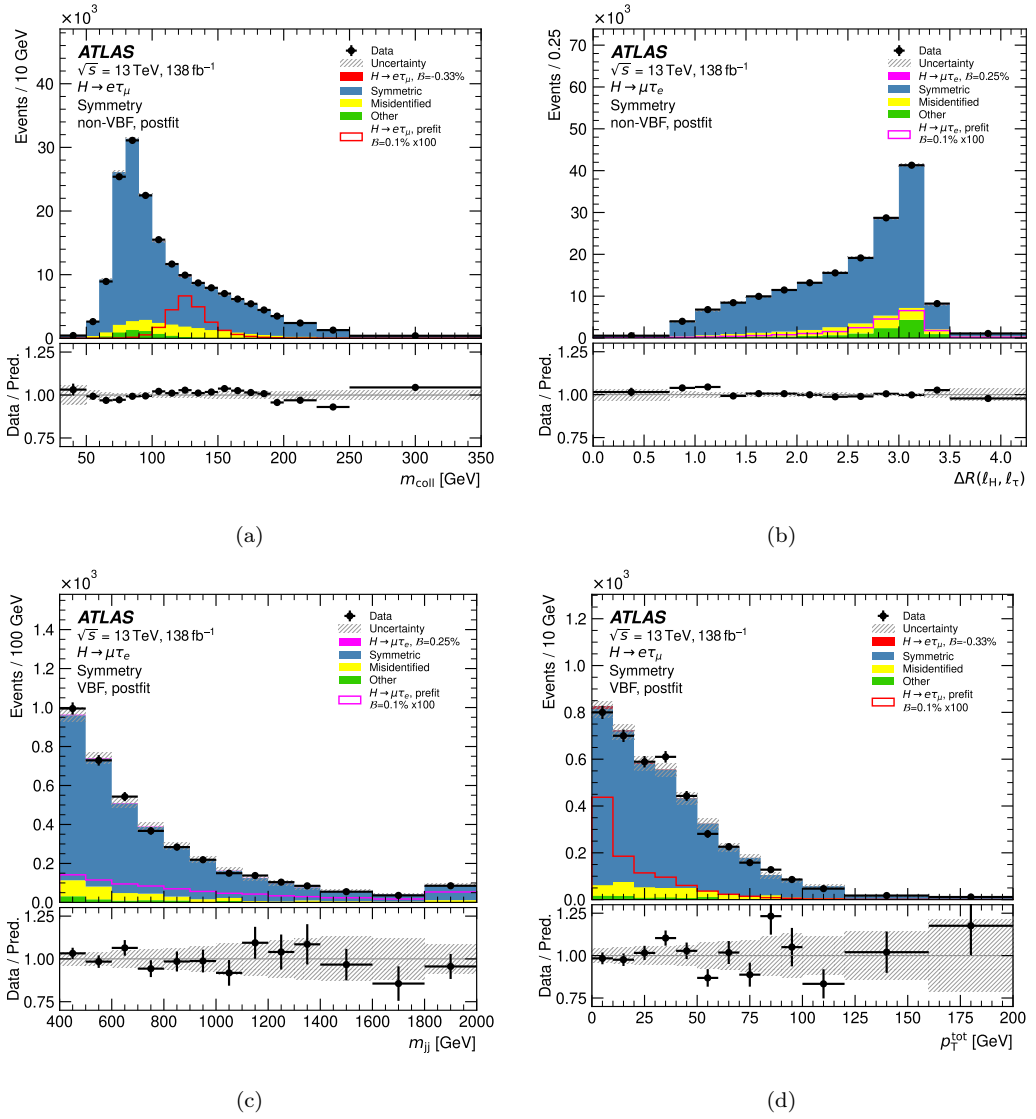
production. These are estimated using the FF method, where the fake factors are determined in a $Z+$jets control region enriched in misidentified leptons. Correction factors, derived from simulation, account for differences between the $Z+$jets and baseline regions.

Fake factors and correction factors are applied as event weights based on the kinematic properties of non-prompt leptons. Misidentified electron and muon contributions are estimated separately to avoid double counting.

In this analysis, the efficiency for detecting a $\mu\tau_e$ event ($\epsilon^{\mu\tau_e}$) or an $e\tau_\mu$ event ($\epsilon^{e\tau_\mu}$) is determined by the product of the event trigger efficiency and the individual muon and electron efficiencies. The trigger efficiency, along with its systematic uncertainties, follows the procedure in Ref. [208], where scale factors are applied to simulated events to adjust for differences in trigger modeling.

Muon and electron efficiencies are products of reconstruction, identification, and isolation efficiencies. Muon efficiencies, which are unaffected by specific event selections, are measured as described in Ref. [208].

The number of selected events dictates the statistical uncertainty in electron efficiency, while systematic uncertainties are assessed by varying selection criteria, such as switching the tagging lepton or adjusting the p_{T} cut of ℓ_1 to 35 GeV. To match the electron efficiencies in data, scale factors and associated uncertainties from Ref. [204] are applied.



(a) (b)

(c) (d)

Figure C.2: Post-fit distributions of kinematic variables for the $e\tau_\mu$ (left) and $\mu\tau_e$ (right) final states are presented after independent fits of the $H \to e\tau$ and $H \to \mu\tau$ signals, respectively, based on data from the Symmetry $\ell\tau_{\ell'}$ channel. For the non-VBF category, the displayed variables are (a) the collinear mass m_{coll} and (b) the angular separation $\Delta R(\ell_H, \ell_\tau)$ between ℓ_H and ℓ_τ. In the VBF category, (c) the total transverse momentum $p_{\text{T}}^{\text{tot}}$, calculated as the magnitude of the vector sum of the transverse momenta of the two leptons, the two leading jets, and $E_{\text{T}}^{\text{miss}}$, and (d) the invariant mass of the dijet system, m_{jj}, are shown. The hashed band indicates the pre-fit statistical uncertainties, as well as experimental and theoretical uncertainties from the likelihood fit. Pre-fit signal shapes are scaled to assume $\mathcal{B}(H \to \ell\tau) = 0.1\%$ and scaled by a factor of 100 for visibility. Post-fit signal contributions are included in the predicted distributions..

Each event is corrected for lepton efficiency according to its kinematics, with $e\tau_\mu$ events rescaled by the efficiency ratio $R_\epsilon = \epsilon^{\mu\tau_e}/\epsilon^{e\tau_\mu}$ to mimic $\mu\tau_e$ events, and vice versa. This efficiency correction, validated with simulated data, is valid only for events with prompt leptons, while simulated events containing misidentified leptons are removed. In data, contributions from both MC-fakes and FF-fakes are

subtracted before applying the efficiency correction. An example of the application of the efficiency correction is shown in Figure C.1, where the distributions of the collinear mass in the non-VBF and VBF categories are shown for the $e\tau_\mu$ and $\mu\tau_e$ datasets, and also for the $e\tau_\mu$ dataset after applying the efficiency correction. The agreement between the $\mu\tau_e$ and $e\tau_\mu$ datasets after the efficiency correction is very good, indicating that the efficiency correction can restore the symmetry between the two datasets.

The number of background events in each dataset ($e\tau_\mu$ and $\mu\tau_e$) is estimated as the sum of three contributions: a symmetric component, determined by applying the efficiency correction to events in the opposite dataset; MC-fakes; and FF-fakes. Since the efficiency correction is valid only for prompt leptons, events with misidentified leptons are removed from the opposite dataset prior to applying the correction.

Figure C.2 compares the data with the predicted background for selected kinematic variables exhibiting strong signal-background discrimination in the $e\tau_\mu$ and $\mu\tau_e$ datasets, within both the non-VBF and VBF categories.The signal and background normalizations are determined through independent fits of the $H \to e\tau$ and $H \to \mu\tau$ signals (discussed in Sections C.4.1), using only $\ell\tau_{\ell'}$ final-state data. The distributions show overall agreement with data within statistical and systematic uncertainties.

C.3 Multivariate techniques in the Symmetry-based $\ell\tau_{\ell'}$ channel

In the $\ell\tau_{\ell'}$ channel of the Symmetry-based method, the m_{jj} threshold in the VBF category has been reduced to 300 GeV to enhance the size of the training dataset. For the non-VBF category, a multiclass classifier is implemented, which yields three distinct output classes: signal, flavor-symmetric SM background, and misidentified background. The contribution from MC-fakes is integrated into the flavor-symmetric background class, as the distribution characteristics of its primary processes—$Z \to \mu\mu$, $Z \to \tau\tau$, and $V\gamma$—are more similar to those of the flavour-symmetric SM background processes rather than those from jets that have been misidentified as leptons.

In the VBF category, three separate binary classification networks are developed. The outputs from these networks are subsequently combined through a linear combination to produce a single discriminant. Each of the three networks is specifically trained to differentiate signal events from different backgrounds: MC-fakes, $Z \to \tau\tau$, and $H \to \tau\tau$ ($\text{NN}_1^{\ell\tau_{\ell'}}$); top-quark events, diboson production, and $H \to WW$ ($\text{NN}_2^{\ell\tau_{\ell'}}$); and FF-fakes ($\text{NN}_3^{\ell\tau_{\ell'}}$). The final discriminant used to evaluate the presence and strength of a signal, is obtained by applying a linear combination of the three networks, as in the MC-based method.

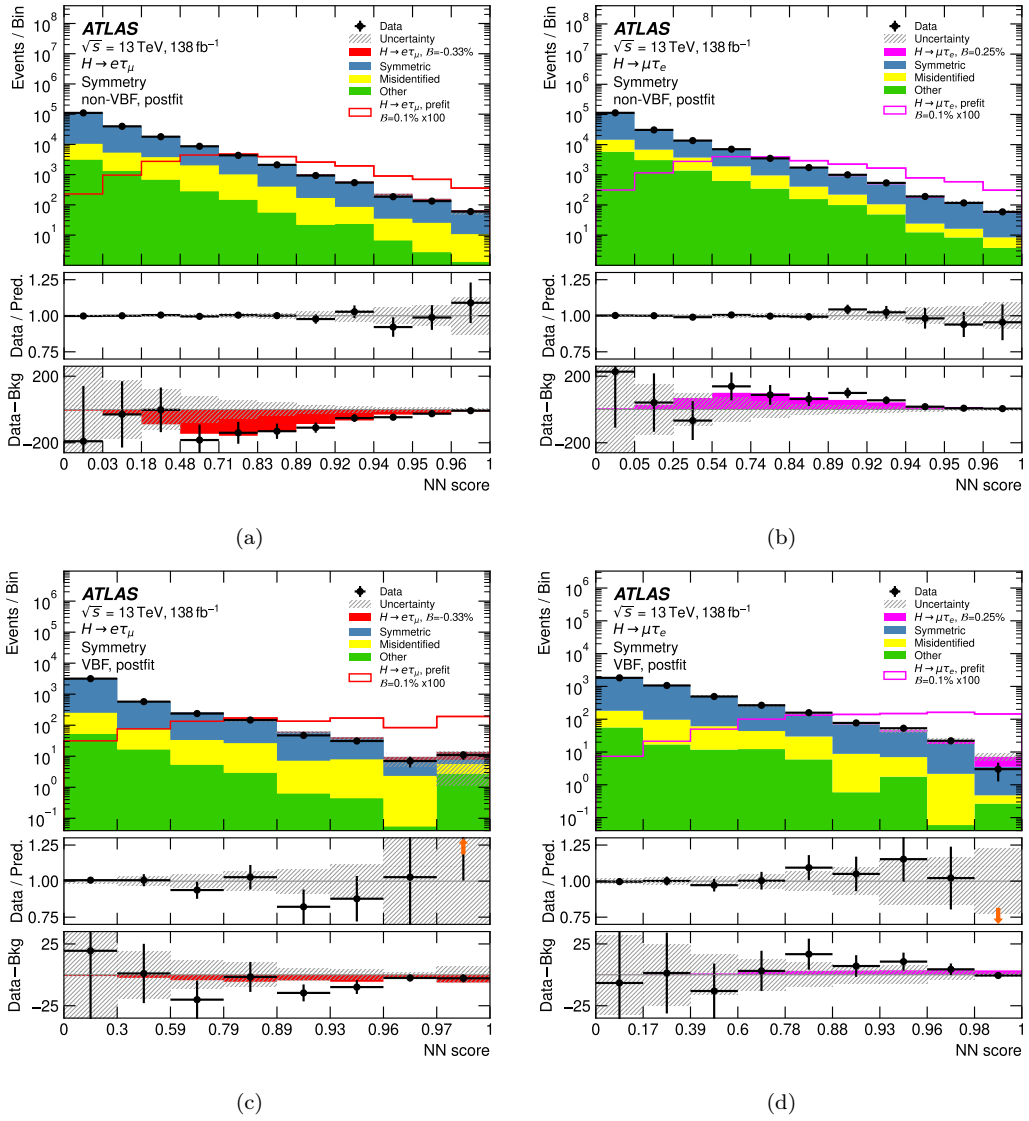


Figure C.3: The NN score distributions for the final states $e\tau_\mu$ (left) and $\mu\tau_e$ (right) are presented for the non-VBF (top) and VBF (bottom) categories within the Symmetry $\ell\tau_{\ell'}$ channel. The central panel displays the ratio of data to predicted yields (both background and signal), while the lower panel shows the residuals between the data and background post-fit. The hashed band represents the total post-fit uncertainty on the overall predicted yields. Each sample's prediction is derived from a likelihood fit performed to measure the $H \to e\tau$ (left) and $H \to \mu\tau$ (right) signals, as described in Section C.4.1. The binning matches that utilized in the likelihood fit, and points beyond the y-axis range are indicated by arrows. The overlayed pre-fit signal shapes, assuming $B(H \to \ell\tau) = 0.1\%$, are scaled by a factor of 100 for visibility. Post-fit signal contributions are incorporated into the predictions. Due to the interpretation in the fits of the branching ratio differences as individual branching ratios, the upward deviation in $\mu\tau e$ corresponds to a downward deviation in $e\tau\mu$.

The input variables used for the NNs are similar to those used for the BDTs in the MC-based $\ell\tau_{\ell'}$ channel. For the non-VBF category, the transverse masses are computed using the lepton from the Higgs boson and the tau lepton, as well as the $E_{\mathrm{T}}^{\mathrm{miss}}$, instead of the leading and subleading leptons. Three ratios between

transverse momenta are included: the ratio between the p_{T} of the lepton from the Higgs boson and the lepton from the tau-decay, $p_{\mathrm{T}}^{\ell_H}/p_{\mathrm{T}}^{\ell_\tau}$; the ratio between the p_{T} of the lepton from the Higgs boson and the $E_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\ell_H}/E_{\mathrm{T}}^{\mathrm{miss}}$; and the ratio between the sum of the lepton from the tau-decay and the $E_{\mathrm{T}}^{\mathrm{miss}}$and the p_{T} of the lepton from the Higgs boson, $(p_{\mathrm{T}}^{\ell_\tau} + E_{\mathrm{T}}^{\mathrm{miss}})/p_{\mathrm{T}}^{\ell_H}$. In addition to $\Delta R(\ell_H, \tau_{\ell'})$ (used also in the MC-based method), the η and ϕ separations between the lepton from the Higgs boson and the lepton from the tau-decay, $\Delta\eta(\ell_H, \tau_{\ell'})$ and $\Delta\phi(\ell_H, \tau_{\ell'})$, are included. The significance of the transverse impact parameter of the lepton from the tau-decay, $\sigma_{d_0}^{\ell_\tau}$, is not used in the training of the NNs.

For the VBF category, in addition to the variables used in the non-VBF category, the variables presented in the MC-based VBF category are included. Furthermore, the training also incorporates as input the p_{T} of the two leading jets.

Figure C.3 shows the NN score distributions for the Symmetry-based $\ell\tau_{\ell'}$ channel. The agreement between data and the prediction is good within the total uncertainties. The predicted signal and background distributions correspond to the post-fit values obtained from the individual fits of the $H \to e\tau$ and $H \to \mu\tau$ signals, using only the Symmetry-based method. The Symmetry method, due to the use of one sample as background estimate of the other, is sensitive to the differences in the signal strength between the two LFV processes. However, if one of the two signals is not present (i.e. $H \to e\tau$ is assumed to be zero in the $H \to \mu\tau$ search, or vice versa), the results can be interpreted as individual branching ratios. The figures show the results of the individual fits, where the upward deviation in $\mu\tau e$ corresponds to a downward deviation in $e\tau\mu$ due to the interpretation of the branching ratio differences as individual branching ratios.

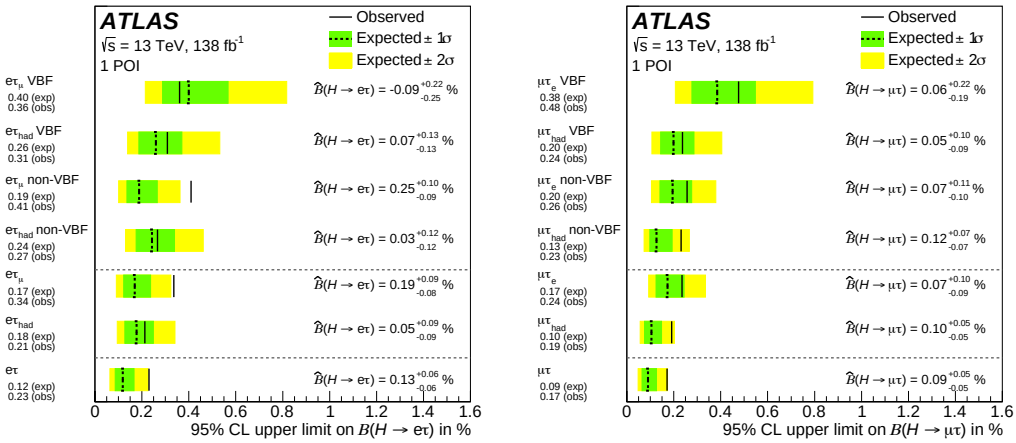
C.4 Results in the Symmetry-based $\ell\tau_{\ell'}$ channel

The Symmetry-based method inherently measures the branching ratio difference, which can then be interpreted as individual branching ratios if one of the signals is assumed to be absent (detailed in Section C.4.1). On the other hand, simultaneous fits from the MC-based method yield branching ratio measurements, providing also the correlation between the two signal strenghts. Therefore, it's possible to compute also the branching ration difference with the MC-based analyses. This fact facilitate direct comparison with the Symmetry-based results. The branching ratio differences, as derived from standalone fits in the $\ell\tau_{\ell'}$ final state under either background estimation method, along with their respective consistencies, are presented in Section C.4.2.

C.4.1 Individual measurements of $H \to e\tau$ and $H \to \mu\tau$

The branching ratio measurements are also performed through single POI fits for the individual $H \to \mu\tau$ and $H \to e\tau$ searches, combining the $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$ final states. The assumption $\mathcal{B}(H \to e\tau) = 0$ applies to $H \to \mu\tau$ analysis and vice versa.

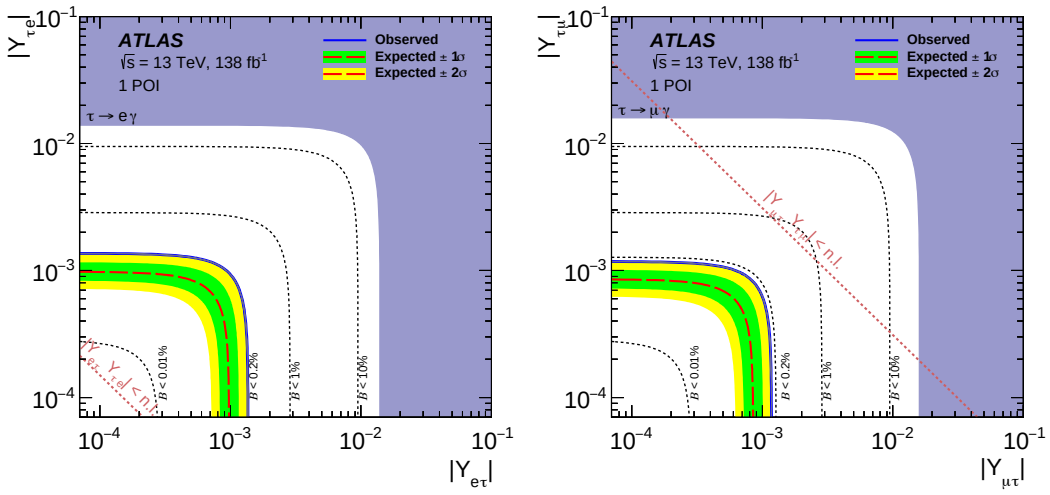
This fitting procedure integrates the $\ell\tau_{\text{had}}$ channel and the non-VBF category from the MC-based $\ell\tau_{\ell'}$ method with the VBF category from the Symmetry-based $\ell\tau_{\ell'}$ approach. The combination of the MC-based and Symmetry-based methods relies on the category-level expected sensitivity using Asimov datasets, where the method with higher significance is selected. The Symmetry-based method yields a sensitivity improvement of 13% and 7% in the VBF category for the $e\tau$ and $\mu\tau$ final states, compared to the MC-based method. Using the $\ell\tau_{\ell'}$ MC-based method in the non-VBF category improves the sensitivity by 28% and 18% for $e\tau$ and $\mu\tau$ final states, with respect to the Symmetry-based method. The $\ell\tau_{\text{had}}$ channel exclusively utilizes the MC-based method.



(a) (b)

Figure C.4: Fit results for the individual 1-POI measurement of the $H \to e\tau$ and $H \to \mu\tau$ signals. 95% CL upper limits on the LFV branching ratios of the Higgs boson are shown. These limits, shown for (a) $H \to e\tau$ and (b) $H \to \mu\tau$, are represented by solid lines for observed results and dashed lines for expected results. The best-fit branching ratio values ($\hat{B}$) are also included in percentage. Each likelihood fit incorporates the channels and categories listed along the y-axis, with signal and control regions from all other channels and categories excluded. Results from individual fits for standalone channels or categories are compared with the combined fit shown in the last row.

Normalization factors are applied to adjust the $Z \to \tau\tau$ production cross-section in the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ SRs estimated with the MC-based method. Specifically, a unified normalization factor is used for both the non-VBF category SR of the $\ell\tau_{\ell'}$ channel and the event yield within the $Z \to \tau\tau$ CR. In the $\ell\tau_{\text{had}}$ channel, separate normalization factors are implemented for the non-VBF and VBF categories, constrained by data within the SRs. Given that top-quark processes represent a minimal background in the $\ell\tau_{\text{had}}$ channel, a top-quark CR is not defined for this channel. Consequently, for the non-VBF category, where the MC-based method is used in both final states, a single normalization factor is applied to scale the top-quark production cross-section across both $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ final states, making use of the $\ell\tau_{\ell'}$ top-quark CR. In the $\ell\tau_{\ell'}$ channel's VBF category, the Symmetry-based



(a) (b)

Figure C.5: The expected (red long-dashed line) and observed (solid blue line) 95% CL upper limits from the simultaneous 1-POI fit of the two searches on the absolute values of the couplings $Y_{\tau\ell}$ and $Y_{\ell\tau}$ are shown, alongside the most stringent indirect limits from $\tau \to \ell\gamma$ searches (dark purple region) for (a) $\ell = e$ or (b) $\ell = \mu$. The short-dashed lines indicate limits corresponding to different branching ratios (0.01%, 0.2%, 1%, and 10%), while the dotted line marks the naturalness limit (denoted by n.l.).

method does not allow to determine a top-quark normalization via a CR for the $\ell\tau_{\text{had}}$ final state, so uncertainties on the top-quark cross-section are incorporated for the $\ell\tau_{\text{had}}$ channel.

The results, shown in Figure C.4, display the 95% CL upper limits for the $H \to e\tau$ and $H \to \mu\tau$ searches, along with independent fits from each channel. An excess corresponding to a 1.9σ significance is observed for the $H \to \mu\tau$ signal and 2.2σ for the $H \to e\tau$ signal. In the $H \to e\tau$ search, this excess arises predominantly from the non-VBF category in the $\ell\tau_{\ell'}$ final state, with inputs from the MC-based method, while the $H \to \mu\tau$ excess is mainly driven by the non-VBF category in the $\ell\tau_{\text{had}}$ final state.

The LFV Higgs boson decay branching ratios relate to non-diagonal elements in the Yukawa coupling matrix. Figure C.5 presents the limits for individual matrix elements $Y_{\tau\ell}$ and $Y_{\ell\tau}$ from independent fits for the $H \to e\tau$ and $H \to \mu\tau$ searches. The figure also includes comparison limits from $\tau \to \ell\gamma$ searches, as in the 2-POI figure presented before.

The impact of the different sources of systematic uncertainties on the observed branching ratios are shown in Table C.1. The largest contributions come from the background sample size, the data sample size and the systematic uncertainties on the misidentified backgrounds, as in the 2-POI fit setup.

1 POI	Impact on observed $[10^{-4}]$	
Source of uncertainty	$\hat{\mathcal{B}}(H \to e\tau)$	$\hat{\mathcal{B}}(H \to \mu\tau)$
Flavour tagging	0.6	0.4
Misidentified background ($\ell\tau_{\text{had}}$)	2.1	1.5
Misidentified background ($\ell\tau_{\ell'}$)	2.9	1.6
Jet and $E_{\text{T}}^{\text{miss}}$	1.1	1.1
Electrons and muons	0.2	0.5
Luminosity	0.6	0.5
Hadronic τ decays	0.9	1.0
Theory (signal)	0.9	0.7
Theory (Z + jets processes)	1.0	1.2
Theory (top-quark processes)	0.3	0.3
Theory (diboson processes)	0.4	0.7
$Z \to \ell\ell$ normalization	0.2	0.7
Symmetric background estimate	0.2	0.1
Background sample size	4.2	2.4
Total systematic uncertainty	5.3	3.9
Data sample size	2.9	2.7
Total	6.1	4.7

Table C.1: Impact of the different sources of systematic uncertainties on the observed branching ratios for the $H \to e\tau$ and $H \to \mu\tau$ signals using the 1 POI fit.

C.4.2 Branching ratio difference in the $\ell\tau_{\ell'}$ channel

The Symmetry-based method uses the $e\tau_\mu$ dataset to estimate the background for the $H \to \mu\tau$ search and the $\mu\tau_e$ dataset for the $H \to e\tau$ search. This approach provides a measurement of the difference between the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ by evaluating $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ in the $H \to \mu\tau$ search and $\mathcal{B}(H \to e\tau) - \mathcal{B}(H \to \mu\tau)$ in the $H \to e\tau$ search. Among these measurements, $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ is chosen as the primary result from the Symmetry-based method due to its smaller expected uncertainty. Combining the non-VBF and VBF categories, the Symmetry-based method determines $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau) = (0.25 \pm 0.10)\%$, showing compatibility with zero within 2.5σ.

For the MC-based method, the branching ratio difference is derived from the simultaneous fit of $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, as outlined in Section 7.8.1, but using only data from the $\ell\tau_{\ell'}$ final state. This approach accounts for the correlation between the two branching ratio parameters. The measured difference $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau) = (0.02 \pm 0.12)\%$ in the combined non-VBF and VBF categories aligns well with zero, within 1σ. Figure C.6 presents the results for the non-VBF and VBF categories both individually and combined. While the

Symmmetry-based method suggests a larger branching ratio for the $H \to \mu\tau$ signal over the $H \to e\tau$ signal with a significance of 2.5σ—mainly from the non-VBF category, the MC-based method finds a branching ratio difference that is compatible with zero, also influenced primarily by the non-VBF category.

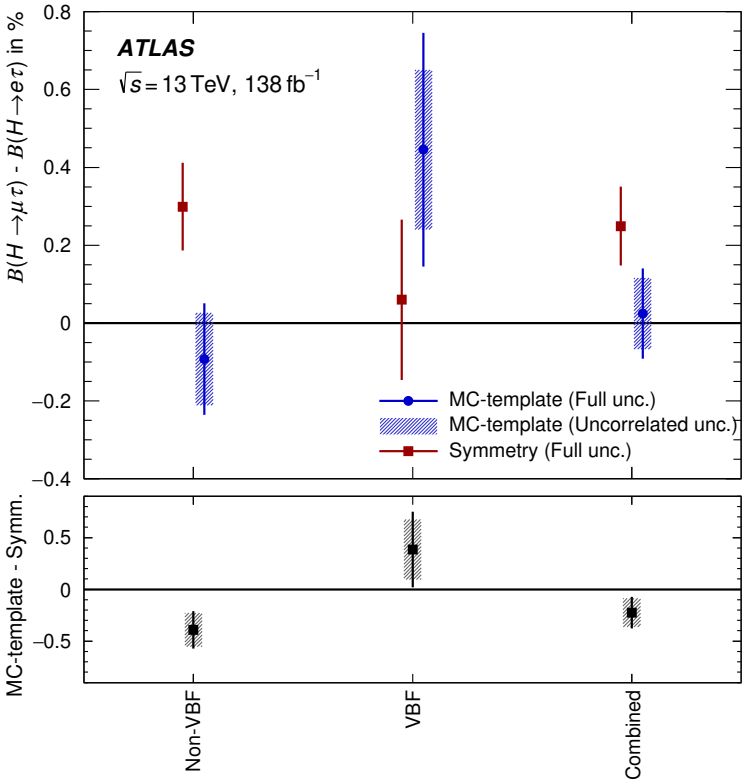


Figure C.6: Best-fit values of $B(H \to \mu\tau) - B(H \to e\tau)$, given in percentage, obtained from the standalone fits of the $\ell\tau\ell'$ final state using either background estimation method. The Symmetry-based result is obtained from the $\mu\tau_e$ 1-POI fit, while the MC-based result is derived from the 2-POI fit. The uncertainty of the difference shown in the lower panel is calculated by considering either only the sources of uncertainty that are uncorrelated between the two analyses (grey bars) or all uncertainties (black lines).

The comparison of branching ratio differences between the Symmetry-based and MC-based methods in the $\ell\tau_{\ell'}$ final state allows a compatibility cross-check of the results from each approach. Compatibility is tested between the methods, accounting for shared uncertainties due to overlapping datasets and common signal simulation samples. Statistical uncertainties are correlated across methods due to the extensive data overlap, while signal uncertainties are similarly correlated. All other uncertainties are assumed uncorrelated. For the MC-based method, correlated nuisance parameters are fixed at their best-fit values, and the uncertainties on the branching ratio differences are recalculated using only uncertainty sources that are uncorrelated between the two methods. The Symmetry-based method includes

the full uncertainty. As shown in Figure C.6, the results from both methods agree within a 2σ level.

C.5 Summary

The Symmetry-based method provides a data-driven approach to estimate the background in the Higgs boson LFV decay searches. The method relies on the symmetry between electrons and muons in the SM background, which would be broken by the presence of LFV decays. In the analysis presented in this thesis, the Symmetry-based method is applied to the $\ell\tau_{\ell'}$ final state.

The analysis measured the branching ratio difference $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ using the symmetry-based method within the $\ell\tau_{\ell'}$ final state, favoring a slightly higher branching ratio for $H \to \mu\tau$. The best-fit value for this difference was determined to be $(0.25 \pm 0.10)\%$, although this result remains compatible with zero within 2.5σ. The independent MC-based estimation in the $\ell\tau_{\ell'}$ final state yielded a difference between the two branching ratios of $(0.02 \pm 0.12)\%$, consistent with the SM prediction of zero. The two results are compatible within a 2σ level.

Resumen

El Modelo Estándar (SM, por sus siglas en inglés) de la física de partículas proporciona el marco teórico para entender las partículas fundamentales y las fuerzas que las gobiernan, con excepción de la gravedad. Esta teoría cuántica de campos relativista describe el comportamiento e interacciones de los fermiones, que constituyen la materia, y de los bosones gauge, que median las interacciones electromagnética, débil y fuerte. Aunque el SM se desarrolló a lo largo de décadas mediante una interacción entre descubrimientos experimentales y avances teóricos, solo se completó en 2012 con el descubrimiento del bosón de Higgs por las colaboraciones ATLAS y CMS en el Gran Colisionador de Hadrones (LHC, por sus siglas en inglés) del CERN. Este descubrimiento confirmó la existencia del campo de Higgs, un campo escalar responsable de otorgar masa a las partículas elementales mediante el mecanismo de Higgs, basado en una ruptura espontánea de la simetría electrodébil.

El descubrimiento de un bosón de Higgs con una masa en torno a 125 GeV no solo validó el SM, sino que también abrió la posibilidad de realizar medidas precisas de sus canales de producción y desintegración, una herramienta clave para poner a prueba las predicciones teóricas y explorar posibles desviaciones de las medidas experimentales respecto a las mismas. A día de hoy, las propiedades medidas del bosón de Higgs son consistentes con las predicciones del SM. A pesar de este éxito, el SM presenta limitaciones. Estas limitaciones impulsan la búsqueda de nueva física más allá del SM, en la cual el bosón de Higgs es una herramienta fundamental debido a sus posibles conexiones con fenómenos más allá del SM.

Esta tesis explora dos vías principales en el estudio del sector de Higgs. La primera es la medida de secciones eficaces diferenciales en el canal de desintegración $H \to \tau^+\tau^-$. Este modo de desintegración ofrece un acceso directo al acoplamiento de Yukawa del Higgs a los leptones, permitiendo analizar tanto las predicciones del SM como posibles modificaciones debidas a nueva física. Además, ofrece una sensibilidad única a la producción del bosón de Higgs mediante fusión de bosones vectoriales.

El análisis de las secciones eficaces diferenciales en este canal de desintegración utiliza el marco de las *Simplified Template Cross-Sections* (STXS), que permite medir la producción del bosón de Higgs en regiones específicas del espacio de fase definidas por la cinemática del propio bosón de Higgs y de las partículas asociadas al mismo en la producción. El marco STXS, junto con la mayor luminosidad obtenida en el conjunto de datos de Run 2 del LHC a $\sqrt{s} = 13$ TeV, amplía la sensibilidad

de las medidas relacionadas con el bosón de Higgs más allá de enfoques inclusivos, en los que únicamente se miden las secciones eficaces totales de producción. Este análisis realiza, por primera vez en el experimento ATLAS, una medida de la sección eficaz diferencial fiducial en el canal de desintegración $H \to \tau^+\tau^-$. Los resultados de esta medida se interpretan en el contexto de una teoría efectiva de campos (EFT, por sus siglas en inglés), que permite estudiar posibles desviaciones de las predicciones del SM en términos de operadores efectivos.

La segunda vía explorada en esta tesis es la búsqueda de desintegraciones del bosón de Higgs que violen el sabor leptónico (LFV, por sus siglas en inglés) a los estados finales $e\tau$ y $\mu\tau$. Aunque el SM conserva el sabor leptónico en el sector de leptones cargados, la observación de oscilaciones de neutrinos indica que el sabor leptónico no se conserva en la naturaleza, lo que plantea la posibilidad de que esta violación se manifieste también en las desintegraciones del bosón de Higgs. Detectar estos modos de desintegración constituiría una evidencia directa de física más allá del SM y ofrecería información sobre posibles mecanismos que generen dicha violación, explicables mediante extensiones del SM. Los límites experimentales actuales en las desintegración LFV han restringido sus fracciones de desintegración, aunque la pequeña anchura de desintegración del bosón de Higgs permite contribuciones de física más allá del SM a niveles detectables por el LHC. El análisis presentado en esta tesis realiza una búsqueda de desintegraciones LFV del bosón de Higgs utilizando dos estrategias diferentes para la estimación de fondos: una basada en simulaciones de Monte Carlo (basado en MC) y otra basada en métodos dependientes de los datos (basado en simetría), aprovechando la simetría bajo intercambio de electrones y muones en procesos del SM.

Esta tesis contribuye a tres áreas principales de investigación dentro del experimento ATLAS: el desarrollo de software de simulación para la electrónica de lectura del Calorímetro de Baldosas en preparación para las mejoras de Alta Luminosidad del LHC (HL-LHC); la medida de las secciones eficaces diferenciales de la producción del bosón de Higgs en el canal de desintegración $H \to \tau^+\tau^-$, con un énfasis en las medidas dentro del marco STXS; y la búsqueda de desintegraciones LFV a estados finales $e\tau$ y $\mu\tau$. Estos análisis utilizan el conjunto completo de datos del Run 2, con 140 fb^{-1} de datos de colisiones pp a una energía en el centro de masas de 13 TeV.

Introducción a la física del bosón de Higgs

El Modelo Estándar es una teoría que describe tres de las cuatro interacciones fundamentales conocidas: la electromagnética, la débil y la fuerte. Estas interacciones se estructuran en torno al grupo de simetría gauge $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, donde $SU(3)_C$ describe la interacción fuerte (cromodinámica cuántica, QCD) y $SU(2)_L \otimes U(1)_Y$ unifica las interacciones electromagnética y débil en la teoría electrodébil.

Las partículas de materia son fermiones con espín 1/2 y están organizadas en tres generaciones de leptones y quarks. La primera generación incluye el electrón (e), el neutrino electrónico (ν_e), el quark up (u) y el quark down (d), mientras que las generaciones sucesivas repiten esta estructura con partículas más masivas. Los bosones de gauge son partículas de espín 1 que median las interacciones y consisten en ocho gluones para la QCD, tres bosones débiles ($W^{\pm}$ y Z) y el fotón para la interacción electromagnética.

El sector electrodébil se basa en el grupo de simetría $SU(2)_L \otimes U(1)_Y$ [10-12]. La simetría gauge impone que los bosones vectoriales y fermiones quirales sean originalmente sin masa, ya que cualquier término de masa violaría la invariancia gauge y comprometería la renormalizabilidad del modelo [1]. El mecanismo de Higgs resuelve este problema introduciendo un campo escalar que, al adquirir un valor de expectación en el vacío, rompe espontáneamente la simetría $SU(2)_L \otimes U(1)_Y$ y otorga masa a los bosones $W^{\pm}$ y Z sin afectar al fotón, que permanece sin masa.

La interacción fuerte es descrita por la QCD, una teoría basada en el grupo de simetría $SU(3)_C$, en la cual los quarks interactúan mediante gluones, los mediadores de la fuerza fuerte. A diferencia de las interacciones débiles y electromagnéticas, los gluones son portadores de carga de color y, por lo tanto, pueden interactuar entre sí [5]. La QCD presenta dos fenómenos fundamentales únicos: la libertad asintótica a altas energías y el confinamiento a bajas energías, el cual restringe a los quarks y gluones a formar hadrones. Un problema significativo en la QCD es el problema CP fuerte: la simetría gauge permite un término en el lagrangiano que viola la simetría CP, pero las medidas experimentales indican que este término debe estar suprimido mediante un parámetro θ_{QCD} extremadamente pequeño.

El mecanismo de Higgs [2-4] introduce en el lagrangiano del SM un potencial para el campo de Higgs, Φ, con la forma $V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$, donde $\mu^2 < 0$ y $\lambda > 0$. Esto da lugar a un mínimo no trivial del potencial, con un valor de expectación en el vacío de $v = \sqrt{-\mu^2/\lambda}$. Este valor de expectación no nulo permite la ruptura espontánea de la simetría electrodébil, generando masas para los bosones $W^{\pm}$ y Z. La incorporación de términos de Yukawa en el lagrangiano permite que los fermiones adquieran masa mediante su interacción con el campo de Higgs. El mecanismo de Higgs es crucial para la consistencia del SM, ya que explica la masa de las partículas sin introducir términos de masa explícitos en el lagrangiano. Su existencia se manifiesta con la predicción de un bosón escalar neutro, el bosón de Higgs. Este bosón se acopla a las partículas del SM proporcionalmente a sus masas (linearmente en el caso de los fermiones, y cuadráticamente en el caso de los bosones de gauge).

En el LHC, el bosón de Higgs se produce principalmente mediante fusión de gluones (ggF). Este proceso representa alrededor del 87 % de la producción total. Los otros modos de producción relevantes son la fusión de bosones vectoriales (VBF), la producción asociada a un bosón vectorial (VH) y la producción en asociación con quarks top (ttH). Cada uno de estos procesos tiene una sección eficaz calculada a órdenes superiores en QCD y electrodébil. En particular, la ggF

se calcula hasta una precisión a N^{3}LO en QCD, mientras que VBF y VH están disponibles con una precisión de NNLO en QCD y NLO en cuanto a correcciones electrodébiles.

El bosón de Higgs predicho por el Modelo Estándar se desintegra en pares de leptones τ con una probabilidad de aproximadamente el 6 %. Este canal de desintegración es importante porque permite medir el acoplamiento de Yukawa del Higgs con los fermiones. También ofrece una buena sensibilidad al modo de producción por fusión de bosones vectoriales (VBF) y permite estudiar las propiedades CP del Higgs en su producción y desintegración. Sin embargo, plantea retos experimentales debido a los neutrinos que se producen en las desintegraciones del τ, que dificultan la reconstrucción del estado final, y a los altos niveles de fondo provenientes del proceso $Z \to \tau\tau$, con una sección eficaz varias órdenes de magnitud mayor que el de la producción total de bosones de Higgs.

Durante Run 1 del LHC (2010-2012), los experimentos ATLAS y CMS realizaron las primeras medidas de las propiedades del Higgs, determinando su masa en $m_H = 125.09 \pm 0.24$ GeV [49]. La razón entre la sección eficaz medida y la predicha por el SM fue $\mu = 1.09 \pm 0.11$ [50]. Se observó la producción del Higgs por ggF y su desintegración a canales como $\gamma\gamma$, ZZ, WW y $\tau\tau$ [51].

Con los datos recolectados durante Run 2 del LHC (2015-2018), que corresponden a 140 fb^{-1} de colisiones pp a $\sqrt{s} = 13$ TeV, se logró mejorar la precisión en las mediciones de las propiedades del Higgs y explorar señales de nueva física. La masa del Higgs se midió con mayor precisión, alcanzando un valor de $m_H = 125.11 \pm 0.11$ GeV en ATLAS utilizando los canales $H \to \gamma\gamma$ y $H \to ZZ^*$ [56]. La anchura total del Higgs, que el SM predice como $\Gamma_H = 4.07$ MeV para $m_H = 125$ GeV [24], no puede medirse directamente debido a la limitada resolución experimental, pero existen límites experimentales utilizando medidas de producción del bosón de Higgs fuera de la capa de masas, con resultados consistentes con el SM [58, 59].

La combinación de análisis de ATLAS más reciente [31] reportó una medida del cociente entre la sección eficaz de producción del Higgs y la predicha por el SM de $\mu = 1.05 \pm 0.06$, consistente con las predicciones del SM. Las secciones eficaces de producción y las fracciones de desintegración del Higgs también se midieron con precisión, con resultados en buen acuerdo con las predicciones teóricas.

Las medidas de secciones eficaces fiduciales diferenciales permiten obtener información más detallada de la producción del bosón de Higgs que las medidas inclusivas. Además, estas medidas permiten minimzar las extrapolaciones de las medidas, definiendo un volumen fiducial del espacio de fases donde se realiza la medida, manteniendo los criterios de selección de los objetos reconstruídos similares a los empleados en la definición del volumen fiducial, maximizando la aceptancia y eficiencia del análisis. Estas medidas permiten minimizar la dependencia con los modelos teóricos, pero requieren de la utilización de criterios de selección simples para mantener una buena alineación entre la selección de eventos y la definición del volumen fiducial.

Para superar estas restricciones y lograr un compromiso entre sensibilidad y dependencia con los modelos teóricos, se utiliza el marco de las *Simplified Template Cross-Sections* (STXS) [63]. Este divide el espacio de fases en regiones definidas por variables como el momento transversal del Higgs p_{T}^{H} o el número de jets asociados.

El marco STXS está diseñado para maximizar la sensibilidad a efectos de nueva física, minimizando al mismo tiempo las incertidumbres teóricas. Las mediciones no necesitan corregir directamente los efectos del detector mediante técnicas de $unfolding$ a un volumen fiducial complejo, permitiendo el uso de técnicas multivariantes avanzadas para canales desafiantes como $H \to \tau\tau$ y $H \to b\bar{b}$. Aunque la posibilidad de utilizar técnicas de análisis multivariante es una ventaja, la ausencia de una definicón de volumen fiducial próximo al espacio de fases reconstruido implica un mayor nivel de extrapolación. Sin embargo, el marco STXS sí proporciona un volumen fiducial simplificado que permite realizar combinaciones entre los diferentes análisis que se enfocan en modos de desintegración específicos. Además, la definición común de las regiones STXS (los bines de las variables con respecto a las cuales se miden las secciones eficaces diferenciales) facilita la combinación de resultados de diferentes canales de desintegración del Higgs, optimizando la sensibilidad experimental.

El marco clasifica la producción del Higgs en cuatro categorías principales: producción por fusión gluón-gluón ($gg \to H$), producción por VBF ($qq' \to qq'H$), producción asociada con bosones vectoriales que decaen a leptones ($VH \to (\ell\ell, \ell\nu)H$) y producción con pares $t\bar{t}$ (ttH). Cada categoría se subdivide según variables como el número de jets o las masas invariantes de los dos jets con mayor p_{T}. Esta granularidad puede ajustarse según las necesidades experimentales, fusionando regiones con baja sensibilidad o dividiéndolas cuando haya suficiente precisión.

La última combinación de datos de ATLAS en el marco STXS [31] incluye medidas en 36 regiones cinemáticas exclusivas. El canal de desintegración $H \to \tau^{+}\tau^{-}$ mostró una sensibilidad significativa en la producción por VBF. Parte de esta tesis se centra en extender estas mediciones dentro del marco STXS en dicho canal.

La conservación del sabor leptónico en el SM es accidental, no siendo una simetría impuesta por la invariancia gauge. En ausencia de neutrinos dextrógiros, el SM asegura dicha conservación, pero esto implica que los neutrinos son partículas sin masa. Sin embargo, la observación de oscilaciones de neutrinos por Super-Kamiokande [67] y por el *Sudbury Neutrino Observatory* [68, 69] demostró que los neutrinos tienen masa. En el SM, las matrices de Yukawa que generan las masas de fermiones se diagonalizan mediante rotaciones unitarias en el espacio de sabor, lo que garantiza la ausencia de corrientes neutras que cambien de sabor y la conservación del sabor leptónico. No obstante, en extensiones mínimas del SM que incluyen neutrinos masivos, surgen mecanismos de violación del sabor leptónico debido al desalineamiento entre las matrices de Yukawa de los leptones cargados y los neutrinos. La matriz PMNS de manera análoga a la matriz CKM en el sector de los quarks.

En el contexto de un SM extendido con neutrinos masivos, procesos como las desintegraciones $\mu \to e\gamma$, $\tau \to \mu\gamma$ y $\tau \to e\gamma$ son posibles, aunque los procesos están altamente suprimidos. Por otro lado, diversos modelos más allá del SM, como las extensiones supersimétricas [70, 71], modelos con más de un doblete de Higgs [72], teorías de Higgs compuesto [73] o modelos de dimensiones extra compactificadas [74, 75], predicen desintegraciones del Higgs que violan el sabor leptónico. Las búsquedas de estas desintegraciones en el LHC no requieren asumir un modelo específico y pueden interpretarse de manera independiente a través de límites efectivos en los acoplamientos de Yukawa no diagonales, como se describe en el capítulo 7.

La parametrización efectiva de estos acoplamientos permite establecer límites en los elementos fuera de la diagonal de la matriz de Yukawa mediante medidas indirectas de procesos de baja energía, como $\mu \to e\gamma$ o $\mu \to e$ en núcleos. Estos límites se traducen en restricciones para las desintegraciones $H \to e\mu$, $H \to e\tau$ y $H \to \mu\tau$ del Higgs. Por ejemplo, los valores actuales para $\mathcal{B}(\mu \to e\gamma)$ [81] y $\mathcal{B}(\mu \to e)_{\mathrm{Au}}$ [80] permiten restringir el producto $\mathcal{B}(H \to e\tau) \times \mathcal{B}(H \to \mu\tau)$ a órdenes de 10^{-4}. Las desintegraciones $H \to e\tau$ y $H \to \mu\tau$ son particularmente relevantes, ya que sus límites indirectos son menos restrictivos, lo que las convierte en objetivos viables para estudios en el LHC, donde los límites directos pueden superar las restricciones indirectas.

Tanto ATLAS como CMS han realizado búsquedas directas de desintegraciones del Higgs que violan el sabor leptónico utilizando los datos de Run 1 y Run 2. Las búsquedas no han observado excesos significativos, pero han establecido límites superiores para los acoplamientos y las fracciones de desintegración de estos procesos. Los límites directos más estrictos para $H \to e\mu$ [82, 83] aún no son competitivos con los límites indirectos (aunque proporcionan resultados independientes de cualquier modelo teórico específico), mientras que los resultados para $H \to e\tau$ y $H \to \mu\tau$ [84, 85] muestran una sensibilidad suficiente para sondear escenarios de violación del sabor leptónico en el sector del Higgs más allá de la que pueden proporcionar las restricciones indirectas.

A pesar de haber pasado más de una década desde el descubrimiento del bosón de Higgs, su estudio sigue siendo un tema abierto. Las medidas actuales han alcanzado una precisión inferior al 10 % para los acoplamientos del Higgs con bosones gauge y entre el 7 % y el 20 % para los acoplamientos con fermiones de tercera generación. Mejorar esta precisión es esencial para investigar posibles extensiones del SM que podrían alterar las secciones eficaces de producción y las fracciones de desintegración del bosón de Higgs debido a efectos de nueva física más allá de la escala de energía del LHC. El programa de física del HL-LHC permitirá alcanzar precisiones del orden del 1 % en estos acoplamientos [86]. Quedan desafíos pendientes, como la confirmación de los acoplamientos del Higgs con muones y quarks charm, o la medida de la producción de pares de Higgs para acceder al autoacoplamiento trilineal, clave para entender el mecanismo de ruptura de la simetría electrodébil. La medida de los acoplamientos a quarks up, down y strange,

a pesar de su interés e importancia, queda más allá del alcance de las medidas deñ HL-LHC.

El sector de Higgs también es una ventana hacia nueva física más allá del SM, con implicaciones cosmológicas significativas. Por ejemplo, modificaciones en el potencial del Higgs podrían permitir una transición de fase de primer orden en el universo temprano, necesaria para explicar la bariogénesis electrodébil [90, 91], o jugar un rol en la inflación cósmica [92]. Asimismo, el potencial del Higgs está relacionado con la estabilidad del vacío y podría conectarse con una explicación para la asimetría materia-antimateria que se manifiesta en el universo observable. El sector del Higgs podría proporcionar nuevas fuentes de violación CP [94]. Por otro lado, el sector de Yukawa podría explicar las jerarquías de masa entre fermiones [95, 96]. Otros problemas actuales que podrían tener una solución relacionada con el sector del Higgs son el origen de las masas de los neutrinos [97] o la interacción de partículas de materia oscura con el contenido del SM. El sector del Higgs podría acoplarse a partículas de materia oscura, proporcionando acceso experimental a este enigma cosmológico [98, 99].

El descubrimiento del bosón de Higgs en 2012 marcó un hito en la física de partículas, pero también abrió una nueva era en la que la precisión en la medida de sus propiedades puede proporcionar información crucial sobre la naturaleza de la materia y las fuerzas fundamentales. La exploración del sector de Higgs en el LHC y en futuros colisionadores de partículas es esencial para comprender la física más allá del SM y responder a algunas de las preguntas más fundamentales de la física de partículas. Los resultados experimentales obtenidos hasta la fecha han confirmado las predicciones del SM, pero la búsqueda de desviaciones con respecto a estas predicciones sigue siendo un objetivo fundamental en el campo.

El experimento ATLAS

El detector ATLAS [102], en el LHC del CERN [100], es un experimento de propósito general diseñado para analizar partículas generadas en colisiones protón-protón con energías de hasta 14 TeV. Está compuesto por varios subsistemas organizados en capas alrededor del punto de interacción, cada uno de los cuales mide características específicas de las partículas emitidas.

ATLAS emplea un sistema de coordenadas cuyo origen se encuentra en el punto de interacción nominal. En este sistema, el eje z sigue la dirección del tubo de haces, el eje x apunta desde el punto de interacción hacia el centro del anillo del LHC y el eje y se orienta hacia arriba. Los ángulos azimutal ϕ y polar θ, junto con la pseudorapidez $\eta = -\ln(\tan(\theta/2))$, se utilizan para describir direcciones y distribuciones angulares de las partículas en el detector.

El subsistema más próximo al punto de interacción es el detector interno (ID), que mide con precisión las trayectorias de partículas cargadas. Consiste en tres subdetectores: el Detector de Pixeles, el *Semiconductor Tracker* (SCT) y *Transition*

Radiation Tracker (TRT) [107, 108]. Los dos primeros emplean tecnología de silicio para registrar las trayectorias, mientras que el TRT utiliza cámaras de gas con tubos de tracción. Este sistema opera bajo un campo magnético de 2 T, permitiendo determinar el momento y la carga de las partículas.

El calorímetro electromagnético de ATLAS mide la energía de electrones y fotones. Utiliza argón líquido (LAr) como medio activo y plomo como material absorbente, en un sistema de muestreo compuesto por un barril y dos tapas (*end-caps*), cubriendo un rango de pseudorapidez de $|\eta| < 3.2$ [111]. Las placas en forma de acordeón proporcionan una cobertura sin fisuras en el ángulo azimutal. La resolución de energía en este calorímetro es $\sigma_E/E = 10\,\%/\sqrt{E/\mathrm{GeV}} \oplus 0.7\,\%$, con una profundidad de 22 longitudes de radiación en el barril.

Rodeando el calorímetro electromagnético, el calorímetro hadrónico mide la energía depositada por partículas hadrónicas que no son absorbidas por el calorímetro electromagnético. Su componente central, el calorímetro de tejas (Tile Calorimeter, TileCal) [112], está compuesto de acero y tejas de plástico centelleador, organizadas en un barril y dos extensiones del mismo, permitiendo la absorción de partículas hadrónicas mediante la luz emitida por las tejas centelleadoras. Esta luz es captada y convertida en señales eléctricas por tubos fotomultiplicadores (PMTs).

TileCal está compuesto de acero y baldosas de plástico centelleador, organizadas en un barril y dos extensiones, permitiendo la absorción de partículas hadrónicas. La luz producida por las baldosas centelleadoras es transmitida mediante fibras ópticas a tubos fotomultiplicadores (PMTs), que convierten la luz en señales eléctricas. El agrupamiento de fibras ópticas que transmiten la luz de distintas tejas define celdas de lectura con una granularidad de $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ en el barril y $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ en las extensiones del barril. Las señales de los PMTs, que leen la luz de una misma celda, se procesan en la tarjeta 3-en-1 [116], donde se amplifican y se moldean, para luego ser digitalizadas en la *Motherboard*. Una vez que un evento es aceptado por el sistema de trigger, las señales digitales son enviadas a la electrónica fuera del detector (*off-detector*) para su procesamiento. Las cuentas de ADC son utilizadas para reconstruir la amplitud del pulso analógico original en los *Readout Drivers* (RODs) utilizando un filtro digital denominado *Optimal Filtering* (OF). La amplitud del pulso, proporcional a la energía depositada en la celda, es posteriormente transformada en energía mediante diversos sistemas de calibración que controlan diferentes efectos en la respuesta del calorímetro.

El espectrómetro de muones [120], en el exterior del detector y ubicado en un campo magnético toroidal de 4 T, detecta los muones que atraviesan los sistemas de calorimetría. El espectrómetro está compuesto por cuatro subsistemas: *Monitored Drift Tubes* (MDT) y *Cathode Strip Chambers* (CSC) para medir el momento, y *Resistive Plate Chambers* (RPC) y *Thin Gap Chambers* (TGC) para generar señales de trigger. La combinación de estos subsistemas permite medir la curvatura de las trayectorias de los muones con una resolución de $\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 10\,\%$ para muones de $p_\mathrm{T} = 1$ TeV.

El sistema de imanes de ATLAS incluye un solenoide de 2 T en el ID que genera un campo en dirección z y un sistema toroidal que genera un campo azimutal en el espectrómetro de muones. Esta configuración optimiza la medida del momento en distintas regiones del detector.

Además de estos subsistemas principales, ATLAS cuenta con detectores adicionales, como LUCID [121] para la medida de luminosidad, AFP [122] y ALFA [123] para la detección de protones dispersados a pequeños ángulos, y ZDC [124] para el análisis de partículas neutras.

El sistema de Triggers y Adquisición de Datos (TDAQ) selecciona y filtra eventos de interés, reduciendo una tasa de 40 MHz a 1 kHz. El trigger de Nivel-1, basado en hardware, analiza información de los sistemas de calorimetría y muones, mientras que el *High-Level Trigger* (HLT) emplea algoritmos de software para reducir aún más la tasa de eventos antes de su almacenamiento.

Finalmente, la Red Mundial de Computación del LHC (WLCG) permite procesar y distribuir los datos generados por ATLAS a nivel global, utilizando una estructura jerárquica de cuatro niveles: Tier-0 en el CERN, Tier-1 en centros nacionales, Tier-2 en universidades e institutos y Tier-3 en instituciones locales. Esta infraestructura facilita el acceso y análisis de los datos a una comunidad amplia de investigadores en todo el mundo.

La mejora del calorímetro TileCal para la fase II del LHC y la simulación de su electrónica de lectura

El proyecto High Luminosity LHC (HL-LHC) [129] busca incrementar la luminosidad instantánea del LHC en un factor de 5 a 7 respecto a su diseño original, con el objetivo de alcanzar una luminosidad integrada de 3000 fb^{-1} al final de su operación. Esto permitirá realizar medidas de precisión y búsquedas de nueva física a una energía en el centro de masa de 14 TeV. Sin embargo, este aumento de luminosidad generará hasta 200 interacciones de protones simultáneas por cruce de bunch, lo que implicará desafíos significativos para los detectores, como el aumento en la ocupación y flujo de partículas.

El detector ATLAS enfrentará estos retos mediante diversas actualizaciones. Entre ellas, el *Inner Tracker* reemplazará completamente al detector interno actual. El nuevo detector estará totalmente basado en silicio para mantener un rendimiento adecuado en la reconstrucción de vértices y trayectorias, a pesar del alto nivel de pile-up. Adicionalmente, se instalará el *High Granularity Timing Detector* en la región delantera para mitigar los efectos del alto pile-up mediante la incorporación de información precisa del tiempo [130].. El sistema de trigger también será mejorado para manejar tasas de hasta 1 MHz y utilizar algoritmos más sofisticados, manteniendo una alta eficiencia en la selección de señales físicas de interés y mejorando la capacidad de rechazo de fondo [131].

En el caso de TileCal, los componentes electrónicos expuestos a la radiación serán reemplazados debido a los niveles de radiación esperados, que alcanzan hasta 1000 Gy de dosis ionizante total y una fluencia de neutrones equivalente de 1 MeV de hasta 10^{13} cm^{-2}. Además, se sustituirán los sistemas de bajo y alto voltaje por versiones más tolerantes a la radiación [132]. La electrónica de lectura incluirá nuevas tarjetas FENICS, que proporcionarán dos rutas de amplificación y conformado de señales para cubrir un amplio rango dinámico. Estas señales serán digitalizadas por las Mainboards y enviadas a través de las Daughterboards al sistema de procesamiento localizado fuera del detector, en la sala de cuentas, compatible con la nueva arquitectura de trigger del HL-LHC.

El nuevo sistema de lectura permitirá una reconstrucción en línea de las amplitudes y tiempos de las señales antes de la decisión del primer nivel de trigger, lo que representa una diferencia fundamental respecto al sistema actual. Este enfoque no solo mejora la compatibilidad del sistema con las tasas de colisiones del HL-LHC, sino que también abre nuevas posibilidades para optimizar el rendimiento del sistema de trigger, por ejemplo, mejorando la eficiencia del trigger en la curva de activación para jets con momentos transversos elevados. Estudios previos han mostrado que esta precisión puede incrementarse significativamente, aunque se necesitan nuevos algoritmos ligeros que puedan funcionar en tiempo real en las FPGAs de la electrónica mejorada para mantener este rendimiento en condiciones de pile-up extremas.

El diseño de las mejoras para los subsistemas de ATLAS ya se encuentra en fase de preproducción o producción. La instalación de los nuevos componentes tendrá lugar durante el Long Shutdown 3 (LS3), programado después de 2025-2026, y su puesta en marcha se realizará durante el LS3 y el Run 4 del LHC, previsto para el inicio de la década de 2030. Estas mejoras son esenciales para garantizar el buen desempeño del experimento en el entorno del HL-LHC.

El Pulse Simulator es una herramienta desarrollada dentro de Athena, el software de reconstrucción offline del experimento ATLAS [140-142]. Su función principal es simular señales físicas bajo diversas condiciones de ruido y parámetros, para luego inyectarlas en la cadena de digitalización de TileCal, generando muestras similares a las que produce el sistema de lectura electrónica. Esto permite estudiar el rendimiento de distintos algoritmos de reconstrucción de señales con control total sobre las características de las señales de entrada, como el ruido electrónico, el pile-up, las imperfecciones en la forma del pulso o variaciones en su fase.

La herramienta se ha actualizado para simular de forma realista las señales en los canales de TileCal bajo las condiciones del HL-LHC. Estas mejoras se basan en medidas realizadas en *test-beams*, datos de laboratorio y simulaciones de la respuesta de TileCal adaptadas a las condiciones del HL-LHC. La nueva versión del simulador, actualmente en uso, permite estudiar con precisión el rendimiento de los algoritmos de reconstrucción de señales en este entorno. Entre las actualizaciones implementadas, el Pulse Simulator ofrece múltiples configuraciones ajustables que facilitan el análisis de una variedad de escenarios experimentales.

La amplitud de las señales generadas por el simulador puede ajustarse para simular señales concretas o para seguir una distribución específica definida por el usuario. Además, se permite configurar la fase de estas señales respecto al tiempo esperado de llegada, ya sea de forma fija o mediante una distribución personalizada. De manera similar, también es posible definir las amplitudes y fases de las señales de pile-up.

El simulador incluye opciones para gestionar cambios de ganancia, simulando la selección entre alta (HG) y baja ganancia (LG) según la saturación de las señales. También permite modelar el ruido electrónico utilizando distribuciones Gaussianas simples o dobles. Los parámetros del ruido pueden configurarse manualmente o extraerse de una base de datos específica para cada canal del detector.

Además, la herramienta permite simular imperfecciones en la forma de los pulsos, lo que es útil para evaluar la robustez de los algoritmos de reconstrucción frente a posibles variaciones. Se puede ajustar el pedestal, de manera uniforme o canal por canal, y configurar la forma del pulso según las características de la electrónica del HL-LHC, incluyendo simulaciones que incorporan colas largas en las señales.

Otro aspecto configurable es el número de muestras digitalizadas, fundamental para analizar la reconstrucción en diferentes condiciones. También se pueden simular el pile-up de eventos cambiando el espaciado entre los *bunches* de protones. Esto permite simular con precisión los efectos de superposición en las señales reconstruidas.

Finalmente, el simulador admite dos paradigmas de lectura. Por un lado, el enfoque tradicional, que utiliza un número fijo de muestras alrededor de la colisión de *bunches* de protones que disparó el trigger. Por otro, el nuevo enfoque basado en lectura continua, en el que todos los cruces de *bunches* se leen, digitalizan y procesan. Esto permite poner a prueba algoritmos de reconstrucción en línea, encargados de determinar la amplitud y el tiempo de las señales en tiempo real para cada cruce de *bunches*. La precisión en la reconstrucción de estas señales es crucial en el entorno del HL-LHC, dado que los algoritmos de trigger dependerán de la calidad de estas medidas para seleccionar eventos interesantes y rechazar el fondo.

Muestras de datos y simulaciones

Los protones son partículas compuestas de quarks y gluones. La interacción de estos componentes se describe mediante la cromodinámica cuántica. En colisiones de alta energía como las del LHC, la libertad asintótica permite aplicar QCD perturbativa para calcular los elementos de matriz de los procesos de dispersión [9]. Esto facilita el cálculo de secciones eficaces, en los cuales las funciones de distribución de partones (PDF) se convolucionan con la sección eficaz partónica $\hat{\sigma}(ij \to X)$. Esto se expresa como $\sigma(pp \to X) = \sum_{ij} \int dx_1 dx_2 f_i(x_1, \mu_F) f_j(x_2, \mu_F) \hat{\sigma}(ij \to X)$,

donde $f_i(x, \mu_F)$ son las PDF de los partones y μ_F y μ_R representan las escalas de factorización y renormalización, respectivamente. Estas PDFs, que se obtienen a partir de ajustes a datos experimentales, dependen también de la escala de energía, pero su evolución está descrita por las ecuaciones de Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [147-149].

La simulación de eventos en el LHC es un proceso complejo y abarca varias etapas [150], comenzando con la dispersión inelástica, que se modela mediante QCD perturbativa, calculando la sección eficaz partónica hasta órdenes elevados en α_s. Dado el alto coste computacional de los cálculos de elementos de matriz en procesos con varios partones salientes, los generadores de elementos de matriz se combinan con algoritmos que simulan cascadas partónicas de una forma fenomenológica. Estos algortimos simulan la emisión de gluones y las divisiones de estos en quarks.

Para eventos con varios jets, el emparejamiento entre el cálculo de los elementos de matriz y la simulación de la cascada partónica es crucial, para evitar tener en cuenta dos veces la misma emisión partónica al calcular la sección eficaz. Métodos como los esquemas MLM [154] y CKKW [155] establecen un umbral que divide el espacio de fase en áreas en los que el cálculo del elemento de matriz se aplica directamente y en las que se simula la cascada partónica. Estos métodos permiten una descripción precisa de eventos con varios jets, evitando contar dos veces la misma emisión partónica y optimizando la precisión en la simulación de eventos con múltiples jets.

A medida que la energía de los partones de la cascada disminuye y se sitúa alrededor de 1 GeV, la QCD entra en un régimen no perturbativo. En esta etapa, los partones se combinan en hadrones mediante modelos fenomenológicos, como el modelo de cuerda de Lund [157] y el de clústeres [158], los cuales describen la formación de estados ligados de hadrones, manteniendo la neutralidad de color.

La simulación de eventos debe considerar también el evento subyacente, que involucra interacciones entre partones adicionales de los protones que no intervienen en la colisión principal. Además, el *pile-up*, la superposición de colisiones adicionales en un mismo cruce de haces, se modela mediante la inclusión de colisiones elásticas junto al evento primario [159]. La simulación incorpora también interacciones de los restos de haz y efectos del halo de haz, junto con colisiones con gas residual en el tubo de haces.

Finalmente, la interacción de partículas en el detector se reproduce mediante el software Geant4, que modela en detalle cómo las partículas interactúan con los distintos materiales del detector [160]. La simulación completa del detector usando Geant4 requiere de un elevado coste computacional, por lo que se han desarrollado métodos, como *AtlFast-II*, que permiten simulaciones rápidas de las interacciones de partículas en los calorímetros y en el detector interno [161, 162]. Los eventos simulados siguen el proceso de reconstrucción que se aplica a los datos reales, permitiendo una comparación directa entre simulaciones y medidas experimentales.

Los datos analizados en esta tesis provienen de colisiones protón-protón a una energía en el centro de masa de $\sqrt{s} = 13$ TeV, recolectados por el detector ATLAS

durante el Run 2 del LHC (2015–2018). La luminosidad integrada total utilizada varía ligeramente entre los dos análisis presentados en esta tesis: 139 fb^{-1} para la búsqueda de desintegraciones del bosón de Higgs violando el sabor leptónico y 140 fb^{-1} para el análisis de $H \to \tau\tau$. Este último valor se corresponde con la calibración de luminosidad más reciente [172].

Ambos análisis dependen de simulaciones de MC para modelar los procesos de señal y fondo. Los procesos de producción del bosón de Higgs simulados abarcan los modos principales: ggF, VBF, VH y $t\bar{t}H$. Además, se consideran contribuciones menores de la producción asociada con un quark top o con dos quarks b. Estas simulaciones se normalizan con cálculos de alta precisión en QCD.

Para el modo de producción ggF, se utilizan muestras generadas con Powheg NNLOPS a NNLO en QCD, escaladas a la sección eficaz calculada en N^3LO, incluyendo correcciones electrodébiles a NLO [25-27, 173-177]. Se emplean el conjunto de PDFs PDF4LHC15nlo [178] y el ajuste AZNLO de Pythia8 para la hadronización.

Los eventos del modo VBF se generan con Powheg a NLO, normalizados a cálculos aproximados a NNLO en QCD [28], incluyendo correcciones electrodébiles a NLO [29, 30]. Para modelar la cascada partónica y la hadronización se utiliza Pythia8.

Los modos VH y $t\bar{t}H$ se simulan con configuraciones específicas. Para VH, se emplea Powheg a NLO con un partón adicional, normalizado a cálculos NNLO en QCD y NLO en correcciones electrodébiles para la producción inducida por quarks [32-38]. En el caso de $t\bar{t}H$, se utilizan PowhegBox v2 a NLO en QCD [40-43] con PDFs NNPDF3.0nlo [181].

En todos los casos, se generan muestras alternativas para estimar incertidumbres asociadas a la modelización de la cascada partónica y la elección del generador, usando Herwig7 [180] y MadGraph5_aMC@NLO [166], respectivamente.

Finalmente, las desintegraciones del bosón de Higgs se normalizan utilizando HDECAY y PROPHECY4F [185-190], asumiendo una masa de 125.09 GeV. Para las desintegraciones con violación de sabor leptónico ($H \to e\tau$ y $H \to \mu\tau$), se considera una fracción de ramificación del 1 %.

La producción de jets asociados a bosones vectoriales (V+jets) mediante interacciones QCD se simula a nivel NLO utilizando Sherpa v2.2.1 con el conjunto de PDFs NNPDF3.0nnlo. Para calcular la matriz de elementos con hasta cuatro partones adicionales a LO, se emplean los generadores Comix y OpenLoops. Las predicciones están normalizadas a secciones eficaces NNLO. La producción electrodébil de V+jets, que genera final states como $\ell\ell jj$, $\ell\nu jj$ y $\nu\nu jj$, utiliza una configuración similar.

La producción de pares $t\bar{t}$ se simula en NLO con PowhegBox v2 utilizando el conjunto NNPDF3.0nlo, y se conecta con Pythia8.230 configurado con el ajuste A14. La incertidumbre de la cascada partónica se evalúa mediante muestras adicionales generadas con Herwig 7.04 y el ajuste H7UE, mientras que las incertidumbres en el generador de elementos de matriz se estiman con

MadGraph5_aMC@NLO conectado con Pythia8. La producción de un solo quarks top en los canales s y t se simula con PowhegBox v2 a nivel NLO en QCD, empleando esquemas de cinco y cuatro sabores, respectivamente, y las secciones eficaces se normalizan utilizando Hathor 2.1.

La producción de pares de bosones (VV) se simula con Sherpa v2.2.1 y v2.2.2. Los cálculos de elementos de matriz se realizan a NLO para procesos con hasta un partón adicional, y a LO para procesos con hasta cuatro partones adicionales. Además, la producción inducida por gluones de pares de bosones WW, WZ y ZZ se simula en LO con un partón adicional. El emparejamiento de los cálculos de matriz de elementos y la simulación de la cascada partónica se realiza mediante el esquema MEPS@NLO, y las predicciones están normalizadas a las secciones eficaces calculadas en NLO.

Reconstrucción de objetos

Las partículas producidas en las colisiones protón-protón en el punto de interacción del detector ATLAS generan señales en sus distintos subdetectores. Estas señales se combinan para reconstruir, calibrar e identificar los diversos objetos físicos utilizados en los análisis, como trazas, vértices, electrones, fotones, muones, jets, leptones τ hadrónicos y el momento transverso faltante.

Las trazas de partículas cargadas se reconstruyen a partir de los *hits* generados en las diferentes capas del detector interno, aprovechando el campo magnético del solenoide, que curva las trayectorias helicoidales de las partículas cargadas. Cada traza se caracteriza mediante cinco parámetros principales: el momento transverso p_{T}, el ángulo polar θ, el ángulo azimutal ϕ, el parámetro de impacto en el plano transverso d_0, y el parámetro de impacto en el plano longitudinal z_0.

La reconstrucción de trazas se lleva a cabo en varias etapas [200]. Inicialmente, los *hits* registrados en el detector de píxeles y el SCT se agrupan en *semillas* de trazas. Estas semillas se extienden posteriormente utilizando un filtro de Kalman combinatorio, que incorpora nuevos *hits* de forma progresiva, generando candidatos de trazas. Este filtro actualiza parámetros como la posición y el momento de las trazas en cada paso. Debido a que pueden existir múltiples candidatos para una misma traza, se aplica un algoritmo de resolución de ambigüedades que selecciona el mejor candidato según criterios como la calidad del ajuste χ^2, el valor de p_{T} y el número de *hits* y *huecos*.

Los vértices primarios de la interacción se reconstruyen utilizando un algoritmo que selecciona el vértice con la mayor suma de momentos transversales cuadrados $\sum p_{\mathrm{T}}^2$, identificándolo como el vértice principal de la colisión inelástica [202, 203].

La reconstrucción de electrones se basa en la identificación de trazas en el detector interno y depósitos de energía en el calorímetro electromagnético [204, 205]. Los electrones producen cascadas electromagnéticas y se identifican mediante proto-clústeres de energía, que se agrupan en clústeres mayores. Si un clúster

no tiene trazas asociadas, un algoritmo de reconocimiento de patrones intenta encontrar trazas considerando las pérdidas de energía debidas a la radiación de frenado.

Tras una calibración inicial de los clústeres, si una traza coincide con un clúster, el objeto se clasifica como electrón; de lo contrario, se considera un fotón. Para reducir falsos positivos, un algoritmo basado en probabilidades evalúa parámetros como la calidad de la coincidencia traza-clúster y la forma de la cascada en el calorímetro.

Los muones producen trazas en el ID, depósitos mínimos en los calorímetros y trazas adicionales en el MS [209]. La reconstrucción en el MS comienza con un reconocimiento de patrones que agrupa los *hits* en segmentos de trayectoria, ajustando luego las trayectorias según el campo magnético.

Existen varios métodos de reconstrucción de muones. El método *Combined* (CB) combina las medidas del ID y del MS para obtener un ajuste óptimo. Métodos alternativos, como *inside-out*, *segment-tagged* y *calorimeter-tagged*, identifican muones según coincidencias de segmentos o depósitos energéticos.

Los jets, generados por la fragmentación de quarks y gluones, se reconstruyen mediante el algoritmo anti-k_t, que es robusto frente a emisiones colineales [210]. La calibración de los jets incluye pasos para ajustar la escala de energía, corregir efectos del *pile-up* y compensar diferencias entre datos y simulaciones [212, 213].

Para identificar jets provenientes de quarks b (*b-jet tagging*), se emplean algoritmos que detectan vértices secundarios y parámetros de impacto. Uno de estos algoritmos, DL1r, utiliza una red neuronal profunda que combina variables discriminantes de distintos métodos para distinguir entre jets b, c y jets de quarks ligeros [215, 216].

Los leptones τ hadrónicos, $\tau_{\text{had-vis}}$, se reconstruyen a partir de jets formados con el algoritmo anti-k_t ($R = 0.4$) y se calibran mediante técnicas de calibración hadrónica locales. Los candidatos $\tau_{\text{had-vis}}$ se clasifican en desintegraciones de uno o tres ramales (*prongs*), según el número de trazas asociadas, y se utiliza una red neuronal recurrente (RNN) para distinguir τ de jets de fondo [217].

El momento transverso perdido, $\vec{E}_{\text{T}}^{\text{miss}}$, se calcula como la suma negativa de los momentos transversales de todos los objetos en el evento, considerando tanto una componente debida a electrones, muones y jets, como una componente derivada de las trazas no asociadas a estos objetos [218].

Medida de secciones eficaces diferenciales de la producción de bosones de Higgs en el canal de desintegración $H \to \tau\tau$

La medida de la sección eficaz de producción del bosón de Higgs en el canal de desintegración $H \to \tau\tau$ permite acceder experimentalmente al acoplamiento de

Yukawa entre el bosón de Higgs y el leptón τ. A día de hoy, es la única medida experimental que permite acceder directamente al acoplamiento entre el bosón de Higgs y los leptones, por lo que es un elemento clave para explorar el sector electrodébil del Modelo Estándar. En el SM, el acoplamiento de Yukawa de los fermiones es proporcional a su masa. Dado que la masa del τ es conocida con precisión [219], la medida directa del acoplamiento de Yukawa en el canal $H \to \tau\tau$ permite contrastar las predicciones del SM con los datos experimentales. Además, la topología de los eventos $H \to \tau\tau$ permite medir con alta sensibilidad la producción del Higgs mediante fusión de bosones vectoriales. La fracción de desintegración de este canal es del 6.3 %, sumada a la topología característica de dos jets de alta energía en la producción VBF permite una discriminación efectiva entre la señal de VBF y los eventos de fondo. Esto permite que, entre los diferentes canales de desintegración que permiten el acceso experimental a la producción del bosón de Higgs, el canal a dos leptones τ es el más preciso para medir la producción VBF.

Desde el descubrimiento del bosón de Higgs en 2012 por las colaboraciones ATLAS y CMS, el estudio detallado de sus propiedades ha sido uno de los principales objetivos en el LHC. Durante el Run 1, ATLAS observó la desintegración $H \to \tau\tau$ con una significancia de 4.5σ [220]. La combinación de los datos de ATLAS y CMS confirmó el proceso con una significancia de 5.5σ [50]. En el Run 2, los análisis se han centrado en obtener medidas de mayor precisión, teniendo como objetivo tanto medidas inclusivas de la sección eficaz total como medidas de secciones eficaces de los diferentes modos de producción del bosón de Higgs en el LHC. Durante los últimos años, a estos enfoques se han sumado estudios más detallados de las propiedades del bosón de Higgs, como la medida de secciones eficaces diferenciales en función de variables cinemáticas clave, como el momento transverso del Higgs (p_{T}^{H}), la multiplicidad de jets y la masa invariante de los dos jets con mayor p_{T} (m_{jj}) [221, 222].

Como se ha indicado anteriormente, el canal $H \to \tau\tau$ es uno de los más relevantes de cara a obtener medidas de precisión de la producción VBF. En el análisis de ATLAS basado en el conjunto completo de datos del Run 2, se obtuvo una significancia observada de 5.3σ para la producción VBF, consistente con las expectativas del SM, y se determinó una sección eficaz de $0.90^{+0.20}_{-0.17}$ veces la predicción teórica [222]. CMS, por su parte, obtuvo resultados similares, con una sección eficaz medida de 0.81 ± 0.17 veces la predicción del SM [223].

Estas medidas, además de medir de forma inclusiva de la producción del bosón de Higgs, proporcionaron medidas de secciones eficaces diferenciales utilizando el marco de las STXS, en el que el espacio de fases se parte en regiones mutuamente excluyentes definidas por variables cinemáticas. Las medidas presentadas en la Ref. [222] incluyen la medida de la sección eficaz diferencial de producción del bosón de Higgs en nueve volúmenes fiduciales. Solo uno de esos volúmenes fiduciales está dedicado a la producción VBF, y se define mediante la selección de eventos con dos jets de alta energía ($p_{\mathrm{T}} > 30$ GeV) y una masa invariante de los dos jets

de al menos 350 GeV. En este volumen fiducial, ATLAS obtuvo una sección eficaz de $0.78^{+0.21}_{-0.18}$ veces la predicción del SM.

El análisis que se presenta en esta tesis, publicado en la Ref. [226], extiende los estudios previos mediante la medida de secciones eficaces diferenciales de la producción de bosones de Higgs en el canal de desintegración $H \to \tau\tau$ en función de p_{T}^{H} y m_{jj}. Se presentan dos medidas diferenciales: una primera utilizando el marco de las STXS, y una segunda medida fiducial diferencial en un espacio de fases enriquecido en eventos VBF. Las secciones eficaces se obtienen en función de p_{T}^{H}, $\Delta\phi_{jj}^{\mathrm{signed}}$, y $p_{\mathrm{T}}(j_0)$. También se realiza una medida bidimensional en función de p_{T}^{H} y $\Delta\phi_{jj}^{\mathrm{signed}}$. Los resultados obtenidos en esta medida fiducial diferencial se interpretan en el marco de la Teoría Efectiva del Modelo Estándar (SMEFT), permitiendo obtener límites en los coeficientes de Wilson de operadores relacionados con el acoplamiento del bosón de Higgs con los bosones vectoriales.

El análisis presenta diversos retos desde el punto de vista experimental. Los leptones τ se desintegran rápidamente, dando lugar a neutrinos en el estado final que no permiten una reconstrucción completa de los eventos. Para mitigar este problema, se utiliza un algoritmo de reconstrucción de masa invariante (Missing Mass Calculator, MMC) que permite obtener una estimación de la masa invariante del par de τ en cada evento.

El análisis está afectado principalmente por dos fuentes de fondo: la producción de $Z \to \tau\tau$ y la producción de eventos con jets mal identificados como leptones τ. En el canal de desintegración dileptónico $\tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$, se añade también la contribución

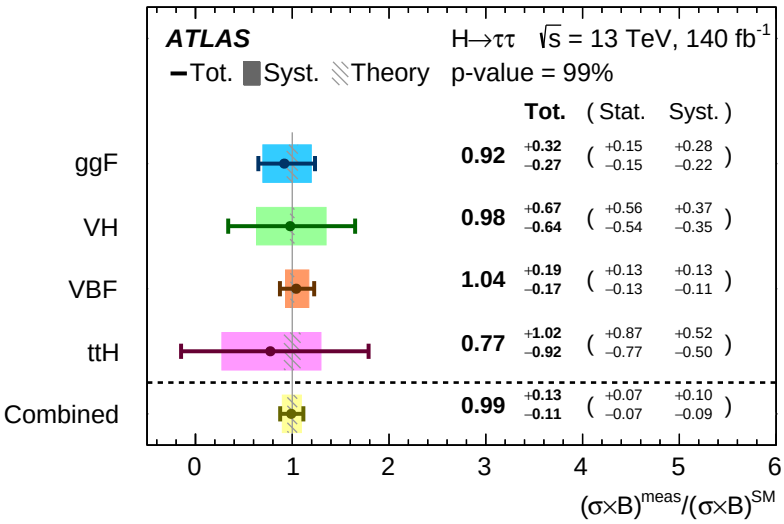


Figura C.7: Medidas de $\sigma_H \times B(H \to \tau\tau)$ relativas a las predicciones del SM en el ajuste que determina la sección eficaz de producción total (indicado como 'Combined' en la gráfica) y en los ajustes estadísticos que determinan las secciones eficaces por modo de producción. La incertidumbre total en cada medida se indica mediante la barra de error, con la contribución debida a las incertidumbres sistemáticas indicada por la banda de color.

de la producción de eventos con quarks top. La normalización de los fondos de $Z \to \tau\tau$ y producción de quarks top se extrae utilizando datos experimentales en regiones de control enriquecidas en esos procesos. Para obtener regiones de control puras en procesos de $Z \to \tau\tau$, se utiliza una técnica denominada *embedding*, en la que se utilizan simulaciones y datos de procesos $Z \to \ell\ell$, siendo ℓ un electrón o un muón, que se corrigen para simular la producción de $Z \to \tau\tau$. Para ello, se tienen en cuenta las diferentes eficiencias experimentales en la identificación, reconstrucción, aislamiento y trigger entre los leptones τ y los leptones ligeros. También se parametriza la energía de los neutrinos en la desintegración de los τ para obtener una estimación de la energía que los productos visibles de la desintegración de los τ ($\tau_{\text{had-vis}}$) deberían tener. La estimación del fondo producido por jets mal identificados como leptones τ se realiza utilizando métodos basados en datos.

El análisis utiliza técnicas de análisis multivariante (MVA) con diferentes objetivos. Por un lado, una red neuronal (NN) se utiliza para reconstruir con precisión p_{T}^{H}, una de las variables que se utiliza para definir las regiones en el marco de las STXS, y que también se utiliza en la medida fiducial diferencial. Por otro lado, se utilizan árboles de decisión (BDT) para discriminar entre señal y fondo en las regiones enriquecidas en señal. Dado que el análisis está diseñado para obtener medidas de secciones eficaces diferenciales para diferentes modos de producción, los BDTs se optimizan para aislar modos de producción específicos, como VBF, VH o $t\bar{t}H$.

Los eventos se clasifican en función del modo de desintegración de los leptones τ: $\tau_e\tau_\mu$, $\tau_{\text{lep}}\tau_{\text{had}}$ y $\tau_{\text{had}}\tau_{\text{had}}$. En el caso de la medida STXS, en cada uno de los canales se definen regiones de señal intentando aislar los diferentes modos de producción, siguiendo los criterios definidos por el marco STXS para aislar las contribuciones

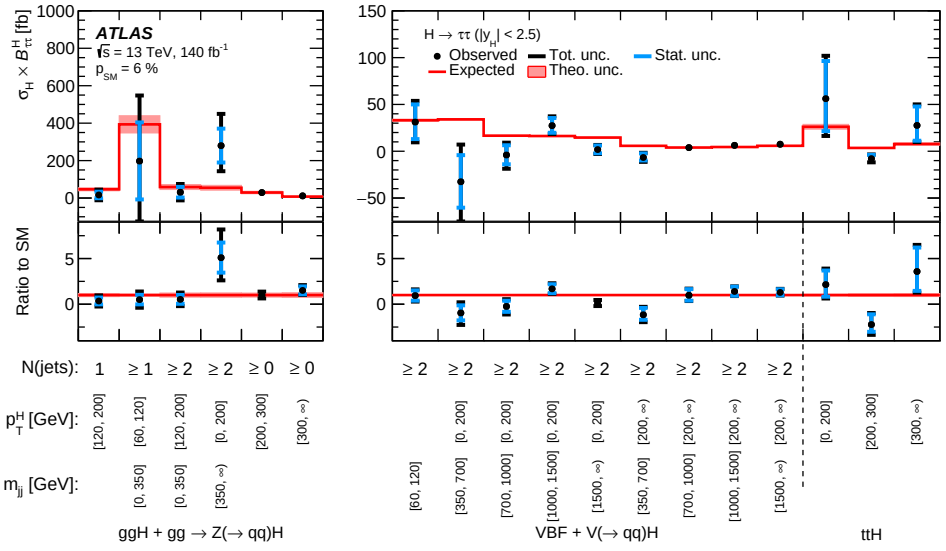


Figura C.8: Medidas de $\sigma_H \times B(H \to \tau\tau)$ en el marco de las STXS. La incertidumbre total en cada medida se indica mediante la barra de error negra, con la contribución debida a la incertiudmbre estadística indicada en azul.

en diferentes rangos de variables cinemáticas. Se definen ocho bines en función de p_{T}^{H} y m_{jj} para la producción VBF, tres bines en función de p_{T}^{H} para $t\bar{t}H$, un bin para VH y seis bines en función de p_{T}^{H}, el número de jets y m_{jj} para ggF. En total, se definen 18 bines dentro del marco STXS.

Las medidas de secciones eficaces para los diferentes modos de producción del bosón de Higgs son compatibles con las predicciones del SM, con un p-valor de 0.99. El valor de la sección eficaz medida con respecto a la predicción del SM en el modo de producción VBF es de $0.93^{+0.17}_{-0.15}$, en buena concordancia con las predicciones del SM. Esta medida de la producción mediante VBF es la más precisa obtenida hasta la fecha en el LHC utilizando únicamente un canal de desintegración. En el caso de la medida en el marco STXS, el p-valor de la comparación entre las medidas y las predicciones del SM es de 0.06, lo que indica una concordancia razonable entre ambas. Para VBF con $p_{\mathrm{T}}^{H} \geq 200$ ($p_{\mathrm{T}}^{H} < 200$) GeV y $m_{\mathrm{jj}} > 1.5$ TeV, se determinó una sección eficaz $1.29^{+0.39}_{-0.34}$ ($0.12^{+0.34}_{-0.33}$) veces la predicción del SM, siendo esta la primera medida en este rango alto de p_{T}^{H} y la más precisa en el rango bajo de p_{T}^{H}.

En la medida fiducial diferencial, se define un volumen fiducial enriquecido en eventos VBF, tratando de reducir la contaminación de otros modos de producción. Las medidas diferenciales se obtienen en función de cuatro variables: p_{T}^{H}, $\Delta\phi_{jj}^{\mathrm{signed}}$, $p_{\mathrm{T}}(j_0)$ y $\Delta\phi_{jj}^{\mathrm{signed}}$ vs. p_{T}^{H}. La técnica de *unfolding* se utiliza para corregir los efectos de resolución y aceptancia del detector en las medidas diferenciales, y proporcionar medidas de secciones eficaces independientes de los efectos experimentales y del modelo.

Estas medidas, que se comparan con diferentes predicciones teóricas que emplean diferentes generadores de MC, también son sensibles a física más allá del SM. Por eso se interpretan en el marco de la SMEFT, permitiendo obtener límites en los coeficientes de Wilson de operadores relacionados con el acoplamiento del bosón de Higgs con los bosones vectoriales, y sus propiedades bajo transformaciones de CP. Estos coeficientes, asociados a operadores en la base de Warsaw, son c_{HW}, c_{HB}, c_{HWB}, $c_{H\tilde{W}}$, $c_{H\tilde{B}}$ y $c_{H\tilde{W}B}$. Los límites al 95 % CL obtenidos para el operador $c_{H\tilde{W}}$ son $[-0.31, +0.88]$, siendo los más restrictivos obtenidos hasta la fecha en cualquier canal de desintegración del bosón de Higgs.

Búsqueda de desintegraciones del bosón de Higgs violando el sabor leptónico a $e\tau$ y $\mu\tau$

La medida precisa de los acoplamientos del bosón de Higgs con fermiones y bosones gauge es una prueba fundamental para evaluar la validez y consistencia del SM. El estudio del sector de Higgs abarca tanto la caracterización de sus propiedades fundamentales como la búsqueda de señales de nueva física más allá del SM. El análisis presentado en esta tesis aborda las búsquedas de desintegraciones del bosón de Higgs que violan el sabor leptónico. La búsqueda se centra en las desintegraciones

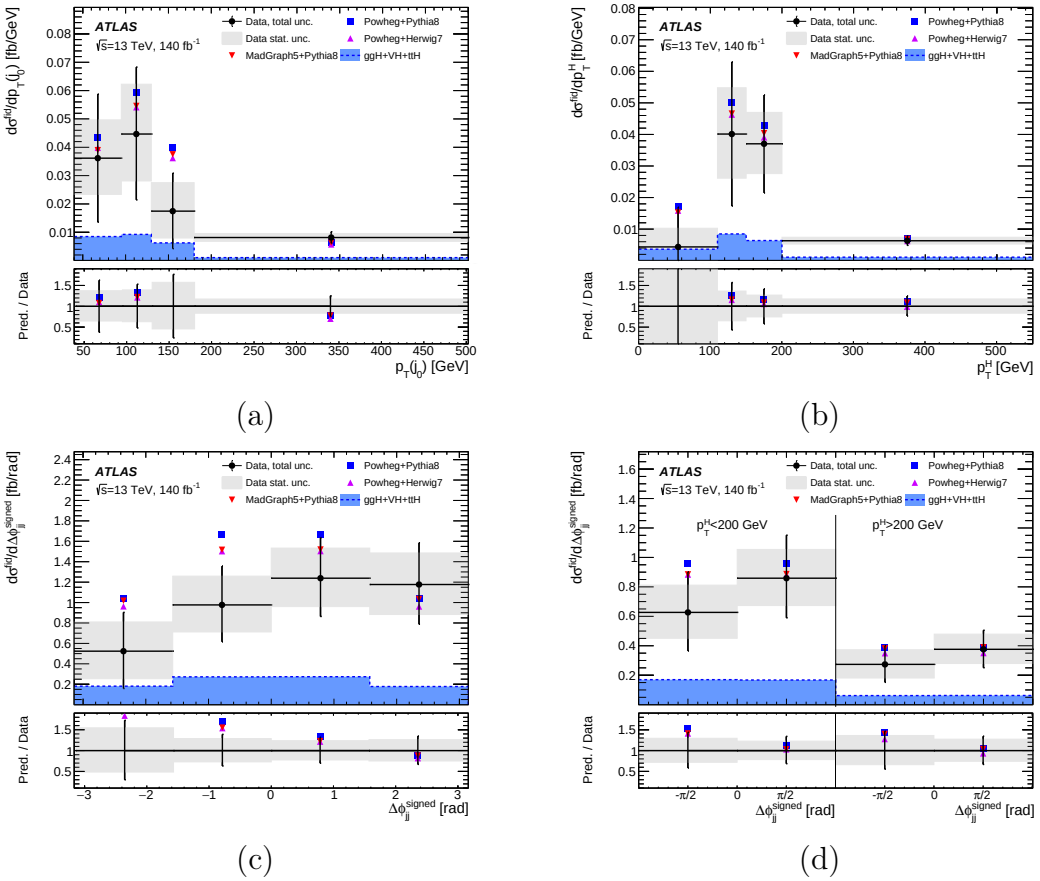
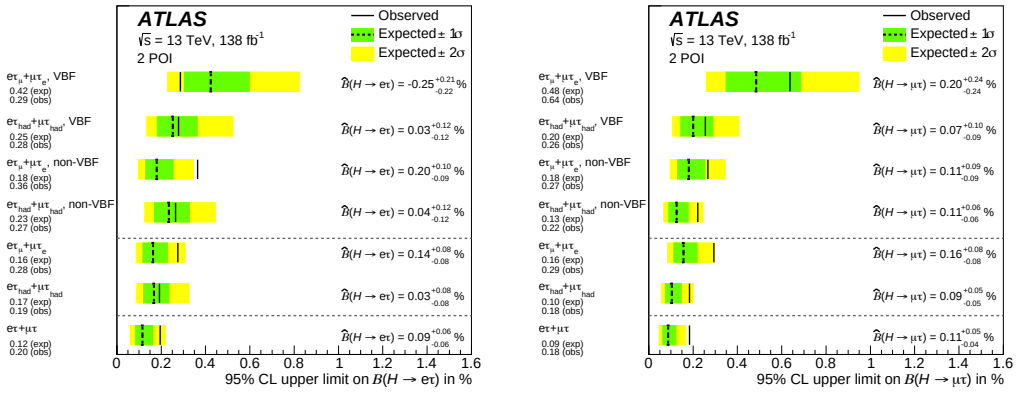


Figura C.9: Medidas de secciones eficaces fiduciales diferenciales para (a) $p_{\mathrm{T}}^{j_0}$, (b) p_{T}^{H}, (c) $\Delta\phi_{jj}^{\mathrm{signed}}$ y (d) $\Delta\phi_{jj}^{\mathrm{signed}}$ vs p_{T}^{H}. Las medidas se comparan con predicciones del SM de diferentes generadores. La caja sombreada alrededor de los puntos de datos representa la incertidumbre estadística, mientras que las barras de error indican la incertidumbre total. También se muestran las contribuciones de ggF, VH y $t\bar{t}H$ predichas por POWHEG + PYTHIA8. Los paneles inferiores muestran el cociente entre las diferentes predicciones y los datos.

del bosón de Higgs a $e\tau$ y $\mu\tau$, utilizando 139 fb^{-1} de colisiones protón-protón a una energía de $\sqrt{s} = 13$ TeV adquiridas por el experimento ATLAS.

Las oscilaciones de neutrinos, observadas experimentalmente, han demostrado que la conservación estricta del sabor leptónico no se cumple en la naturaleza. Esto abre la posibilidad de que se produzcan procesos LFV en el sector de los leptones cargados. Las búsquedas directas son fundamentales para establecer límites experimentales en estas desintegraciones del bosón de Higgs. Si bien experimentos de baja energía han impuesto límites estrictos a procesos como $\mu \to e\gamma$, $\tau \to \mu\gamma$ y $\tau \to e\gamma$, los límites indirectos en las desintegraciones $H \to e\tau$ y $H \to \mu\tau$ son mucho menos estrictos que los que se pueden obtener mediante búsquedas directas en los experimentos del LHC.

Las búsquedas directas de $H \to e\tau$ y $H \to \mu\tau$ son esenciales para investigar posibles valores no nulos en los elementos fuera de la diagonal de la matriz de Yukawa. Análisis previos realizados por ATLAS han establecido límites a un intervalo de



(a) (b)

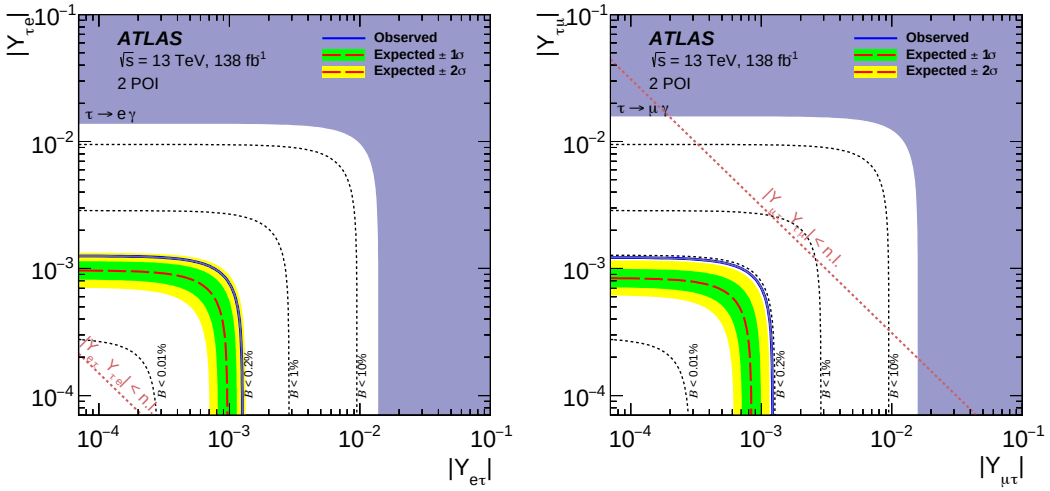
Figura C.10: Límites a un intervalo de confianza del 95 % para las fracciones de desintegración de los procesos (a) $H \to e\tau$ y (b) $H \to \mu\tau$. Los límites observados y esperados se muestran con líneas sólidas y discontinuas, respectivamente. El valor estimado de la fracción de desintegración se indica porcentualmente. Los resultados se muestran utilizando ajustes estadísticos en los diferentes canales y categorías. Los resultados utilizando la combinación de todos los canales y categorías se muestran en la última fila.

confianza del 95 % en las fracciones de desintegración 0.47 % y 0.28 % para los procesos $H \to e\tau$ y $H \to \mu\tau$, respectivamente. Estos resultados se obtuvieron utilizando 36.1 fb^{-1} de datos de colisiones protón-protón a una energía de $\sqrt{s} = 13$ TeV [84]. Por otro lado, la colaboración CMS, utilizando 137 fb^{-1} de datos a la misma energía, obtuvo límites más estrictos de 0.22 % y 0.15 % para los procesos $H \to e\tau$ y $H \to \mu\tau$, respectivamente [85]. ATLAS también ha reportado un límite de 6.1×10^{-5} al 95 % CL para el proceso $H \to e\mu$ utilizando 139 fb^{-1} de datos a $\sqrt{s} = 13$ TeV [82].

Uno de los principales objetivos de esta tesis es estudiar las desintegraciones $H \to e\tau$ y $H \to \mu\tau$ utilizando el conjunto completo de datos del Run 2 del LHC. Estos resultados se encuentran publicados en la Ref. [238].

Para cada canal, se analizan dos tipos de estados finales: $\ell\tau_{\ell'}$ y $\ell\tau_{\text{had}}$. En el caso de $\ell\tau_{\ell'}$, los leptones τ se identifican mediante sus decaimientos leptónicos, $\tau \to \ell'\nu\bar{\nu}$, mientras que en $\ell\tau_{\text{had}}$ se investigan los decaimientos hadrónicos, $\tau \to$ hadrón $+ \nu$.

En el estado final $\ell\tau_{\ell'}$, solo se utilizán eventos en los que los dos leptones del estado final tienen sabor diferente para evitar el fondo de Drell-Yan, que opacaría la búsqueda de señales de LFV. Se emplean dos métodos para estimar el fondo en este canal. El primero, basado en simulaciones MC, utiliza plantillas generadas por simulación para modelar los fondos principales, mientras que las contribuciones de leptones mal identificados se estiman a partir de técnicas basadas en datos. La normalización de los fondos principales ($Z \to \tau\tau$ y fondos relacionados con la producción de quarks top) se extrae de regiones de control enriquecidas en esos procesos. El segundo método explota la simetría del SM bajo el intercambio de electrones y muones para estimar el fondo directamente a partir de datos [237]. Para



(a) (b)

Figura C.11: Límites esperados (línea discontinua roja) y observados (línea sólida azul) a un intervalo de confianza del 95 % en los valores absolutos de los acoplamientos $Y_{\tau e}$ y $Y_{e\tau}$ (a) y $Y_{\tau\mu}$ y $Y_{\mu\tau}$ (b). Los límites indirectos más restrictivos de búsquedas de $\tau \to \ell\gamma$ se muestran en la región morada oscura. Las líneas discontinuas indican límites correspondientes a diferentes fracciones de desintegración (0.01 %, 0.2 %, 1 % y 10 %), mientras que la línea punteada marca el límite de naturalidad.

el canal $\ell\tau_{\text{had}}$, se aplica exclusivamente el método basado en simulaciones, utilizando plantillas generadas por MC para modelar los fondos principales y técnicas basadas en datos para estimar las contribuciones de jets mal identificados como leptones τ.

El análisis se basa en una selección no estricta de eventos, que son posteriormente clasificados en dos categorías: uno enfocada en la producción VBF y otra en el resto de modos de producción del bosón de Higgs (llamada non-VBF, dominada por producción ggF). Esto perimite mejorar la separación entre señal y fondo aprovechando las características de los jets en los eventos producidos mediante VBF. La discriminación entre señal y fondo se mejora mediante técnicas de análisis multivariante, como redes neuronales (en el caso del análisis basado en simetría) y árboles de decisión (en el análisis basado en simulaciones MC).

El resultado final se extrae mediante un ajuste estadístico que combina los resultados de los canales $\ell\tau_{\ell'}$ y $\ell\tau_{\text{had}}$. El ajuste determina simultáneamente las fracciones de desintegración de $H \to e\tau$ y $H \to \mu\tau$. La figura C.10 muestra los límites al 95 % CL obtenidos para las fracciones de desintegración de los procesos $H \to e\tau$ y $H \to \mu\tau$, mostrando los límites observados y esperados en los diferentes canales y categorías. Se observan excesos de 2.4σ y 1.6σ para las señales $H \to \mu\tau$ y $H \to e\tau$, respectivamente. En el caso de $H \to e\tau$, el exceso se debe principalmente a la categoría non-VBF del estado final $\ell\tau_{\ell'}$, utilizando resultados del método basado en MC. Por otro lado, el exceso observado en $H \to \mu\tau$ proviene principalmente de la categoría non-VBF del estado final $\ell\tau_{\text{had}}$.

Los resultados del ajuste simultáneo de las señales $H \to e\tau$ y $H \to \mu\tau$, incluyendo los contornos al 68 % y 95 % CL, muestran compatibilidad con la

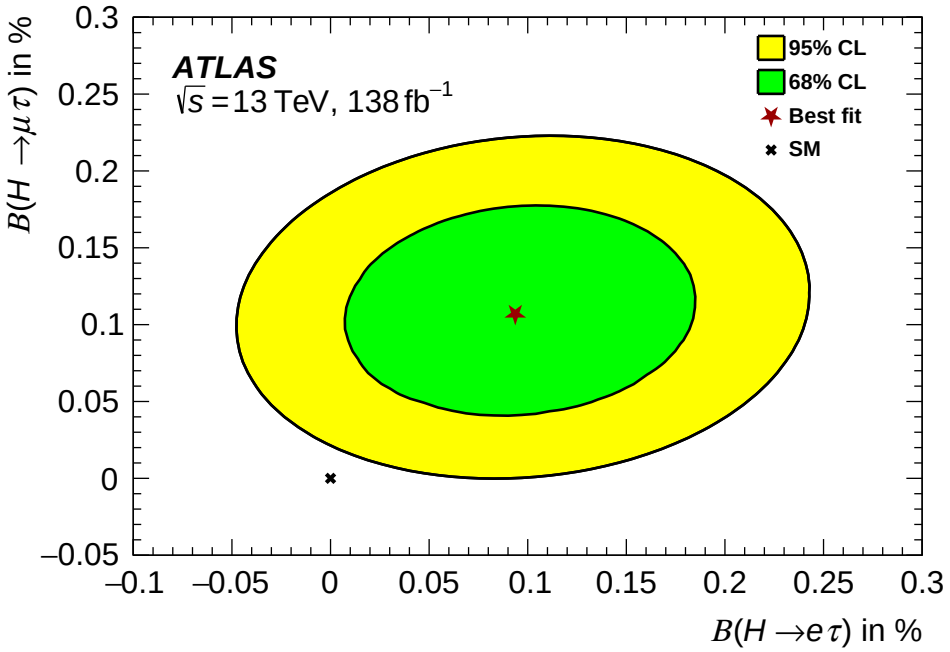


Figura C.12: Valores obtenidos en el ajuste estadístico para las fracciones de desintegración $\mathcal{B}(H \to e\tau)$ y $\mathcal{B}(H \to \mu\tau)$, expresadas en porcentaje. Se muestra el valor de la predicción del SM (cruz negra) y el valor obtenido en el ajuste a los datos experimentales (estrella roja). Se indican las regiones de 68 % y 95 % CL.

predicción del Modelo Estándar dentro de 2.1σ. Esto se ilustra en la Figura C.12, donde los contornos en el plano $\mu_{e\tau}$-$\mu_{\mu\tau}$ revelan que el exceso está impulsado principalmente por la señal $\mu\tau$. Además, los límites superiores al 95 % CL en los elementos de la matriz de Yukawa $Y_{\tau\ell}$ y $Y_{\ell\tau}$ son $\sqrt{|Y_{\tau e}|^2 + |Y_{e\tau}|^2} < 0.0013$ y $\sqrt{|Y_{\tau\mu}|^2 + |Y_{\mu\tau}|^2} < 0.0012$, como se muestra en la Figura C.11.

Conclusiones

Esta tesis aborda tres áreas principales de investigación relacionadas con el experimento ATLAS del LHC. La primera está relacionada con el desarrollo de un simulador de señales para la electrónica de lectura de TileCal como parte de las mejoras del detector de cara al HL-LHC. Las otras dos se centran en la exploración experimental del sector de Higgs: la medida de secciones eficaces diferenciales de producción del bosón de Higgs en el canal de desintegración $H \to \tau^+\tau^-$ y la búsqueda de desintegraciones del bosón de Higgs con violación del sabor leptónico ($H \to e\tau$ y $H \to \mu\tau$), utilizando el conjunto completo de datos del Run 2, correspondiente a 140 fb^{-1} de colisiones pp a $\sqrt{s} = 13$ TeV.

El simulador de pulsos de TileCal permite modelar la respuesta de los canales de TileCal en las condiciones del HL-LHC, en el que la luminosidad instantánea llegará a 7.5×10^{34} cm^{-2}s^{-1}. La actualización de este paquete incluye características

como las formas de pulso generadas por la electrónica de lectura mejorada, simulación de los efectos del pile-up y condiciones de ruido realistas. Una de sus implementaciones clave es el modo de lectura continua, que permite simular cadenas ininterrumpidas de cuentas de ADC, emulando las entradas que recibirán los algoritmos de reconstrucción en línea implementados en la electrónica fuera del detector. Este simulador es crucial para validar y optimizar los algoritmos de reconstrucción de señales en preparación para el HL-LHC.

Las medidas diferenciales de la sección eficaz de producción del bosón de Higgs en el canal $H \to \tau^+\tau^-$ se realizaron con los 140 fb^{-1} del conjunto de datos del Run 2. Estas medidas no solo incluyen la sección eficaz total de producción y la sección eficaz por modo de producción, sino también resultados diferenciales según variables cinemáticas dentro del marco de las STXS. Por ejemplo, en la producción por VBF, el cociente entre la sección eficaz medida y la predicha por el SM fue $0.93^{+0.17}_{-0.15}$.

Se midieron 18 regiones cinemáticas STXS, obteniendo una compatibilidad razonable con el SM (p-valor del 6 %). En particular, se lograron las primeras medidas de VBF en el rango $p_{\mathrm{T}}^{H} \geq 200$ GeV, mientras que para eventos con $p_{\mathrm{T}}^{H} < 200$ GeV y $m_{\mathrm{jj}} > 1.5$ TeV se alcanzaron precisiones sin precedentes. Además, se realizaron medidas desdobladas en distribuciones específicas para comparar de manera independiente con predicciones teóricas y para interpretaciones en EFT.

La medida en el marco de las STXS se complementó con una medida fiducial diferencial en un espacio de fases enriquecido en eventos VBF. Esta medida fue reinterpretada en SMEFT, permitiendo obtener límites en los coeficientes de Wilson de operadores relacionados con el acoplamiento del bosón de Higgs con los bosones gauge. Para el coeficiente $c_{H\tilde{W}}$, se obtuvieron límites al 95 % CL de $[-0.31, +0.88]$, los más restrictivos obtenidos hasta la fecha en cualquier canal de desintegración del bosón de Higgs.

La extrapolación de estas medidas para el HL-LHC, asumiendo 3000 fb^{-1} a $\sqrt{s} = 14$ TeV, indica mejoras significativas en la precisión de las secciones eficaces, donde la precisión de la medida del modo de producción VBF podría alcanzar el 7 %. Estas mejoras se producirán no solo por el aumento en la energía y la luminosidad, sino también por mejoras en el detector y en los avances en la reconstrucción de objetos.

Finalmente, esta tesis presenta búsquedas de desintegraciones del bosón de Higgs con violación del sabor leptónico en los canales $H \to e\tau$ y $H \to \mu\tau$. Los resultados obtenidos mejoran los límites previos más allá de las previsiones debidas al aumento de la luminosidad integrada. Esto se debe a diversas mejoras en el análisis, que incluyen una utilización de algoritmos multivariante más avanzados, una mejora en el análisis estadístico que permite obtener simultáneamente las fracciones de desintegración de los dos procesos LFV o mejoras en la asignación de leptones (la determinación de qué lepton proviene directamente del Higgs y cuál del leptón τ). Se observaron excesos no significativos en ambos canales de desintegración, con límites superiores al 95 % de confianza en las fracciones de

desintegración de 0.20 % para $H \to e\tau$ y 0.18 % para $H \to \mu\tau$. Estos límites se trasladaron a los coeficientes no diagonales de las matrices de Yukawa.

Estos resultados son de particular relevancia dado que imponen límites más estrictos en las desintegraciones LFV del Higgs que los obtenidos mediante restricciones indirectas basadas en los resultados de búsquedas LFV a baja energía, en procesos como $\tau \to \mu\gamma$ y $\tau \to e\gamma$.

Las extrapolaciones para el HL-LHC indican que, con 3000 fb^{-1} a $\sqrt{s} = 14$ TeV, se podrían alcanzar límites al 95 % CL de 0.024 % para ambas desintegraciones, mejorando en un factor de 4.8 y 3.9 respectivamente los límites obtenidos en el análisis de los datos del Run 2.

Los estudios realizados en esta tesis amplían significativamente el conocimiento sobre el bosón de Higgs y sus interacciones, reforzando la capacidad del experimento ATLAS para investigar sus propiedades en el canal de desintegración $H \to \tau^+\tau^-$ y abriendo nuevas perspectivas en la búsqueda de desintegraciones que violen el sabor leptónico, un posible indicio de física más allá del Modelo Estándar. Con los datos del Run 3 y los futuros registros del HL-LHC, se espera alcanzar una precisión sin precedentes en la medida de las secciones eficaces de producción del bosón de Higgs y establecer límites aún más estrictos para procesos que desafían las predicciones actuales, acercándonos un paso más a comprender el rol fundamental del bosón de Higgs tanto dentro como más allá del Modelo Estándar.

References

[1] P. Langacker
The Standard Model and Beyond
CRC Press, 2010 (cit. on pp. 6, 8, 305).

[2] P. W. Higgs
Broken Symmetries and the Masses of Gauge Bosons
Phys. Rev. Lett. **13** (1964) 508, ed. by J. C. Taylor (cit. on pp. 6, 305).

[3] F. Englert and R. Brout
Broken Symmetry and the Mass of Gauge Vector Mesons
Phys. Rev. Lett. **13** (1964) 321, ed. by J. C. Taylor (cit. on pp. 6, 305).

[4] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble
Global Conservation Laws and Massless Particles
Phys. Rev. Lett. **13** (1964) 585, ed. by J. C. Taylor (cit. on pp. 6, 305).

[5] H. Georgi
Lie algebras in particle physics
2nd ed., vol. 54, Perseus Books, 1999 (cit. on pp. 7, 305).

[6] Abel, C. et al.
Measurement of the Permanent Electric Dipole Moment of the Neutron
Phys. Rev. Lett. **124** (2020) 081803, arXiv: 2001.11966 [hep-ex]
(cit. on p. 8).

[7] F. Wilczek
Problem of Strong P and T Invariance in the Presence of Instantons
Phys. Rev. Lett. **40** (1978) 279 (cit. on p. 8).

[8] R. D. Peccei and H. R. Quinn
CP Conservation in the Presence of Instantons
Phys. Rev. Lett. **38** (1977) 1440 (cit. on p. 8).

[9] D. J. Gross and F. Wilczek
Ultraviolet Behavior of Nonabelian Gauge Theories
Phys. Rev. Lett. **30** (1973) 1343, ed. by J. C. Taylor
(cit. on pp. 8, 83, 313).

[10] Sheldon L. Glashow
Partial-symmetries of weak interactions
Nuclear Physics **22** (1961) 579, URL: https://www.sciencedirect.com/science/article/pii/0029558261904692 (cit. on pp. 8, 305).

[11] S. Weinberg
A Model of Leptons
Phys. Rev. Lett. **19** (1967) 1264 (cit. on pp. 8, 11, 305).

[12] A. Salam and J.C. Ward
Electromagnetic and weak interactions
Physics Letters **13** (1964) 168, URL: https://www.sciencedirect.com/science/article/pii/0031916364907115 (cit. on pp. 8, 305).

[13] E. Noether
Invariant Variation Problems
Gott. Nachr. **1918** (1918) 235, arXiv: physics/0503066 (cit. on p. 9).

[14] M. Gell-Mann
The interpretation of the new particles as displaced charge multiplets
Nuovo Cim. **4** (1956) 848 (cit. on p. 9).

[15] K. Nishijima
*Charge Independence Theory of V Particles**
Progress of Theoretical Physics **13** (1955) 285,
eprint: https://academic.oup.com/ptp/article-pdf/13/3/285/5425869/13-3-285.pdf,
URL: https://doi.org/10.1143/PTP.13.285 (cit. on p. 9).

[16] Carlo Giunti and Chung W. Kim
Fundamentals of Neutrino Physics and Astrophysics
Oxford University Press, 2007,
URL: https://doi.org/10.1093/acprof:oso/9780198508717.001.0001 (cit. on p. 9).

[17] A. Pich
The Standard Model of Electroweak Interactions
2010 European School of High Energy Physics, 2012 1,
arXiv: 1201.0537 [hep-ph] (cit. on p. 13).

[18] Barate, R. et al.
Search for the standard model Higgs boson at LEP
Phys. Lett. B **565** (2003) 61, arXiv: hep-ex/0306033 (cit. on p. 15).

[19] CDF and D0 Collaborations
Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data
2012, arXiv: 1207.0449 [hep-ex] (cit. on p. 15).

[20] David Smith
Higgs Boson searches at LEP2.
PoS **silafae-III** (2000) 045, URL: https://cds.cern.ch/record/851161 (cit. on p. 15).

[21] 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. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on pp. 15, 16).

[22] CMS Collaboration
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC
Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on p. 15).

[23] ATLAS and CMS Collaborations
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] (cit. on p. 15).

[24] de Florian, D. et al.
Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector
2/2017 (2016), arXiv: 1610.07922 [hep-ph] (cit. on pp. 15, 20, 21, 154, 193, 306).

[25] C. Anastasiou et al.
Higgs Boson Gluon-Fusion Production in QCD at Three Loops
Phys. Rev. Lett. **114** (2015) 212001, arXiv: 1503.06056 [hep-ph] (cit. on pp. 15, 91, 315).

[26] C. Anastasiou et al.
High precision determination of the gluon fusion Higgs boson cross-section at the LHC
JHEP **05** (2016) 058, arXiv: 1602.00695 [hep-ph] (cit. on pp. 15, 91, 315).

[27] S. Actis et al.
NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders
Phys. Lett. B **670** (2008) 12, arXiv: 0809.1301 [hep-ph] (cit. on pp. 15, 91, 315).

[28] P. Bolzoni et al.
Higgs production via vector-boson fusion at NNLO in QCD
Phys. Rev. Lett. **105** (2010) 011801, arXiv: 1003.4451 [hep-ph]
(cit. on pp. 16, 92, 315).

[29] M. Ciccolini, A. Denner, and S. Dittmaier
Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC
Phys. Rev. D **77** (2008) 013002, arXiv: 0710.4749 [hep-ph]
(cit. on pp. 16, 92, 315).

[30] M. Ciccolini, A. Denner, and S. Dittmaier
Strong and electroweak corrections to the production of Higgs + 2jets via weak interactions at the LHC
Phys. Rev. Lett. **99** (2007) 161803, arXiv: 0707.0381 [hep-ph]
(cit. on pp. 16, 92, 315).

[31] ATLAS Collaboration
A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery
Nature **607** (2022) 52, [Erratum: Nature 612, E24 (2022)],
arXiv: 2207.00092 [hep-ex] (cit. on pp. 17, 19, 22, 25, 26, 28, 306, 307).

[32] O. Brein, R. V. Harlander, and T. J. E. Zirke
vh@nnlo - Higgs Strahlung at hadron colliders
Comput. Phys. Commun. **184** (2013) 998, arXiv: 1210.5347 [hep-ph]
(cit. on pp. 17, 92, 315).

[33] R. V. Harlander et al.
vh@nnlo-v2: New physics in Higgs Strahlung
JHEP **05** (2018) 089, arXiv: 1802.04817 [hep-ph]
(cit. on pp. 17, 92, 315).

[34] O. Brein et al.
Top-Quark Mediated Effects in Hadronic Higgs-Strahlung
Eur. Phys. J. C **72** (2012) 1868, arXiv: 1111.0761 [hep-ph]
(cit. on pp. 17, 92, 315).

[35] L. Altenkamp et al.
Gluon-induced Higgs-strahlung at next-to-leading order QCD
JHEP **02** (2013) 078, arXiv: 1211.5015 [hep-ph] (cit. on pp. 17, 92, 315).

[36] R. V. Harlander et al.
Soft gluon resummation for gluon-induced Higgs Strahlung
JHEP **11** (2014) 082, arXiv: 1410.0217 [hep-ph] (cit. on pp. 17, 92, 315).

[37] A. Denner et al.
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, arXiv: 1412.5390 [hep-ph]
(cit. on pp. 17, 92, 315).

[38] M. L. Ciccolini, S. Dittmaier, and M. Kramer
Electroweak radiative corrections to associated WH and ZH production at hadron colliders
Phys. Rev. D **68** (2003) 073003, arXiv: hep-ph/0306234
(cit. on pp. 17, 92, 315).

[39] A. Denner et al.
Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with HAWK
JHEP **03** (2012) 075, arXiv: 1112.5142 [hep-ph] (cit. on p. 17).

[40] W. Beenakker et al.
NLO QCD corrections to t anti-t H production in hadron collisions
Nucl. Phys. B **653** (2003) 151, arXiv: hep-ph/0211352
(cit. on pp. 18, 92, 315).

[41] S. Dawson et al.
Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections
Phys. Rev. D **68** (2003) 034022, arXiv: hep-ph/0305087
(cit. on pp. 18, 92, 315).

[42] Y. Zhang et al.
QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider
Phys. Lett. B **738** (2014) 1, arXiv: 1407.1110 [hep-ph]
(cit. on pp. 18, 92, 315).

[43] S. Frixione et al.
Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons
JHEP **06** (2015) 184, arXiv: 1504.03446 [hep-ph]
(cit. on pp. 18, 92, 315).

[44] S. Dawson et al.
Exclusive Higgs boson production with bottom quarks at hadron colliders
Phys. Rev. D **69** (2004) 074027, arXiv: hep-ph/0311067 (cit. on p. 18).

[45] S Dittmaier, M. Krämer, and M. Spira
Higgs radiation off bottom quarks at the Tevatron and the CERN LHC
Phys. Rev. D **70** (2004) 074010, arXiv: hep-ph/0309204 (cit. on p. 18).

[46] R. Harlander, M. Kramer, and M. Schumacher
Bottom-quark associated Higgs-boson production: reconciling the four- and five-flavour scheme approach
(2011), arXiv: 1112.3478 [hep-ph] (cit. on p. 18).

[47] F. Demartin et al.
Higgs production in association with a single top quark at the LHC
Eur. Phys. J. C **75** (2015) 267, arXiv: 1504.00611 [hep-ph]
(cit. on p. 18).

[48] F. Demartin et al.
tWH associated production at the LHC
Eur. Phys. J. C **77** (2017) 34, arXiv: 1607.05862 [hep-ph] (cit. on p. 18).

[49] ATLAS 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]
(cit. on pp. 20, 306).

[50] ATLAS 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]
(cit. on pp. 20, 114, 306, 318).

[51] ATLAS 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] (cit. on pp. 20, 306).

[52] ATLAS Collaboration
Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector
Eur. Phys. J. C **75** (2015) 335, arXiv: 1503.01060 [hep-ex]
(cit. on p. 20).

[53] CMS Collaboration
Search for Higgs boson off-shell production in proton-proton collisions at 7 and 8 TeV and derivation of constraints on its total decay width
JHEP **09** (2016) 051, arXiv: 1605.02329 [hep-ex] (cit. on p. 20).

[54] ATLAS Collaboration
Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector
Eur. Phys. J. C **75** (2015) 476, [Erratum: Eur.Phys.J.C 76, 152 (2016)],
arXiv: 1506.05669 [hep-ex] (cit. on p. 20).

[55] CMS Collaboration
Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV
Phys. Rev. D **92** (2015) 012004, arXiv: 1411.3441 [hep-ex]
(cit. on p. 20).

[56] ATLAS Collaboration
Combined Measurement of the Higgs Boson Mass from the $H\to\gamma\gamma$ and $H\to ZZ^*\to 4\ell$ Decay Channels with the ATLAS Detector Using s=7, 8, and 13 TeV pp Collision Data
Phys. Rev. Lett. **131** (2023) 251802, arXiv: 2308.04775 [hep-ex]
(cit. on pp. 21, 306).

[57] F. Caola and K. Melnikov
Constraining the Higgs boson width with ZZ production at the LHC
Phys. Rev. D **88** (2013) 054024, arXiv: 1307.4935 [hep-ph]
(cit. on p. 21).

[58] ATLAS Collaboration
Evidence of off-shell Higgs boson production from ZZ leptonic decay channels and constraints on its total width with the ATLAS detector
Phys. Lett. B **846** (2023) 138223, arXiv: 2304.01532 [hep-ex]
(cit. on pp. 21, 306).

[59] CMS Collaboration
Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production
Nature Phys. **18** (2022) 1329, arXiv: 2202.06923 [hep-ex]
(cit. on pp. 21, 306).

[60] CMS Collaboration
A portrait of the Higgs boson by the CMS experiment ten years after the discovery.
Nature **607** (2022) 60, [Erratum: Nature 623, (2023)],
arXiv: 2207.00043 [hep-ex] (cit. on p. 22).

[61] ATLAS and CMS Collaborations
Evidence for the Higgs Boson Decay to a Z Boson and a Photon at the LHC
Phys. Rev. Lett. **132** (2024) 021803, arXiv: 2309.03501 [hep-ex]
(cit. on p. 22).

[62] CMS Collaboration
Evidence for Higgs boson decay to a pair of muons
JHEP **01** (2021) 148, arXiv: 2009.04363 [hep-ex] (cit. on p. 22).

[63] Andersen, J. R. et al.
Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report
9th Les Houches Workshop on Physics at TeV Colliders, 2016,
arXiv: 1605.04692 [hep-ph] (cit. on pp. 23, 131, 307).

[64] Berger, N. et al.
Simplified Template Cross Sections - Stage 1.1
(2019), arXiv: 1906.02754 [hep-ph] (cit. on pp. 23, 24).

[65] Andersen, J. R. et al.
Handbook of LHC Higgs Cross Sections: 3. Higgs Properties
(2013), ed. by S Heinemeyer et al., arXiv: 1307.1347 [hep-ph]
(cit. on pp. 25, 154).

[66] ATLAS Collaboration
Combination of searches for Higgs boson pair production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
(2024), arXiv: 2406.09971 [hep-ex] (cit. on p. 28).

[67] Super-Kamiokande Collaboration
Evidence for oscillation of atmospheric neutrinos
Phys. Rev. Lett. **81** (1998) 1562, arXiv: hep-ex/9807003
(cit. on pp. 28, 307).

[68] SNO Collaboration
Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory
Phys. Rev. Lett. **89** (2002) 011301, arXiv: nucl-ex/0204008
(cit. on pp. 28, 307).

[69] SNO Collaboration
Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory
Phys. Rev. Lett. **87** (2001) 071301, arXiv: nucl-ex/0106015
(cit. on pp. 28, 307).

[70] M. Arana-Catania, E. Arganda, and M. J. Herrero
Non-decoupling SUSY in LFV Higgs decays: a window to new physics at the LHC
JHEP **09** (2013) 160, [Erratum: JHEP 10, 192 (2015)],
arXiv: 1304.3371 [hep-ph] (cit. on pp. 29, 308).

[71] A. Arhrib, Y. Cheng, and O. C. W. Kong
Comprehensive analysis on lepton flavor violating Higgs boson to $\mu^\mp\tau^\pm$ decay in supersymmetry without R parity
Phys. Rev. D **87** (2013) 015025, arXiv: 1210.8241 [hep-ph]
(cit. on pp. 29, 308).

[72] A. Vicente
Higgs lepton flavor violating decays in Two Higgs Doublet Models
Front. in Phys. **7** (2019) 174, arXiv: 1908.07759 [hep-ph]
(cit. on pp. 29, 308).

[73] K. Agashe and R. Contino
Composite Higgs-Mediated FCNC
Phys. Rev. D **80** (2009) 075016, arXiv: 0906.1542 [hep-ph]
(cit. on pp. 29, 308).

[74] G. Perez and L. Randall
Natural Neutrino Masses and Mixings from Warped Geometry
JHEP **01** (2009) 077, arXiv: 0805.4652 [hep-ph] (cit. on pp. 29, 308).

[75] A. Azatov, M. Toharia, and L. Zhu
Higgs Mediated FCNC's in Warped Extra Dimensions
Phys. Rev. D **80** (2009) 035016, arXiv: 0906.1990 [hep-ph]
(cit. on pp. 29, 308).

[76] R. Harnik, J. Kopp, and J. Zupan
Flavor Violating Higgs Decays
JHEP **03** (2013) 026, arXiv: 1209.1397 [hep-ph]
(cit. on pp. 30, 264, 265).

[77] MEG Collaboration
New limit on the lepton-flavour violating decay $\mu^+ \to e^+\gamma$
Phys. Rev. Lett. **107** (2011) 171801, arXiv: 1107.5547 [hep-ex]
(cit. on p. 30).

[78] BaBar Collaboration
Searches for Lepton Flavor Violation in the Decays tau+- —> e+- gamma and tau+- —> mu+- gamma
Phys. Rev. Lett. **104** (2010) 021802, arXiv: 0908.2381 [hep-ex]
(cit. on pp. 30, 264).

[79] I. Doršner et al.
New Physics Models Facing Lepton Flavor Violating Higgs Decays at the Percent Level
JHEP **06** (2015) 108, arXiv: 1502.07784 [hep-ph] (cit. on p. 31).

[80] Bertl, Wilhelm H. et al.
A Search for muon to electron conversion in muonic gold
Eur. Phys. J. C **47** (2006) 337 (cit. on pp. 31, 308).

[81] Adam, J. et al.
New constraint on the existence of the $\mu^+ \to e^+\gamma$ decay
Phys. Rev. Lett. **110** (2013) 201801, arXiv: 1303.0754 [hep-ex]
(cit. on pp. 31, 308).

[82] ATLAS Collaboration
Search for the Higgs boson decays $H \to ee$ and $H \to e\mu$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
Phys. Lett. B **801** (2020) 135148, arXiv: 1909.10235 [hep-ex]
(cit. on pp. 31, 214, 264, 308, 323).

[83] CMS Collaboration
Search for the lepton-flavor violating decay of the Higgs boson and additional Higgs bosons in the $e\mu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV
Phys. Rev. D **108** (2023) 072004, arXiv: 2305.18106 [hep-ex]
(cit. on pp. 31, 308).

[84] ATLAS Collaboration
Searches for lepton-flavour-violating decays of the Higgs boson in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector
Phys. Lett. B **800** (2020) 135069, arXiv: 1907.06131 [hep-ex]
(cit. on pp. 31, 213, 308, 323).

[85] CMS Collaboration
Search for lepton-flavor violating decays of the Higgs boson in the $\mu\tau$ and $e\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV
Phys. Rev. D **104** (2021) 032013, arXiv: 2105.03007 [hep-ex]
(cit. on pp. 31, 214, 275, 308, 323).

[86] *Snowmass White Paper Contribution: Physics with the Phase-2 ATLAS and CMS Detectors*
(2022), ATL-PHYS-PUB-2022-018 (cit. on pp. 32, 308).

[87] ATLAS Collaboration
Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector
Eur. Phys. J. C **82** (2022) 717, arXiv: 2201.11428 [hep-ex]
(cit. on p. 32).

[88] CMS Collaboration
Search for Higgs Boson Decay to a Charm Quark-Antiquark Pair in Proton-Proton Collisions at s=13 TeV
Phys. Rev. Lett. **131** (2023) 061801, arXiv: 2205.05550 [hep-ex]
(cit. on p. 32).

[89] ATLAS Collaboration
Measurements of WH and ZH Higgs production with decays into bottom quarks and direct constraints on the charm Yukawa coupling with 13 TeV collisions in the ATLAS detector.
Tech. rep., ATLAS-CONF-2024-010: CERN, 2024,
URL: https://cds.cern.ch/record/2905263 (cit. on p. 32).

[90] G. Degrassi et al.
Higgs mass and vacuum stability in the Standard Model at NNLO
JHEP **08** (2012) 098, arXiv: 1205.6497 [hep-ph] (cit. on pp. 32, 309).

[91] M. Reichert et al.
Probing baryogenesis through the Higgs boson self-coupling
Phys. Rev. D **97** (2018) 075008, arXiv: 1711.00019 [hep-ph]
(cit. on pp. 32, 309).

[92] F. L. Bezrukov and M. Shaposhnikov
The Standard Model Higgs boson as the inflaton
Phys. Lett. B **659** (2008) 703, arXiv: 0710.3755 [hep-th]
(cit. on pp. 32, 309).

[93] A. V. Bednyakov et al.
Stability of the Electroweak Vacuum: Gauge Independence and Advanced Precision
Phys. Rev. Lett. **115** (2015) 201802, arXiv: 1507.08833 [hep-ph]
(cit. on p. 33).

[94] D. E. Morrissey and M. J. Ramsey-Musolf
Electroweak baryogenesis
New J. Phys. **14** (2012) 125003, arXiv: 1206.2942 [hep-ph]
(cit. on pp. 33, 309).

[95] S. M. Barr and A. Zee
A New Approach to the electron-Muon Mass Ratio
Phys. Rev. D **15** (1977) 2652 (cit. on pp. 33, 309).

[96] K. S. Babu and S. Nandi
Natural fermion mass hierarchy and new signals for the Higgs boson
Phys. Rev. D **62** (2000) 033002, arXiv: hep-ph/9907213
(cit. on pp. 33, 309).

[97] A. de Gouvêa
Neutrino Mass Models
Ann. Rev. Nucl. Part. Sci. **66** (2016) 197 (cit. on pp. 33, 309).

[98] B. Patt and F. Wilczek
Higgs-field portal into hidden sectors
(2006), MIT-CTP-3745, arXiv: hep-ph/0605188 (cit. on pp. 33, 309).

[99] C. P. Burgess, M. Pospelov, and T. ter Veldhuis
The Minimal model of nonbaryonic dark matter: A Singlet scalar
Nucl. Phys. B **619** (2001) 709, arXiv: hep-ph/0011335
(cit. on pp. 33, 309).

[100] *LHC Machine*
JINST **3** (2008) S08001, ed. by L. Evans and P. Bryant
(cit. on pp. 35, 309).

[101] E. A. Mobs
The CERN accelerator complex. Complexe des accélérateurs du CERN (2016), General Photo, URL: https://cds.cern.ch/record/2225847 (cit. on p. 36).

[102] ATLAS Collaboration
The ATLAS Experiment at the CERN Large Hadron Collider
JINST **3** (2008) S08003 (cit. on pp. 36, 37, 42, 47, 48, 309).

[103] CMS Collaboration
The CMS Experiment at the CERN LHC
JINST **3** (2008) S08004 (cit. on p. 36).

[104] LHCb Collaboration
The LHCb Detector at the LHC
JINST **3** (2008) S08005 (cit. on p. 36).

[105] ALICE Collaboration
The ALICE experiment at the CERN LHC
JINST **3** (2008) S08002 (cit. on p. 36).

[106] G. C. Strong
On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case
Mach. Learn. Sci. Tech. **1** (2020) 045006,
arXiv: 2002.01427 [physics.data-an] (cit. on p. 38).

[107] ATLAS Collaboration
ATLAS inner detector: Technical Design Report, 1
Technical design report. ATLAS, CERN, 1997,
URL: https://cds.cern.ch/record/331063 (cit. on pp. 38, 39, 310).

[108] ATLAS Collaboration
ATLAS inner detector: Technical Design Report, 2
Technical design report. ATLAS, CERN, 1997,
URL: https://cds.cern.ch/record/331064 (cit. on pp. 38, 310).

[109] M. Capeans et al.
ATLAS Insertable B-Layer Technical Design Report
Tech. rep., 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 39).

[110] ATLAS Collaboration
Alignment of the ATLAS Inner Detector in Run-2
Eur. Phys. J. C **80** (2020) 1194, arXiv: 2007.07624 [hep-ex] (cit. on p. 40).

[111] ATLAS Collaboration
ATLAS liquid-argon calorimeter: Technical Design Report
Technical design report. ATLAS, CERN, 1996,
URL: https://cds.cern.ch/record/331061 (cit. on pp. 41, 310).

[112] ATLAS Collaboration
ATLAS Tile Calorimeter: Technical Design Report
Technical design report. ATLAS, CERN, 1996,
URL: https://cds.cern.ch/record/331062 (cit. on pp. 41, 310).

[113] ATLAS Collaboration
Approved Plots for the Tile Calorimeter (DAQ)
[Accessed: 11-August-2024], 2024, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTile (cit. on pp. 43, 46, 58).

[114] M. Crouau et al.
Characterization of 8-stages Hamamatsu R5900 photomultipliers for the TILE calorimeter
(1997) (cit. on p. 44).

[115] C. A. Solans Sánchez
Implementation of the ROD Crate DAQ Software for the ATLAS Tile Calorimeter and a Search for a MSSM Higgs Boson decaying into Tau pairs
PhD thesis: Valencia U., IFIC, 2010 (cit. on p. 44).

[116] Anderson, K. et al.
Design of the front-end analog electronics for the ATLAS tile calorimeter
Nucl. Instrum. Meth. A **551** (2005) 469 (cit. on pp. 45, 310).

[117] Berglund, S. et al.
The ATLAS tile calorimeter digitizer
JINST **3** (2008) P01004 (cit. on p. 45).

[118] J. Christiansen, A. Marchioro, and P. Moreira
TTCrx, an ASIC for timing, trigger and control distribution in LHC experiments
2nd Workshop on Electronics for LHC Experiments, 1996 161
(cit. on p. 45).

[119] K. Anderson et al.
ATLAS tile calorimeter interface
8th Workshop on Electronics for LHC Experiments, 2002 (cit. on p. 45).

[120] ATLAS Collaboration
ATLAS muon spectrometer: Technical Design Report
Technical design report. ATLAS, CERN, 1997,
URL: https://cds.cern.ch/record/331068 (cit. on pp. 47, 310).

[121] P. Jenni et al.
ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity
(2008), CERN-LHCC-2008-004, ATLAS-TDR-18 (cit. on pp. 49, 311).

[122] L. Adamczyk et al.
Technical Design Report for the ATLAS Forward Proton Detector
Tech. rep., CERN-LHCC-2015-009, ATLAS-TDR-024, 2015,
URL: https://cds.cern.ch/record/2017378 (cit. on pp. 50, 311).

[123] Abdel Khalek, S. et al.
The ALFA Roman Pot Detectors of ATLAS
JINST **11** (2016) P11013, arXiv: 1609.00249 [physics.ins-det]
(cit. on pp. 50, 311).

[124] ATLAS Collaboration
Zero degree calorimeters for ATLAS
(2007), CERN-LHCC-2007-01 (cit. on pp. 50, 311).

[125] ATLAS Collaboration
Operation of the ATLAS trigger system in Run 2
JINST **15** (2020) P10004, arXiv: 2007.12539 [physics.ins-det]
(cit. on p. 50).

[126] ATLAS Collaboration
Approved Plots for the Data Acquisition (DAQ)
[Accessed: 11-August-2024], 2024, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ
(cit. on p. 51).

[127] K. Bos et al.
LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)
Technical design report. LCG, CERN, 2005,
URL: https://cds.cern.ch/record/840543 (cit. on p. 51).

[128] *Worldwide LHC Computing Grid*
[Online; accessed 15-Sep-2024], 2024, URL: https://wlcg-docs.web.cern.ch/outreach/images/WLCG-Tiers-2021_v3.png
(cit. on p. 52).

[129] Zurbano Fernandez, I. et al.
High-Luminosity Large Hadron Collider (HL-LHC): Technical design report
10/2020 (2020), ed. by I. Béjar Alonso et al. (cit. on pp. 53, 311).

[130] ATLAS Collaboration
Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment
(2012), LHCC-I-023, CERN-LHCC-2012-022 (cit. on pp. 53, 311).

[131] ATLAS Collaboration
Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System
Tech. rep., CERN-LHCC-2017-020, ATLAS-TDR-029: CERN, 2017, URL: https://cds.cern.ch/record/2285584 (cit. on pp. 54, 311).

[132] ATLAS Collaboration
Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter
Tech. rep., CERN-LHCC-2017-019, ATLAS-TDR-028: CERN, 2017, URL: https://cds.cern.ch/record/2285583 (cit. on pp. 54, 56, 78, 312).

[133] F. Tang et al.
Upgrade Analog Readout and Digitizing System for ATLAS TileCal Demonstrator
IEEE Trans. Nucl. Sci. **62** (2015) 1045 (cit. on p. 55).

[134] E. Valdes Santurio, S. Silverstein, and C. Bohm
Readiness of the ATLAS Tile Calorimeter link daughterboard for the High Luminosity LHC era
PoS **TWEPP2019** (2020) 087 (cit. on p. 55).

[135] F. Carrió and A. Valero
The PreProcessor module for the ATLAS Tile Calorimeter at the HL-LHC
Nucl. Instrum. Meth. A **958** (2020) 162487, ed. by Manfred Krammer et al. (cit. on p. 55).

[136] F. Carrió
Design of the Compact Processing Module for the ATLAS Tile Calorimeter
IEEE Trans. Nucl. Sci. **68** (2021) 1944,
arXiv: 2011.02684 [physics.ins-det] (cit. on p. 55).

[137] A. Cervelló et al.
A new data transfer scheme for the HL-LHC upgrade of the ATLAS Tile Hadronic Calorimeter
J. Phys. Conf. Ser. **2374** (2022) 012097 (cit. on p. 55).

[138] ATLAS Collaboration
Operation and performance of the ATLAS tile calorimeter in LHC Run 2 (2024), arXiv: 2401.16034 [hep-ex] (cit. on p. 57).

[139] Fullana, E. et al.
Optimal Filtering in the ATLAS Hadronic Tile Calorimeter (2005) (cit. on p. 57).

[140] Barrand, G. et al.
GAUDI - A software architecture and framework for building HEP data processing applications
Comput. Phys. Commun. **140** (2001) 45 (cit. on pp. 59, 312).

[141] P. Calafiura et al.
The StoreGate: A Data model for the Atlas software architecture
eConf **C0303241** (2003) MOJT008, arXiv: cs/0306089
(cit. on pp. 59, 312).

[142] P. Calafiura et al.
The Athena control framework in production, new developments and lessons learned
14th International Conference on Computing in High-Energy and Nuclear Physics, 2005 456 (cit. on pp. 59, 312).

[143] I. Jen-La Plante and M. Tylmad
Pulse shapes for signal reconstruction in the ATLAS Tile Calorimeter
Nucl. Instrum. Meth. A **617** (2010) 96, ed. by G. Chiarelli et al.
(cit. on p. 70).

[144] J. C. Collins and D. E. Soper
Parton Distribution and Decay Functions
Nucl. Phys. B **194** (1982) 445 (cit. on p. 83).

[145] J. C. Collins and D. E. Soper
The Theorems of Perturbative QCD
Ann. Rev. Nucl. Part. Sci. **37** (1987) 383 (cit. on p. 84).

[146] Peskin, M. and Schroeder, D.
An Introduction to quantum field theory
Addison-Wesley, 1995,
URL: http://www.slac.stanford.edu/~mpeskin/QFT.html
(cit. on p. 84).

[147] V. N. Gribov and L. N. Lipatov
Deep inelastic e p scattering in perturbation theory
Sov. J. Nucl. Phys. **15** (1972) 438 (cit. on pp. 84, 314).

[148] G. Altarelli and G. Parisi
Asymptotic Freedom in Parton Language
Nucl. Phys. B **126** (1977) 298 (cit. on pp. 84, 314).

[149] 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 (cit. on pp. 84, 314).

[150] Buckley, A. et al.
General-purpose event generators for LHC physics
Phys. Rept. **504** (2011) 145, arXiv: 1101.2599 [hep-ph]
(cit. on pp. 84, 314).

[151] T. Gleisberg et al.
Event generation with SHERPA 1.1
JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on p. 85).

[152] P. Z. Skands
QCD for Collider Physics
2010 European School of High Energy Physics, 2011,
arXiv: 1104.2863 [hep-ph] (cit. on p. 85).

[153] G. C. Fox and S. Wolfram
A Model for Parton Showers in QCD
Nucl. Phys. B **168** (1980) 285 (cit. on p. 86).

[154] M. L. Mangano, M. Moretti, and R. Pittau
Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study
Nucl. Phys. B **632** (2002) 343, arXiv: hep-ph/0108069
(cit. on pp. 87, 314).

[155] S. Catani et al.
QCD matrix elements + parton showers
JHEP **11** (2001) 063, arXiv: hep-ph/0109231 (cit. on pp. 87, 314).

[156] Rikkert Frederix and Stefano Frixione
Merging meets matching in MC@NLO
JHEP **12** (2012) 061, arXiv: 1209.6215 [hep-ph] (cit. on p. 87).

[157] B. Andersson et al.
Parton Fragmentation and String Dynamics
Phys. Rept. **97** (1983) 31 (cit. on pp. 87, 314).

[158] J. C. Winter, F. Krauss, and G. Soff
A Modified cluster hadronization model
Eur. Phys. J. C **36** (2004) 381, arXiv: hep-ph/0311085
(cit. on pp. 87, 314).

[159] Z. Marshall
Simulation of Pile-up in the ATLAS Experiment
J. Phys. Conf. Ser. **513** (2014) 022024, ed. by D. L. Groep and D. Bonacorsi (cit. on pp. 88, 314).

[160] Agostinelli, S. et al.
GEANT4–a simulation toolkit
Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on pp. 89, 314).

[161] C. ATLAS et al.
The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim
Tech. rep., ATL-PHYS-PUB-2010-013: CERN, 2010,
URL: https://cds.cern.ch/record/1300517 (cit. on pp. 89, 314).

[162] K. Edmonds et al.
The Fast ATLAS Track Simulation (FATRAS)
Tech. rep., CERN, 2008, URL: https://cds.cern.ch/record/1091969
(cit. on pp. 89, 314).

[163] T. Sjöstrand et al.
An introduction to PYTHIA 8.2
Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph]
(cit. on pp. 89, 203).

[164] Bahr, M. et al.
Herwig++ Physics and Manual
Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph]
(cit. on pp. 90, 203).

[165] Bothmann, E. et al.
Event Generation with Sherpa 2.2
SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph]
(cit. on pp. 90, 93).

[166] J. Alwall et al.
The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations
JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph]
(cit. on pp. 90, 92, 203, 315).

[167] S. Frixione, P. Nason, and G. Ridolfi
A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction
JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph] (cit. on pp. 90, 203).

[168] S. Alioli et al.
A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX
JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on pp. 90, 203).

[169] 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] (cit. on p. 90).

[170] P. Nason
A New method for combining NLO QCD with shower Monte Carlo algorithms
JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on p. 90).

[171] ATLAS Collaboration
ATLAS data quality operations and performance for 2015–2018 data-taking
JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det]
(cit. on p. 90).

[172] ATLAS Collaboration
Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC
Eur. Phys. J. C **83** (2023) 982, arXiv: 2212.09379 [hep-ex]
(cit. on pp. 90, 160, 315).

[173] R. V. Harlander and K. J. Ozeren
Finite top mass effects for hadronic Higgs production at next-to-next-to-leading order
JHEP **11** (2009) 088, arXiv: 0909.3420 [hep-ph] (cit. on pp. 91, 315).

[174] R. V. Harlander and K. J. Ozeren
Top mass effects in Higgs production at next-to-next-to-leading order QCD: Virtual corrections
Phys. Lett. B **679** (2009) 467, arXiv: 0907.2997 [hep-ph]
(cit. on pp. 91, 315).

[175] R. V. Harlander et al.
Higgs production in gluon fusion at next-to-next-to-leading order QCD for finite top mass
Eur. Phys. J. C **66** (2010) 359, arXiv: 0912.2104 [hep-ph]
(cit. on pp. 91, 315).

[176] A. Pak, M. Rogal, and M. Steinhauser
Finite top quark mass effects in NNLO Higgs boson production at LHC
JHEP **02** (2010) 025, arXiv: 0911.4662 [hep-ph] (cit. on pp. 91, 315).

[177] M. Bonetti, K. Melnikov, and L. Tancredi
Higher order corrections to mixed QCD-EW contributions to Higgs boson production in gluon fusion
Phys. Rev. D **97** (2018) 056017, [Erratum: Phys.Rev.D 97, 099906 (2018)], arXiv: 1801.10403 [hep-ph] (cit. on pp. 91, 315).

[178] Butterworth, J. et al.
PDF4LHC recommendations for LHC Run II
J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph]
(cit. on pp. 92, 154, 315).

[179] 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] (cit. on p. 92).

[180] Bellm, J. et al.
Herwig 7.1 Release Note
(2017), arXiv: 1705.06919 [hep-ph] (cit. on pp. 92, 315).

[181] Ball, R. D. et al.
Parton distributions for the LHC Run II
JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on pp. 92, 315).

[182] A. Buckley
ATLAS Pythia 8 tunes to 7 TeV data
6th International Workshop on Multiple Partonic Interactions at the LHC, 2014 29 (cit. on p. 92).

[183] D. J. Lange
The EvtGen particle decay simulation package
Nucl. Instrum. Meth. A **462** (2001) 152, ed. by S. Erhan, P. Schlein, and Y. Rozen (cit. on p. 92).

[184] Bellm, J. et al.
Herwig 7.0/Herwig++ 3.0 release note
Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph]
(cit. on p. 92).

[185] A. Djouadi, J. Kalinowski, and M. Spira
HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension
Comput. Phys. Commun. **108** (1998) 56, arXiv: hep-ph/9704448
(cit. on pp. 93, 315).

[186] M. Spira
QCD effects in Higgs physics
Fortsch. Phys. **46** (1998) 203, arXiv: hep-ph/9705337 (cit. on pp. 93, 315).

[187] A. Djouadi, M. M. Muhlleitner, and M. Spira
Decays of supersymmetric particles: The Program SUSY-HIT (SUspect-SdecaY-Hdecay-InTerface)
Acta Phys. Polon. B **38** (2007) 635, ed. by K. Fialkowski and B. Muryn, arXiv: hep-ph/0609292 (cit. on pp. 93, 315).

[188] A. Bredenstein et al.
Radiative corrections to the semileptonic and hadronic Higgs-boson decays H —> W W / Z Z —> 4 fermions
JHEP **02** (2007) 080, arXiv: hep-ph/0611234 (cit. on pp. 93, 315).

[189] A. Bredenstein et al.
Precise predictions for the Higgs-boson decay H —> WW/ZZ —> 4 leptons
Phys. Rev. D **74** (2006) 013004, arXiv: hep-ph/0604011
(cit. on pp. 93, 315).

[190] A. Bredenstein et al.
Precision calculations for the Higgs decays H —> ZZ/WW —> 4 leptons
Nucl. Phys. B Proc. Suppl. **160** (2006) 131, ed. by J. Blumlein, S. Moch, and T. Riemann, arXiv: hep-ph/0607060 (cit. on pp. 93, 315).

[191] T. Gleisberg and S. Hoeche
Comix, a new matrix element generator
JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 93).

[192] F. Buccioni et al.
OpenLoops 2
Eur. Phys. J. C **79** (2019) 866, arXiv: 1907.13071 [hep-ph]
(cit. on p. 93).

[193] F. Cascioli, P. Maierhofer, and S. Pozzorini
Scattering Amplitudes with Open Loops
Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph]
(cit. on p. 93).

[194] A. Denner, S. Dittmaier, and L. Hofer
Collier: a fortran-based Complex One-Loop LIbrary in Extended Regularizations
Comput. Phys. Commun. **212** (2017) 220, arXiv: 1604.06792 [hep-ph]
(cit. on p. 93).

[195] C. Anastasiou et al.
High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO
Phys. Rev. D **69** (2004) 094008, arXiv: hep-ph/0312266 (cit. on p. 93).

[196] ATLAS Collaboration
Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO
Tech. rep., ATL-PHYS-PUB-2017-007: CERN, 2017,
URL: https://cds.cern.ch/record/2261938 (cit. on p. 93).

[197] L. A. Harland-Lang et al.
Parton distributions in the LHC era: MMHT 2014 PDFs
Eur. Phys. J. C **75** (2015) 204, arXiv: 1412.3989 [hep-ph] (cit. on p. 93).

[198] M. Aliev et al.
HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR
Comput. Phys. Commun. **182** (2011) 1034, arXiv: 1007.1327 [hep-ph]
(cit. on p. 94).

[199] T. G. Cornelissen et al.
Updates of the ATLAS Tracking Event Data Model (Release 13)
Tech. rep., ATL-SOFT-PUB-2007-003, ATL-COM-SOFT-2007-008: CERN, 2007, URL: https://cds.cern.ch/record/1038095 (cit. on p. 96).

[200] ATLAS Collaboration
Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2
Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on pp. 95, 316).

[201] R. Fruhwirth
Application of Kalman filtering to track and vertex fitting
Nucl. Instrum. Meth. A **262** (1987) 444 (cit. on p. 95).

[202] ATLAS Collaboration
Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC
Eur. Phys. J. C **77** (2017) 332, arXiv: 1611.10235 [physics.ins-det] (cit. on pp. 96, 316).

[203] ATLAS Collaboration
Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s}$ = 13 TeV
Tech. rep., ATL-PHYS-PUB-2015-026: CERN, 2015, URL: https://cds.cern.ch/record/2037717 (cit. on pp. 96, 316).

[204] ATLAS Collaboration
Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data
JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on pp. 97, 160, 215, 256, 293, 316).

[205] ATLAS Collaboration
Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment
JHEP **05** (2024) 162, arXiv: 2308.13362 [hep-ex] (cit. on pp. 97, 99, 100, 316).

[206] ATLAS Collaboration
Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1
Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 97).

[207] ATLAS Collaboration
Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung
Tech. rep., CERN, 2012, URL: https://cds.cern.ch/record/1449796 (cit. on p. 97).

[208] ATLAS Collaboration
Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV
Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [physics.ins-det] (cit. on pp. 98, 293).

[209] ATLAS Collaboration
Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV
Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex] (cit. on pp. 100, 102, 104, 317).

[210] M. Cacciari, G. P. Salam, and G. Soyez
The anti-k_t jet clustering algorithm
JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on pp. 104, 317).

[211] ATLAS Collaboration
Jet reconstruction and performance using particle flow with the ATLAS Detector
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on pp. 105, 256).

[212] ATLAS Collaboration
Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
Eur. Phys. J. C **81** (2021) 689, arXiv: 2007.02645 [hep-ex] (cit. on pp. 105, 158, 317).

[213] 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, arXiv: 1703.09665 [hep-ex] (cit. on pp. 106, 317).

[214] ATLAS Collaboration
Tagging and suppression of pileup jets
Tech. rep., ATL-PHYS-PUB-2014-001: CERN, 2014, URL: http://cds.cern.ch/record/1643929 (cit. on p. 106).

[215] ATLAS Collaboration
ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV
Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex]
(cit. on pp. 106, 160, 317).

[216] ATLAS Collaboration
ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset
Eur. Phys. J. C **83** (2023) 681, arXiv: 2211.16345 [physics.data-an]
(cit. on pp. 106, 107, 317).

[217] ATLAS Collaboration
Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment
Tech. rep., CERN, 2019, URL: https://cds.cern.ch/record/2688062
(cit. on pp. 108, 109, 256, 317).

[218] ATLAS Collaboration
The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using 140 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions
(2024), arXiv: 2402.05858 [hep-ex] (cit. on pp. 110, 256, 317).

[219] Workman, R. L. et al.
Review of Particle Physics
PTEP **2022** (2022) 083C01 (cit. on pp. 113, 318).

[220] 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]
(cit. on pp. 113, 143, 318).

[221] 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. D **99** (2019) 072001, arXiv: 1811.08856 [hep-ex]
(cit. on pp. 114, 318).

[222] ATLAS Collaboration
Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
JHEP **08** (2022) 175, arXiv: 2201.08269 [hep-ex]
(cit. on pp. 114, 115, 125–127, 135, 137, 200, 274, 318).

[223] CMS Collaboration
Measurements of Higgs boson production in the decay channel with a pair of τ leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV
Eur. Phys. J. C **83** (2023) 562, arXiv: 2204.12957 [hep-ex]
(cit. on pp. 114, 273, 318).

[224] ATLAS Collaboration
Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector
Phys. Rev. D **97** (2018) 072003, arXiv: 1712.08891 [hep-ex]
(cit. on p. 114).

[225] B. Grzadkowski et al.
Dimension-Six Terms in the Standard Model Lagrangian
JHEP **10** (2010) 085, arXiv: 1008.4884 [hep-ph] (cit. on pp. 116, 208).

[226] ATLAS Collaboration
Differential cross-section measurements of Higgs boson production in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
(2024), arXiv: 2407.16320 [hep-ex] (cit. on pp. 116, 123, 319).

[227] M. Hübner
Higgs Boson Production Cross-Section Measurements in the Di-Tau Final State at ATLAS
PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn U., 2022 (cit. on p. 122).

[228] ATLAS Collaboration
Interpretations of the ATLAS measurements of Higgs boson production and decay rates and differential cross-sections in pp collisions at $\sqrt{s} = 13$ TeV
(2024), arXiv: 2402.05742 [hep-ex] (cit. on p. 131).

[229] ATLAS Collaboration
Measurement of the VH $H \to \tau\tau$ process with the ATLAS detector at 13 TeV
Phys. Lett. B **855** (2024) 138817, arXiv: 2312.02394 [hep-ex]
(cit. on p. 134).

[230] ATLAS Collaboration
Formulae for Estimating Significance
Tech. rep., ATL-PHYS-PUB-2020-025: CERN, 2020,
URL: https://cds.cern.ch/record/2736148 (cit. on p. 136).

[231] ATLAS Collaboration
Evaluation of QCD uncertainties for Higgs boson production through gluon fusion and in association with two top quarks for simplified template cross-section measurements
Tech. rep., CERN, 2023, URL: http://cds.cern.ch/record/2878797 (cit. on p. 155).

[232] I. W. Stewart and F. J. Tackmann
Theory Uncertainties for Higgs and Other Searches Using Jet Bins
Phys. Rev. D **85** (2012) 034011, arXiv: 1107.2117 [hep-ph] (cit. on p. 155).

[233] T. Appelquist and J. Carazzone
Infrared Singularities and Massive Fields
Phys. Rev. D **11** (1975) 2856 (cit. on p. 207).

[234] C. Hays et al.
On the impact of dimension-eight SMEFT operators on Higgs measurements
JHEP **02** (2019) 123, arXiv: 1808.00442 [hep-ph] (cit. on p. 208).

[235] I. Brivio, Y. Jiang, and M. Trott
The SMEFTsim package, theory and tools
JHEP **12** (2017) 070, arXiv: 1709.06492 [hep-ph] (cit. on pp. 208, 209).

[236] I. Brivio
SMEFTsim 3.0 — a practical guide
JHEP **04** (2021) 073, arXiv: 2012.11343 [hep-ph] (cit. on p. 209).

[237] S. Bressler, A. Dery, and A. Efrati
Asymmetric lepton-flavor violating Higgs boson decays
Phys. Rev. D **90** (2014) 015025, arXiv: 1405.4545 [hep-ph] (cit. on pp. 214, 323).

[238] ATLAS Collaboration
Searches for lepton-flavour-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector
JHEP **07** (2023) 166, arXiv: 2302.05225 [hep-ex] (cit. on pp. 215, 323).

[239] Hocker, A. et al.
TMVA - Toolkit for Multivariate Data Analysis
(2007), CERN-OPEN-2007-007, arXiv: physics/0703039 (cit. on p. 234).

[240] G. Cowan et al.
Asymptotic formulae for likelihood-based tests of new physics
Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur.Phys.J.C 73, 2501 (2013)], arXiv: 1007.1727 [physics.data-an] (cit. on pp. 235, 258).

[241] ATLAS Collaboration
Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV
Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex]
(cit. on p. 256).

[242] ATLAS Collaboration
Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC
Tech. rep., CERN, 2019, URL: https://cds.cern.ch/record/2677054
(cit. on p. 256).

[243] A. L. Read
Presentation of search results: The CL_s technique
J. Phys. G **28** (2002) 2693, ed. by M. R. Whalley and L. Lyons
(cit. on pp. 258, 288).

[244] S. S. Wilks
The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses
Annals Math. Statist. **9** (1938) 60 (cit. on pp. 267, 287).

[245] ATLAS Collaboration
Measurements of Higgs boson production by gluon-gluon fusion and vector-boson fusion using $H \to WW^* \to e\nu\mu\nu$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
Phys. Rev. D **108** (2023) 032005, arXiv: 2207.00338 [hep-ex]
(cit. on p. 273).

[246] ATLAS Collaboration
Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s}$ =13 TeV with the ATLAS detector
Eur. Phys. J. C **80** (2020) 957, [Erratum: Eur.Phys.J.C 81, 29 (2021), Erratum: Eur.Phys.J.C 81, 398 (2021)], arXiv: 2004.03447 [hep-ex]
(cit. on p. 273).

[247] ATLAS Collaboration
Measurement of the properties of Higgs boson production at $\sqrt{s} = 13$ TeV in the $H \to \gamma\gamma$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment
JHEP **07** (2023) 088, arXiv: 2207.00348 [hep-ex] (cit. on p. 273).

[248] ATLAS Collaboration
Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = 13$ TeV
Eur. Phys. J. C **81** (2021) 537, arXiv: 2011.08280 [hep-ex]
(cit. on p. 273).

[249] CMS Collaboration
An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data
JINST **14** (2019) P06032, arXiv: 1903.01216 [hep-ex] (cit. on p. 276).

[250] ATLAS Collaboration
Searches of lepton-flavour-violating decays of the Higgs bosons with the ATLAS detector at the HL-LHC
Tech. rep., CERN, 2022, URL: https://cds.cern.ch/record/2841245 (cit. on p. 276).

[251] F. James and M. Winkler
MINUIT User's Guide
(2004) (cit. on p. 280).

[252] Roger J. Barlow and Christine Beeston
Fitting using finite Monte Carlo samples
Comput. Phys. Commun. **77** (1993) 219 (cit. on p. 282).

[253] J. Neyman and E. S. Pearson
On the Problem of the Most Efficient Tests of Statistical Hypotheses
Phil. Trans. Roy. Soc. Lond. A **231** (1933) 289 (cit. on p. 287).

Agradecimientos

Como no podría ser de otra manera, escribo estos agradecimientos apresurado, casi a última hora, poco antes de depositar esta tesis. Es la única forma en la que puedo hacerlo de manera auténtica. Si estos agradecimientos hubieran sido escritos con más calma, quizá no serían realmente míos. Lamento si olvido a alguien, porque no solo debo recordar a quienes han estado directamente implicados en esta tesis, sino también a todos aquellos que, a lo largo de mi vida, han encendido en mí la curiosidad. Perdonadme si no os menciono, o si lo hago con menos efusividad de la que merecéis. Quizá sea mejor así: siempre he sido una persona introvertida y prefiero que mis palabras de gratitud lleguen en persona.

En primer lugar, tengo que darle las gracias a Luca, quien no solo me ha guiado por el camino de la investigación, sino que me ha dado la libertad y la oportunidad de explorar siempre que lo he necesitado. Gracias por tu paciencia, por tu apoyo y por tu confianza. Si tuviera que volver a comenzar esta tesis, te volvería a elegir sin duda.

No podrían faltar palabras de agradecimiento para mis compañeros del IFIC y para el resto de seniors del grupo de ATLAS. Debo hacer una mención especial a todas las personas relacionadas con el grupo de TileCal. En primer lugar, a los juniors que me acogieron al llegar, Adam e Iván. También a los que han ido llegando después a mi vida, y que han hecho que mi estancia en el grupo haya sido siempre agradable. A los seniors: Victoria, Alberto, Arantxa, y sobre todo a Ximo. Gracias, Ximo, por tu apoyo durante estos últimos años, por los cafés y por tu capacidad para apaciguar el caos que siempre me rodea.

Tengo que dar gracias con un punto y aparte a Fernando y a Toni, con quienes he compartido innumerables horas de despacho en el CERN. Gracias por la compañía, los cafés y las conversaciones, no siempre inofensivas, pero siempre necesarias.

My deepest gratitude goes to the TileCal team at CERN. I have grown as a physicist (or at least tried) within this group, and I truly feel that without the TileCal team, this thesis would not have been the same. My term as Run Coordinator marked a before and after in my PhD, and I am grateful for the opportunity to work with such a talented and dedicated group of people. I feel honored to have been part of this team, and I hope I have been able to contribute even a small fraction of what I have received. Mentioning names is always dangerous, but I have to specially thank Sasha Solodkov, the first TileCal person I met, who has been not only an infinite source of knowledge but also an example of dedication and hard work. I also want to thank Oleg Solovyanov, Henric Wilkens, Jalal Abdallah, Danijela Bogavac, Filipe Martins, Irakli Minashvili, Helena Santos, Tomas Davidek and Jana Faltova for their support and help during these years.

A big part of this thesis has been developed in collaboration with the amazing teams working on physics analysis in the ATLAS experiment. Trying to respect the chronological order, I have to thank the people I met during my first adventures

with the LFV analysis, and I would like to mention especially Katharina and Valerie, who were always there to help and answer all my questions, and Julia, who was always a great support to panic with before every approval. I only have words of gratitude for the $H \to \tau\tau$ team, with special mention to Antonio de Maria, Chris Young, and Kathrin Becker, who have always been examples of dedication and excellence in physics.

Tamar deserves a separate paragraph. Her good mood and her enthusiasm have always been a source of inspiration for me since the first time we met in the ugly, cold, and dirty test beam area. Thanks a lot for the coffee breaks and the laughs (and the gossiping). These last years would have been much more boring and harder without your support, your company, and your friendship.

Another special mention goes to my dear friend and colleague, Shalini. Thanks a lot for the ice creams in front of a cult temple, the endless gossiping and chats about life.

I cannot end the CERN section of these acknowledgements without mentioning my wonderful friend and office mate, Michaela. CERN would have been a much darker place without the coffees, the dinners we shared and the safe space you always provided. Thanks for making it possible to believe that a normal life is achievable even in the middle of the LHC.

Vuelvo ahora al castellano para entrar en la parte más personal de estos agradecimientos. Me gustaría empezar, aunque sin alejarme demasiado de la física, agradeciendo a mis amigos de la carrera, quienes han sido un apoyo constante durante el desquiciante y alborotado viaje que fue el grado en Física. Gracias a Calata y a Berni por los años compartidos y por las penurias vividas juntos. Pero, sobre todo, quiero dedicar unas palabras especiales a David. Gracias por estar siempre ahí durante tantos años de camino académico, y gracias por las innumerables comidas compartidas en las que, jamás, bajo ninguna circunstancia, se criticó a un tercero. Sin esos momentos de desahogo y risas, no estoy seguro de si habría llegado vivo a este punto.

Cuántas cosas se pueden decir del Colegio Mayor San Juan de Ribera. Quién sabe qué habría sido de mí si no hubiera llegado a ese lugar. Cada una de las personas con las que he compartido estos años en el Colegio ha dejado una huella imborrable en mí. Gracias a todos, y especialmente a aquellos que han sabido aguantarme y que incluso, tal vez, me han cogido cariño.

En primer lugar, quiero dar las gracias a mis compañeros de generación: Carlos, Rasmi, Francís, Quico y Álex. Carlos, gracias por ser siempre mi compañero en las desdichas, y por las infinitas risas en la distopía que era el estudio durante los exámenes. A mi querido amigo Francís, gracias por ser siempre un apoyo incondicional, por las conversaciones interminables a pesar de la distancia, y por sostenerme en los momentos más bajos.

A María, el alma del Colegio, gracias por tu energía y tu capacidad de aportar alegría y entusiasmo a un cascarrabias como yo. Eres una inspiración para todos los que te rodeamos. A mi queridísima hija Karina, con quien he compartido locuras

y gritos en medio de las cenas de fin de semana. A Max, amigo y, me atrevo a decir, discípulo converso (quizás incluso debería pedirte perdón por ello), gracias por enseñarme tanto, por tu curiosidad inagotable sobre la vida y la realidad, por tu amistad y por tu mente siempre abierta. Finalmente, gracias a Adrián, quien ha sido un apoyo incondicional y fundamental desde hace muchos años. Gracias por tu cariño y amistad, por ser mi compañero en decenas de cenas y copas dignas de señoras de bien.

Quiero dedicar unas palabras (aunque seguramente insuficientes) a Santiago Pons y a José Vicente Puig. Santi, gracias por todos estos años compartidos en el Colegio, por tu sensatez y por tu capacidad de hacernos frenar, aunque solo fuera por un instante, en medio de la locura. Gracias por tu amistad, que espero conservar durante muchos años más. José Vicente, querido amigo, gracias por tu buen humor y por hacer que todo parezca más sencillo. Pero, sobre todo, gracias por tu profundidad y tu sensibilidad. Gracias por tu paz, que siempre me ha ayudado a encontrar la mía.

Gracias a mis amigas y amigos de Albacete, por soportar mis interminables monólogos cada vez que regresaba con una nueva obsesión.

Gracias a Cristian, el Orfeo capaz de calmar y apaciguar a mi Cancerbero interior. Y a mis padres y abuelos, a los que siempre miro y miraré como ejemplo de esfuerzo y superación.

Gracias a todas y a todos por sostenerme. LDVM.

València, a 29 de noviembre de 2024.