



Lancaster  
University

---

DOCTORAL THESIS

---

Measuring the CP properties of the  
Higgs boson in its decay to  $\tau$  leptons  
with the ATLAS detector

*A thesis submitted in fulfilment of the requirements  
for the degree of Doctor of Philosophy*

*in the*

*Experimental Particle Physics Research Group, Physics Department*

*Author:*

Sebastiano Spinali

*Supervisor:*

Harald Fox

26th June 2023

*To Mum, Dad mentors of my life.*

*To my brother and my sisters supporting columns always present.*

*To my grandmother, a constant upright inspiration.*

*To all my friends, the real reasons to carry on.*

I, Sebastiano Spinali, hereby declare that this thesis has not been and will not be, submitted in whole or in part to another university for the award of any other degree.

*Lancaster University 26th June 2023*

---

Sebastiano Spinali

Lancaster University  
Physics Department  
Experimental Particle Physics Research Group

## DOCTORAL THESIS

---

# Measuring the CP properties of the Higgs boson in its decay to $\tau$ leptons with the ATLAS detector

---

by Sebastiano Spinali

## ABSTRACT

This thesis presents a study of the CP properties of the Standard Model (SM) Higgs boson interaction with  $\tau$  leptons. In those SM interactions the CP symmetry is conserved. The measurement of the properties is performed by defining a CP sensitive angular observable  $\varphi_{CP}^*$ , which can be kinematically constructed from the Higgs visible decay products. The interactions occurring between the SM Higgs particle and the  $\tau$  particles are described by making use of a mixing angle  $\phi_\tau$  appearing in the Yukawa coupling. The two channels considered in the thesis are the  $\tau_{\text{had}}\tau_{\text{had}}$  and the  $\tau_{\text{lep}}\tau_{\text{had}}$  channels, where  $\tau_{\text{had}}$  denotes a hadronic decay of the  $\tau$  and  $\tau_{\text{lep}}$  denotes a leptonic  $\tau$  decay. The study is performed using proton-proton collision data recorded between 2015 and 2018 with the ATLAS detector at the Large Hadron Collider. The measured (expected)  $\phi_\tau$  value is equal to  $8.68^\circ \pm 16^\circ$  ( $0^\circ \pm 28^\circ$ ) at the 68% confidence level (CL), and  $8.68^\circ \pm 34^\circ$  ( $-70^\circ / +75^\circ$ ) at the 95.5% CL. The pure CP-odd hypothesis is rejected at a value of  $3.4\sigma$  (standard deviation), which implies a compatibility between the measurement and the Higgs prediction in the Standard Model.

# CONTENTS

<b>List of Figures</b>	<b>ix</b>
<b>Introduction</b>	<b>1</b>
<b>1 Theoretical Background</b>	<b>4</b>
1.1 The Standard Model . . . . .	4
1.1.1 Quantum Electrodynamics . . . . .	6
1.1.2 Weak Interaction . . . . .	7
1.1.3 Electroweak Unification . . . . .	7
1.2 CP-symmetry violation . . . . .	8
1.3 Brout-Englert-Higgs mechanism . . . . .	11
1.3.1 Lepton Higgs Coupling . . . . .	14
1.4 Higgs Production . . . . .	15
1.4.1 Gluon-gluon fusion . . . . .	15
1.4.2 Vector boson fusion . . . . .	16
1.4.3 Vector boson associated production $Vh$ . . . . .	17
$t\bar{t}$ associated production . . . . .	17
1.5 Higgs Boson Decay . . . . .	18
<b>2 The ATLAS Experiment at the LHC</b>	<b>20</b>
2.1 The Large Hadron Collider . . . . .	20
2.2 The ATLAS Detector . . . . .	22
2.2.1 The Magnet System . . . . .	24
2.2.2 The Inner Detector . . . . .	25
2.2.3 The Calorimeters . . . . .	28
2.2.4 The Muon Spectrometer . . . . .	30
2.3 The ATLAS Trigger System . . . . .	31

<b>3</b>	<b>Object reconstruction</b>	<b>33</b>
3.1	Primary vertex and track parameters . . . . .	33
3.2	Object Reconstruction . . . . .	34
3.2.1	Electron Reconstruction . . . . .	34
3.2.2	Muon Reconstruction . . . . .	35
3.2.3	Tau Hadronic Reconstruction . . . . .	36
3.2.4	Jet Reconstruction . . . . .	38
3.2.5	Missing Transverse Energy . . . . .	38
3.2.6	Overlap Removal . . . . .	39
3.3	Tau Decay Plane Reconstruction . . . . .	39
<b>4</b>	<b>Analysis strategy</b>	<b>43</b>
4.1	CP-measurements via $\phi_\tau$ mixing angle. . . . .	45
4.1.1	IP method . . . . .	48
4.1.2	$\rho$ method . . . . .	50
4.1.3	IP- $\rho$ method . . . . .	52
4.1.4	$a_1$ method . . . . .	54
4.2	Analysis Decay mode . . . . .	56
4.3	Background estimation . . . . .	58
<b>5</b>	<b>Data and Simulated samples</b>	<b>60</b>
5.1	Data . . . . .	60
5.2	Monte Carlo simulations . . . . .	62
5.2.1	Signal generator . . . . .	62
5.2.2	Background generator . . . . .	65
<b>6</b>	<b>Event selection</b>	<b>67</b>
6.1	Trigger selection . . . . .	67
6.2	Preselection . . . . .	68
6.3	Signal regions . . . . .	71
	Event categorisation regions . . . . .	71
6.3.1	Decay mode and signal optimisation regions . . . . .	72
6.3.2	Fit Signal Regions . . . . .	72
6.4	Control regions for $Z\tau\tau$ and multijet background . . . . .	73
<b>7</b>	<b>Systematic Uncertainties</b>	<b>74</b>
7.1	Experimental Uncertainties . . . . .	75
7.1.1	Jet uncertainties . . . . .	75
7.1.2	Missing transverse energy uncertainties . . . . .	76

7.1.3	Lepton uncertainties . . . . .	77
7.1.4	Luminosity uncertainties . . . . .	77
7.2	Theoretical Uncertainties . . . . .	78
7.3	Analysis-specific uncertainties . . . . .	80
	Tau decay mode classification uncertainties . . . . .	80
	$\varphi_{CP}^*$ systematic uncertainties . . . . .	81
	$\tau_{lep}\tau_{had}$ - Fake Systematic Uncertainties. . . . .	81
	$\tau_{had}\tau_{had}$ - Fake Systematic Uncertainties. . . . .	82
7.4	Data-driven fake estimation uncertainties . . . . .	82
7.5	RNN shape uncertainties . . . . .	85
7.6	Primary vertex uncertainties . . . . .	89
7.7	Smoothing Options . . . . .	100
<b>8</b>	<b>Results</b>	<b>109</b>
8.1	Fit Model . . . . .	109
	8.1.1 CP templates . . . . .	111
	8.1.2 Treatment of nuisance parameters (NP) . . . . .	111
	8.1.3 Sensitivity regions . . . . .	112
	8.1.4 Binning Optimisation . . . . .	115
	8.1.5 Prefit distribution . . . . .	115
8.2	Sensitivity . . . . .	121
8.3	Fit results . . . . .	121
	8.3.1 Asimov fit results . . . . .	122
	8.3.2 Unblinded fit results . . . . .	127
<b>9</b>	<b>Conclusion</b>	<b>132</b>
<b>10</b>	<b>Appendix A</b>	<b>134</b>
10.1	$\varphi_{CP}^*$ distributions at preselection level . . . . .	134
10.2	ZCR $\varphi_{CP}^*$ distributions at preselection level . . . . .	143
10.3	Observed post-fit distributions . . . . .	152
<b>11</b>	<b>Appendix B</b>	<b>155</b>
11.1	Lepton systematic uncertainties list . . . . .	155
11.2	Jet systematic uncertainties list . . . . .	157
11.3	$\phi^*$ distributions for the RNN shape uncertainties studies . . . . .	158
11.4	Primary vertex distributions studies . . . . .	161
11.5	Coordinate residual plots . . . . .	165
11.6	2D Reconstructed vertex coordinates . . . . .	169

---

11.7 $\varphi^*$ distributions . . . . .	172
11.8 $\varphi^*$ residual comparison plots . . . . .	173
11.9 Smoothing Options studies . . . . .	175
<b>12 Glossary</b>	<b>180</b>
<b>Bibliography</b>	<b>183</b>



# LIST OF FIGURES

1.1	Standard Model of particles. . . . .	6
1.2	Representation of the scalar field potential $V(\phi)$ , commonly known as the Mexican hat potential. . . . .	12
1.3	Leading order Feynman diagram for Higgs Boson decay to two fermions (left), two bosons (center) and two $\gamma$ (right). . . . .	18
1.4	Standard Model Higgs decay branching ratio as a function of the Higgs mass. . . . .	19
2.1	CERN Accelerator complex. The LHC is the last ring (dark gray line). Smaller machines are used for early-stage acceleration and also to provide beams for other experiments [1]. . . . .	22
2.2	Cut-away view of the ATLAS detector. The dimensions of the detector are 25 m in height and 44 m in length. The overall weight of the detector is approximately 7000 tonnes [1]. . . . .	23
2.3	The ATLAS magnet system. . . . .	24
2.4	Overview of the ATLAS ID with labels and dimensions. [2]. . . . .	26
2.5	Diagram of the ATLAS ID and its sub-detectors. . . . .	26
2.6	A computer generated image of the full calorimeter. . . . .	28
2.7	Cut-away view of the ATLAS muon system [3]. . . . .	31
3.1	Feynman diagram of the $\tau$ hadronic and leptonic decays. . . . .	36
3.2	A transverse view of the detector. The blue arrows represent observed particles and their momentum while the red arrow represent the missing energy of the particle(s) that did not interact with the detector but due to the conservation laws can be calculated. . . . .	39

3.3	Efficiency matrix of the $\tau$ particle flow decay mode classification [4]. Each element in the efficiency matrix is normalized by the total number of events in its corresponding generated decay mode (column), while an element in the purity matrix is normalized by the total events in the corresponding reconstructed decay mode (row). . . . .	40
3.4	Efficiency matrix of the Sherpa 2.2.1 $Z \rightarrow \tau\tau$ MC sample used in the analysis. Each element in the efficiency matrix is normalized by the total number of events in its corresponding generated decay mode (column), while an element in the purity matrix is normalized by the total events in the corresponding reconstructed decay mode (row). . . . .	41
3.5	Efficiency matrix of the Powheg + Pythia8 $H \rightarrow \tau\tau$ ggH signal sample used in the analysis. Each element in the efficiency matrix is normalized by the total number of events in its corresponding generated decay mode (column), while an element in the purity matrix is normalized by the total events in the corresponding reconstructed decay mode (row). . . . .	42
4.1	Summary of Higgs boson production and decay modes by the ATLAS collaboration [5]. . . . .	44
4.2	Angular correlations of the tau decay products from $H/A \rightarrow \tau^+\tau^-$ decays. H represents the scalar Higgs boson while A is a pseudoscalar Higgs boson. . . . .	46
4.3	Illustration of a decay $H \rightarrow \tau^+\tau^- \rightarrow \pi^+\pi^- + 2\nu$ in the Zero Momentum Frame. The decay planes spanned by the momenta of the tau leptons and the pions are shown, together with the angle they form [6]. . . . .	47
4.4	Normalised $\phi_{CP}^*$ distribution for $pp \rightarrow H/Z^*/\gamma^* \rightarrow \tau^+\tau^-$ production [6]. . . . .	47
4.5	Definition of the Impact Parameter vector [7] $n^\pm$ in the decay plane of either $\tau^\pm \rightarrow \pi^\pm\nu_\tau$ or $\tau^\pm \rightarrow l^\pm\nu_\tau$ [6]. . . . .	49
4.6	Geometrical view of the $\varphi^*$ angle construction in the IP-method [8]. . . . .	49
4.7	$\phi_{CP}^*$ shape distribution for 1p0n-1p0n VBF preselection high(left) and low(right). . . . .	50
4.8	The $\phi_{CP}^*$ shape distributions for 1p1n-1p1n VBF preselection high(left) and low(right). . . . .	51
4.9	The $\phi_{CP}^*$ shape distributions 1p1n-1pXn 1pXn-1p1n VBF preselection high(left) and low(right). . . . .	52
4.10	$\phi_{CP}^*$ shape distribution for 1p0n-1p1n 1p1n-1p0n VBF preselection high(left) and low(right). . . . .	53
4.11	$\phi_{CP}^*$ shape distribution for 1p0n-1pXn 1pXn-1p0n VBF preselection high(left) and low(right). . . . .	53
4.12	$\phi_{CP}^*$ shape distribution for l-1p1n VBF preselection high(left) and low(right). . . . .	54

4.13	$\phi_{CP}^*$ shape distribution for 1-1pXn VBF preselection high(left) and low(right). . .	54
4.14	$\phi_{CP}^*$ shape distribution for 1p0n-3p0n VBF preselection high(left) and low(right). . .	55
4.15	$\phi_{CP}^*$ shape distribution for 1-3p0n VBF preselection high(left) and low(right). . .	56
4.16	Higgs decay in a $\tau^+\tau^-$ couple. The $\tau^+$ decays hadronically producing a quark-antiquark while the $\tau^-$ decays leptonically producing an electron and electron-antineutrino. . . . .	57
4.17	Figure 4.17a shows the Feynman diagram of a $\tau$ decays into an electron, an electron-antineutrino and a tau-antineutrino. Figure 4.17b shows the Feynman diagram of the hadronic $\tau$ decay in a couple quark-antiquark plus a tau-antineutrino and a $\pi^+$ particle (1p0n decay mode) while figure 4.17c shows the hadronic $\tau$ decays into a $\pi^+$ and a $\pi^0$ (1p1n decay mode). . . . .	57
7.1	Calibration stages for EM-scale jets. Other than the origin correction, each stage of the calibration is applied to the four-momentum of the jet [9]. . . . .	75
7.2	Illustration of the track-based soft term projections with regard to $p_T^{hard}$ for the systematic uncertainty calculation, ref. [10]. . . . .	77
7.3	Boost Loose $\phi^*$ shape distributions shown for the different RNN bins. . . . .	87
7.4	The 3 bins $\phi^*$ envelopes and the "ztt_rnn_shape" systematic uncertainty are shown for the different signal regions. . . . .	88
7.5	PV coordinates (x, y, z) and the distance in the detector transverse plane (Rxy) for the VBF CP-even sample. Comparison between the default BS constraint, no BS constraint and 55% of the BS constraint is performed. . . . .	92
7.6	Coordinate residual plots of the VBF CP-even sample. . . . .	93
7.7	First row shows true z position (y-axis) vs z position (x-axis) plots, second row shows true z position (y-axis) vs z residual (x-axis) plots, third row shows true x position (y-axis) vs x position (x-axis) plots and fourth row shows true x position (y-axis) vs x residual (x-axis) plots. From left to right the plots correspond to the noBS case, default case and soft BS case. All of them are calculated using the VBF even sample. . . . .	94
7.8	$\varphi^*$ distributions for MC truth for the four signal samples. . . . .	96
7.9	From left to right, in the upper row, the VBF CP-even $\varphi^*$ hybrid method distributions for default and no BS while in the lower row soft BS. . . . .	96
7.10	$\varphi^*$ residual plots of the three different hybrid scenarios against the MC for the ggH sample (top(left CP-even, right CP-odd)) and the VBF sample (bottom(left CP-even, right CP-odd)). . . . .	97

7.11	$\varphi^*$ residual plots of the three different scenarios against themselves for the ggH sample (top(left CP-even, right CP-odd)) and the VBF sample (bottom(left CP-even, right CP-odd)). . . . .	97
7.12	From left to right, in the upper row, the VBF CP-even $\varphi^*$ residual plot when considering (hybrid default - MC) and (hybrid noBS - MC) while in the lower row the (hybrid soft - MC) case. . . . .	98
7.13	From left to right, in the upper row, the VBF CP-even $\varphi^*$ residual plot when considering (hybrid default - hybrid noBS) and (hybrid soft - hybrid default) while in the lower row the (hybrid soft - hybrid noBS) case. . . . .	98
7.14	The <code>JET_JER_EffectiveNP_4</code> is shown, for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, in the boost tight high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic. . . . .	102
7.15	The <code>JET_JER_EffectiveNP_4</code> is shown, for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, in the boost loose high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic. . . . .	103
7.16	$\Delta\text{NLL}$ distribution showing the comparison among the different smoothing options. . . . .	107
8.1	Schematic summary of the fit models used in the analysis. They are grouped by topologies (Boosted (green) and VBF (red)) and by decay channels (HH for $\tau_{\text{had}}\tau_{\text{had}}$ channel and LH for $\tau_{\text{lep}}\tau_{\text{had}}$ channel). The boxes with dashed lines as boundaries indicate the free floating normalisation factors which are acting on various regions. . . . .	111
8.2	Expected $\Delta\text{NLL}$ curves for the $\tau_{\text{had}}\tau_{\text{had}}$ plus $\tau_{\text{lep}}\tau_{\text{had}}$ combined fit with only the statistical uncertainty of the Asimov dataset. The sensitivity of each fit can be estimated by $\Delta\text{NLL}$ at $\phi_\tau = \pm 90^\circ$ . The intercepts of the two $\Delta\text{NLL}$ curves with $\Delta\text{NLL}=0.5$ (grey horizontal dashed line) correspond to the 68% confidence levels on $\phi_\tau$ from the combined fit. . . . .	113
8.3	Maximum $\Delta\text{NLL}$ as a function of the number of bins of observable for the Asimov fit using 3, 4, 5, 7, and 9 bins in the observable $\phi_{CP}^*$ [6]. . . . .	115
8.4	Pre-fit distributions in the $H \rightarrow \tau\tau$ signal regions in the $\tau_{\text{had}}\tau_{\text{had}}$ channel, which will be included in the likelihood model. ‘OtherBkgs’ includes W, diboson, top, $Z \rightarrow \text{ll}, H \rightarrow \text{WW}$ samples. . . . .	117
8.5	Pre-fit distributions in the $H \rightarrow \tau\tau$ signal regions in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel, which will be included in the likelihood model. ‘OtherBkgs’ includes W, diboson, top, $Z \rightarrow \text{ll}, H \rightarrow \text{WW}$ samples. . . . .	118

8.6	Pre-fit distributions in the $Z \rightarrow \tau\tau$ control regions in both semi-leptonic and full-hadronic channels, which will be included in the likelihood model. ‘OtherBkgs’ includes W, diboson, top, $Z \rightarrow ll, H \rightarrow WW$ samples. . . . .	119
8.7	Pre-fit distributions in the $Z \rightarrow \tau\tau$ control regions for the $\rho$ mass distribution. . . . .	120
8.8	Expected post-fit values and uncertainties from the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ combined fit on the Asimov dataset for the signal strength and background normalisations (NF) for $Z \rightarrow \tau\tau$ samples in VBF and Boost regions. . . . .	123
8.9	Expected ranking of 40 systematic uncertainties with the highest post-fit impact on $\varphi_\tau$ in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ combined fit on the Asimov dataset. . . . .	124
8.10	Expected $\Delta LNN$ curves for the $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$ combined Asimov fit with all statistical and systematic uncertainties. The expected value for $\varphi_\tau$ is $0 \pm 29$ at one standard deviation. . . . .	125
8.11	Expected 2D $\Delta NLL$ contours as functions of $H \rightarrow \tau\tau$ signal strength $\mu^{\tau\tau}$ versus CP mixing angle $\varphi_\tau$ for the $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$ had combined Asimov fit with all statistical and systematic uncertainties (stat+syst). . . . .	125
8.12	Expected post-fit values and uncertainties from the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ combined fit on the unblinded data for the signal strength and background normalisations (NF) for $Z \rightarrow \tau\tau$ samples in VBF and Boost regions. . . . .	128
8.13	Expected ranking of 40 systematic uncertainties with the highest post-fit impact on $\varphi_\tau$ in the $\tau_{\text{had}}\tau_{\text{had}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$ combined fit on the unblinded data. . . . .	129
8.14	Expected $\Delta LNN$ curves for the $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$ combined data fit with all statistical and systematic uncertainties. The expected value for $\varphi_\tau$ is $0^\circ \pm 29^\circ$ at one standard deviation. . . . .	130
8.15	Expected 2D $\Delta NLL$ contours as functions of $H \rightarrow \tau\tau$ signal strength $\mu^{\tau\tau}$ versus CP mixing angle $\varphi_\tau$ for the $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$ had combined data fit with all statistical and systematic uncertainties (stat+syst). . . . .	130
8.16	Data - background prefit distribution for the HH VBF High region, data and background distribution and the relative $\Delta LNN$ distribution comparison between expected and observed. . . . .	131
10.1	Semi-leptonic $\varphi_{CP}^*$ distribution, for $0^\circ$ and $90^\circ$ degree templates, of different CP-hypotheses in the di- $\tau$ mode combination for the high optimisation regions. . . . .	135
10.2	Fully hadronic $\varphi_{CP}^*$ distribution, for $0^\circ$ and $90^\circ$ degree templates, of different CP-hypotheses in the di- $\tau$ mode combination for the high optimisation regions. . . . .	136
10.3	Semi-leptonic $\varphi_{CP}^*$ distribution, for $0^\circ$ and $90^\circ$ degree templates, of different CP-hypotheses in the di- $\tau$ mode combination for the low optimisation regions. . . . .	137

10.4 Fully hadronic $\varphi_{CP}^*$ distribution, for $0^\circ$ and $90^\circ$ degree templates, of different CP-hypotheses in the di- $\tau$ mode combination for the low optimisation regions. . . . .	138
10.5 Semi-leptonic $\varphi_{CP}^*$ distribution, for $40^\circ$ and $140^\circ$ degree templates, of different CP-hypotheses in the di- $\tau$ mode combination for the high optimisation regions. . . . .	139
10.6 Fully hadronic $\varphi_{CP}^*$ distribution, for $40^\circ$ and $140^\circ$ degree templates, of different CP-hypotheses in the di- $\tau$ mode combination for the high optimisation regions. . . . .	140
10.7 Semi-leptonic $\varphi_{CP}^*$ distribution, for $40^\circ$ and $140^\circ$ degree templates, of different CP-hypotheses in the di- $\tau$ mode combination for the low optimisation regions. . . . .	141
10.8 Fully hadronic $\varphi_{CP}^*$ distribution, for $40^\circ$ and $140^\circ$ degree templates, of different CP-hypotheses in the di- $\tau$ mode combination for the low optimisation regions. . . . .	142
10.9 $\phi^*$ in preselection and Z control regions in the $\tau_{had}\tau_{had}$ channel. . . . .	144
10.10 $\phi^*$ in preselection and Z control regions in the $\tau_{lep}\tau_{had}$ channel. . . . .	145
10.11 $\phi^*$ in different decay channels in preselection events in the $\tau_{had}\tau_{had}$ channel. . . . .	146
10.12 $\phi^*$ in different decay channels in preselection events in the $\tau_{lep}\tau_{had}$ channel. . . . .	147
10.13 $\phi^*$ by decay channels in the boost Z control region in the $\tau_{had}\tau_{had}$ channel. . . . .	148
10.14 $\phi^*$ by decay channels in the boost Z control region in the $\tau_{lep}\tau_{had}$ channel. . . . .	149
10.15 $\phi^*$ by decay channels in the VBF Z control region in the $\tau_{had}\tau_{had}$ channel. . . . .	150
10.16 $\phi^*$ by decay channels in the VBF Z control region in the $\tau_{lep}\tau_{had}$ channel. . . . .	151
10.17 Observed post-fit distributions in the $H \rightarrow \tau\tau$ signal regions in $\tau_{had}\tau_{had}$ channels, which will be included in the likelihood model. ‘OtherBkgs’ includes $W$ , diboson, top, $Z \rightarrow ll$ , $H \rightarrow WW$ samples. . . . .	152
10.18 Observed post-fit distributions in the $H \rightarrow \tau\tau$ signal regions in $\tau_{lep}\tau_{had}$ channels, which will be included in the likelihood model. ‘OtherBkgs’ includes $W$ , diboson, top, $Z \rightarrow ll$ , $H \rightarrow WW$ samples. . . . .	153
10.19 Observed post-fit distributions in the Z( $\rightarrow \tau\tau$ ) control regions in $\tau_{had}\tau_{had}$ and $\tau_{lep}\tau_{had}$ channels, which will be included in the likelihood model. ‘OtherBkgs’ includes $W$ , diboson, top, $Z \rightarrow ll$ , $H \rightarrow WW$ samples. . . . .	154
11.1 Boost Tight $\phi^*$ shape distributions shown for the different RNN bins. . . . .	159
11.2 VBF $\phi^*$ shape distributions shown for the different RNN bins. . . . .	160
11.3 PV coordinates (x, y, z) and the distance in the detector transverse plane (Rxy) for the VBF CP-odd sample. Comparison between the MC values and the default BS constraint, no BS constraint and 55% of the BS constraint is performed. . . . .	162
11.4 PV coordinates (x, y, z) and the distance in the detector transverse plane (Rxy) for the ggH CP-even sample. Comparison between the MC values and the default BS constraint, no BS constraint and 55% of the BS constraint is performed. . . . .	163

11.5	PV coordinates (x, y, z) and the distance in the detector transverse plane ( $R_{xy}$ ) for the ggH CP-odd sample. Comparison between the MC values and the default BS constraint, no BS constraint and 55% of the BS constraint is performed. . . . .	164
11.6	Coordinate residual plots of the VBF CP-odd sample. . . . .	166
11.7	Coordinate residual plots of the ggH CP-even sample. . . . .	167
11.8	Coordinate residual plots of the ggH CP-odd sample. . . . .	168
11.9	First row shows true z position (y-axis) vs z position (x-axis) plots, second row shows row true z position (y-axis) vs z residual (x-axis) plots, third row shows true x position (y-axis) vs x position (x-axis) plots and forth row shows true x position (y-axis) vs x residual (x-axis) plots. From left to right the plots correspond to the noBS case, default case and soft BS case. All of them are calculated using the VBF odd sample. . . . .	169
11.10	First row shows true z position (y-axis) vs z position (x-axis) plots, second row shows row true z position (y-axis) vs z residual (x-axis) plots, third row shows true x position (y-axis) vs x position (x-axis) plots and forth row shows true x position (y-axis) vs x residual (x-axis) plots. From left to right the plots correspond to the noBS case, default case and soft BS case. All of them are calculated using the ggH even sample. . . . .	170
11.11	First row shows true z position (y-axis) vs z position (x-axis) plots, second row shows row true z position (y-axis) vs z residual (x-axis) plots, third row shows true x position (y-axis) vs x position (x-axis) plots and forth row shows true x position (y-axis) vs x residual (x-axis) plots. From left to right the plots correspond to the noBS case, default case and soft BS case. All of them are calculated using the ggH odd sample. . . . .	171
11.12	From left to right the VBF CP-odd $\varphi^*$ hybrid method distributions for default, no BS and soft BS. . . . .	172
11.13	From left to right the ggH CP-even $\varphi^*$ hybrid method distributions for default, no BS and soft BS. . . . .	172
11.14	From left to right the ggH CP-odd $\varphi^*$ hybrid methods distributions for default, no BS and soft BS. . . . .	172
11.15	From left to right the VBF CP-odd $\varphi^*$ residual plot when considering (hybrid default - MC), (hybrid noBS - MC) and (hybrid soft - MC) respectively. . . . .	173
11.16	From left to right the ggH CP-even $\varphi^*$ residual plot when considering (hybrid default - MC), (hybrid noBS - MC) and (hybrid soft - MC) respectively. . . . .	173
11.17	From left to right the ggH CP-odd $\varphi^*$ residual plot when considering (hybrid default - MC), (hybrid noBS - MC) and (hybrid soft - MC) respectively. . . . .	173

11.18	From left to right the VBF CP-odd $\varphi^*$ residual plot when considering (hybrid default - hybrid noBS), (hybrid soft - hybrid default) and (hybrid soft - hybrid noBS) respectively. . . . .	174
11.19	From left to right the ggH CP-even $\varphi^*$ residual plot when considering (hybrid default - hybrid noBS), (hybrid soft - hybrid default) and (hybrid soft - hybrid noBS) respectively. . . . .	174
11.20	From left to right the ggH CP-odd $\varphi^*$ residual plot when considering (hybrid default - hybrid noBS), (hybrid soft - hybrid default) and (hybrid soft - hybrid noBS) respectively. . . . .	174
11.21	The <code>TAUS_TRUEHADTAU_EFF_RNNID_HIGHPT</code> is shown, for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, in the boost tight high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic. . . . .	176
11.22	The <code>TAUS_TRUEHADTAU_EFF_RNNID_HIGHPT</code> is shown, for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, in the boost loose high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic. . . . .	177
11.23	The <code>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_1P0N_RECO_1P0N_TOTAL</code> is shown, for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, in the boost tight high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic. . . . .	178
11.24	The <code>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_1P0N_RECO_1P0N_TOTAL</code> is shown, for the $\tau_{\text{had}}\tau_{\text{had}}$ channel, in the boost loose high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic. . . . .	179



# INTRODUCTION

During the last five decades the SM [11, 12, 13, 14] of particle physics has been tested by many experiments confirming, by their results, a successful description of the high energy particle physics interactions. However, it was only in 2012 that the mechanism of electroweak symmetry breaking in the SM have been verified by the discovery of a resonance having properties consistent with those, predicted by the Higgs mechanism, of the Higgs boson. In particular, it is the mechanism itself which gives mass to massive elementary particles, that implies the existence of a scalar particle, the SM Higgs boson.

The Standard Model predict the Higgs boson to be purely CP-even, even though the existence of a CP-odd component is not ruled out yet and therefore the existence of a CP-mixed Higgs boson is still taken into account and it could be considered as evidence of physics beyond the Standard Model. There are several reasons for which obtaining evidence for CP violation is an important milestone to achieve. In particular it has a dominant role in the understanding of the matter-antimatter asymmetry in our universe by providing important information about the processes involved in its generation.

Different CP scenarios can be investigated by defining a model-independent variable derived from the geometry of the Higgs decay products. This analysis exploits the Higgs decay into two  $\tau$  leptons, the next to heaviest fermions into which the Higgs boson can decay. The  $\tau$  particle has a relatively short lifetime as a consequence; the signal in the detector is given by their decay products and thus tau decays can be classified hadronic or leptonic. In the hadronic case we detect neutral and charged pions while in the leptonic case electrons and muons. The reconstruction of the decay products allow the geometrical construction of the two  $\tau$ 's decay angle which is then correlated to a crucial variable of this analysis  $\phi_{CP}$  which permits one to determine the CP phase of the Higgs $\rightarrow\tau\tau$  decay. A mixing angle between CP-even and CP-odd states can be extracted by studying the transverse spin correlation of the two  $\tau$ 's, and eventually a likelihood fit allows one to

differentiate between the different CP scenarios.

An important role to evaluate the results is played by the uncertainties that affect the measurement and therefore need to be accounted for as precisely as possible. Within the whole set of uncertainties, both statistical fluctuations and systematic uncertainties have to be studied. Statistical fluctuations can impact the shape of the histograms that are used in the likelihood fit, while systematic uncertainties take into account all the persistent effects that can lead to an unvarying uncertainty during the whole analysis. The major sources of systematic uncertainties come from the reconstruction of objects, such as jets,  $\tau$ ,  $\pi$ ,  $e$  and  $\mu$ . They are therefore treated as additional free parameters in the likelihood fit with the possibility to be evaluated by the fit itself.

This thesis is organised as follows: After an introduction to the theoretical foundations in chapter 1, an overview on the detector is given in chapter 2. Then, the attention is pointed to the object reconstruction in chapter 3 which will highlight the reconstructed objects used in the following thesis. The different methods of measuring the CP-sensitive observables are then presented in chapter 4, followed by a description of which data, as well as Monte Carlo simulation samples, are used in the analysis in chapter 5. Subsequently, the individual steps of this analysis are outlined, beginning with the selection criteria applied to choose the events of interest in chapter 6. The estimation of the above mentioned systematic uncertainties, as well as several ongoing studies are presented in chapter 7. Ultimately, the actual likelihood fit and its results are presented in chapter 8 while conclusions are reported in chapter 9.

## Contributions

This analysis sets a limit on the CP mixing angle  $\phi_\tau$  which offers a new opportunity to find new sources of CP violation in the Higgs interaction with other SM particles which can give a hint on the nature of the universe, with a more precise look on the difference of matter and antimatter amount.

During my Ph.D. I have been involved in different aspect of the analysis, and my contributions can be summarised as follow:

- Recurrent Neural Network (RNN)  $Z \rightarrow \tau\tau$  uncertainty
- $\tau$  uncertainties
- smoothing and pruning studies
- fit group member
- 1D and 2D likelihood function development
- primary vertex beam spot constraint
- development of the fit package

# THEORETICAL BACKGROUND

# 1

A brief introduction of the theoretical framework of particle physics is shown in this section. The theory behind, known as SM, is discussed in section 1.1 with a major focus on the spontaneous symmetry breaking mechanism in section 1.3. Section 1.4 and section 1.5 are dedicated to the Higgs Boson, to its production modes and its decays respectively.

## 1.1 The Standard Model

The most general theory able to describe how fundamental particles behave and to explain the interactions occurring between them is known as the SM of particle physics. It has been very useful during the last decades for its power to predict important "placeholders" in the discovery of nature's laws. As an example, the SM is not able to predict the mass of the Higgs Boson, as well as individual masses of any particles, but it is very powerful to predict the ratio of the Z and W boson masses that is linked to the ratio of the weak and electromagnetic force strength and therefore the SM describes the electroweak interaction and Quantum Chromodynamics (QCD).

The mathematical theory behind the SM is a non-abelian gauge theory where two different sectors, the strong interaction and the electroweak interaction, take place. The strong sector is described by the non-abelian group  $SU(3)_C$  while the electroweak sector by the  $SU(2)_L \otimes U(1)_Y$  that lead the SM to be described by the following non-abelian group:  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ .

The three known forces, electromagnetic, strong and weak, except gravity, interact by the mediation of a vector boson. as shown in 1.1. The properties of a vector boson are very peculiar. There is no baryonic number conservation applied to them, apart from the charged  $W^\pm$ , they are all electrically chargeless and the vector boson spin is an integer that makes the interaction more favored due to the possibility that the spin S can be coupled

to the orbital angular momentum  $L$ . The total angular momentum  $J$  is defined as the sum  $L + S$  while the exponent represents the parity of the particle that can be either  $+$  or  $-$ ,  $J^P$ .

Interaction	exchanged particle	mass	$J^P$
strong	gluons (g)	0	$1^-$
electromagnetic	photon ( $\gamma$ )	0	$1^-$
weak	$W^\pm, Z^0$	$\approx 10^2$	1

Table 1.1: The three known forces in the Universe described by the SM.

The strong interaction is mediated by the gluons which, due to the color charge they carry, can interact with themselves. Similarly to it, the weak interaction boson vectors carry the weak charge and therefore can interact with themselves as well.

In addition to those bosons, there are other fundamental particles, quarks and leptons, which have semi-integer spin ( $\frac{1}{2}$ ) that can be grouped together, by following the mass rule, in three different families or generations, as shown in table 1.2.

Fermions	Generation	electric charge	color	weak isospin	spin
leptons	$\nu_e \nu_\mu \nu_\tau$	0	-	$\frac{1}{2}$	$\frac{1}{2}$
	$e \mu \tau$	-1	-	$\frac{1}{2}$	$\frac{1}{2}$
Quarks	u c t	$\frac{2}{3}$	r b g	$\frac{1}{2}$	$\frac{1}{2}$
	d s b	$-\frac{1}{3}$	r b g	$\frac{1}{2}$	$\frac{1}{2}$

Table 1.2: The three fermionic generations of matter.

For each fermion, a particle having an half-odd-integer spin and therefore obeying the Pauli exclusion principle, there is an associated anti-fermion having the same mass but with opposite electric and color charge, and an opposite z Isospin component. The gluons are massless but, at the same time, the strong interaction range has a very short finite value due to the gluon self-interaction. This can be explained by looking at the color field associated energy, which increases with increasing distance. When the distance is  $\geq 1$  fm, the energy is enough to create a quark-antiquark pair, meaning that the "free" particle must be, from color point of view, neutral. An overview of the SM is given in Figure 1.1.

Even if the Standard Model, is the most powerful theory we have to describe the fundamental nature around us, it is not describing several observed phenomena such as neutrino oscillations, matter-antimatter asymmetry and the gravitational interaction. It is important to emphasize that, due to its weaker nature, compared to the other fundamental forces, the gravitational interaction is dozens of orders weaker than the other forces accessible to the LHC experiments at European Organization for Nuclear Research (CERN).

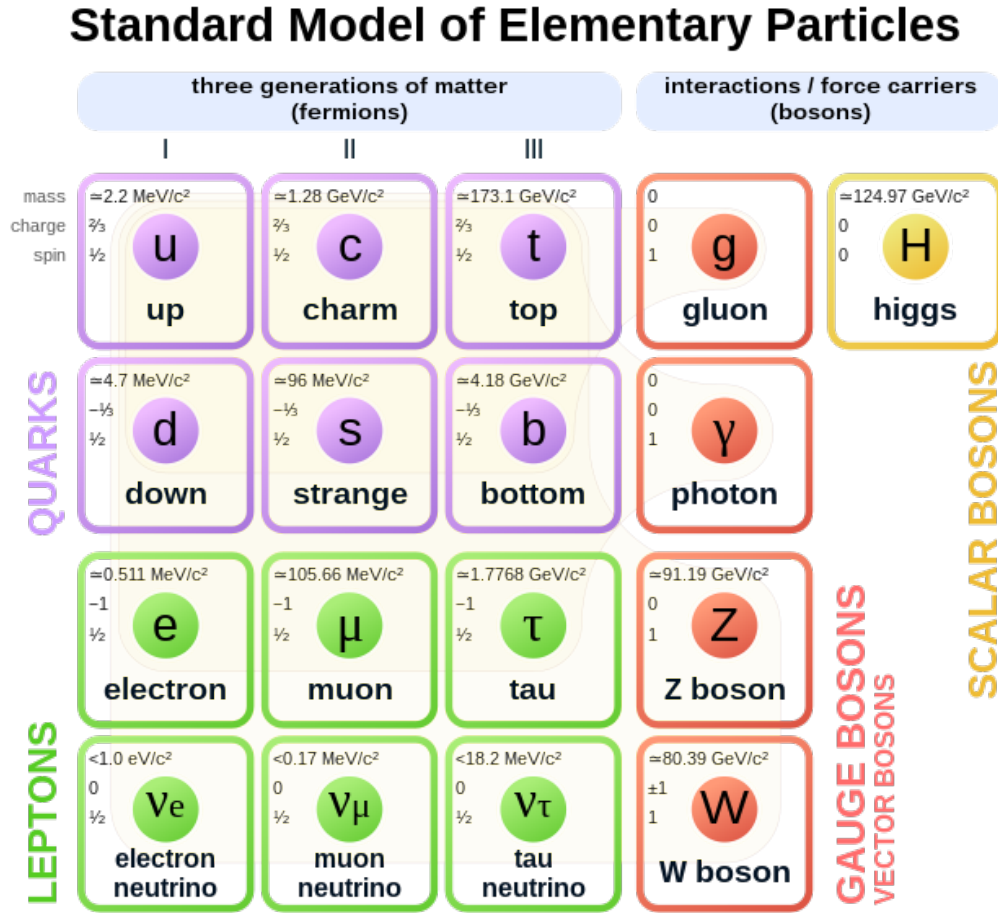


Figure 1.1: Standard Model of particles.

### 1.1.1 Quantum Electrodynamics

The use of gauge bosons to describe particle interactions was first developed as a part of the QED theory. The principal reason behind the formulation of this theory, made by Feynman, Schwinger and Tomonaga [15] was to expand the electromagnetism formalism to the elementary particle scheme.

If we consider a charged particle that satisfies the Dirac equation

$$(i\gamma^\mu \partial_\mu - m)\psi = 0, \quad (1.1)$$

we can state that the U(1) global symmetry is conserved while the gauge transformation is not conserved when local. This is overcome by considering an interaction term between the electromagnetic field and the particle itself in the Hamiltonian. This additional term simplifies the QED matrix elements calculations in a specific two body scatter process where two particles interact through the emission of a gauge boson, in this case a photon. The strength of such an interaction is proportional to the momentum exchange  $q^2$ . When the  $q^2$  assumes low values it is equal to the fine structure constant  $\alpha = 1/137$ .

### 1.1.2 Weak Interaction

Several symmetries are conserved in the domain of the Quantum Field Theory (QFT). Of particular importance is the quantum symmetry which is responsible for the uniformity of space, and this symmetry is called parity. In quantum mechanics, parity is described with an operator which, when applied to any spatial coordinate  $\mathbf{x}$  of a wave function at a specific time  $t$ , flips its sign into its mirror image:

$$\Psi(\mathbf{x}, t) \rightarrow \Psi'(\mathbf{x}, t) = \hat{P}\Psi(\mathbf{x}, t) = \Psi(-\mathbf{x}, t) \quad (1.2)$$

When the parity operator is applied in the QED framework it acts on the fermions by flipping the sign of the current and this leads to QED parity conservation.

However, when Wu carried out the experiment on the beta decay measurement of Co-60 [16], the data showed a non-conservation of the parity in the process and in general in the weak interaction processes. This led to the conclusion that there is a difference between the fermionic currents acting in the weak force and the fermionic currents occurring in QED processes.

### 1.1.3 Electroweak Unification

Both QED and the weak interaction are QFT theories which are not modified under certain symmetries of the universe. In particular, when applying the U(1) symmetry, a unitary group consisting of all complex numbers with an absolute value 1, to the QED process, the conserved quantity is the electric charge of the particle, while the similar quantity conserved in weak processes is known as Weak Isospin ( $I_W$ ), which is associated with the SU(2) symmetry. The Electroweak (EW) theory unifies the  $SU(2)_L$  weak isospin symmetry with the  $U(1)_Y$  group of weak hypercharge.

The violation of parity conservation in the  $SU(2)_L$  gauge group implies that the weak isospin current couples only to left-handed fermions where the spin projection onto the particle's momentum is negative. Neutral currents conserve lepton flavour, whereas the charged currents change the flavour of left-handed fermion fields.

The SU(2) symmetry introduces three possible weak bosons for the weak interaction that are usually referred to as  $W^1$ ,  $W^2$  and  $W^3$  [17]. In this scenario,  $W^1$  and  $W^2$  can represent the physical  $W^\pm$  bosons while  $W^3$  can be associated with a weak neutral current. That means that only left handed fermions could couple with the neutral current and this leads to a discrepancy with the experiments conducted so far. This is also the reason why

the  $W^3$  can not be associated with the physical Z boson discovered at CERN in 1973 [18] that is allowed to couple to either left-handed or right-handed fermions.

The solution was first suggested by Glashow, Salam and Weinberg [11, 13] by proposing the electroweak unification. The symmetry formed by the electroweak unification is therefore conserved under U(1) transformations, and is called  $U(1)_Y$ , where the Y represents the hypercharge that can be expressed in the following way:

$$Y = 2(q - T_3) \tag{1.3}$$

where  $q$  represent the electric charge and  $T_3$  represent the third component of the weak isospin. The unified electroweak theory suggests the existence of four massless mediators to let the EW interaction happen.

At the same time, the short range of the weak interaction would suggest a massive mediator. This inconsistency led to the idea of symmetry breaking implemented by Kibble, Guralnik and Hagen [19], Brout and Englert [20] and Higgs [21, 22] to give what is now known as the Brout-Englert-Higgs mechanism (BEH).

## 1.2 CP-symmetry violation

The connection between the geometric conservation laws (along with conservation of energy, of momentum and of angular momentum) and the symmetries (uniformity of time, homogeneity and isotropy of space) is of fundamental importance in contemporary physics. Until the 1950's, parity conservation was believed to be one of the fundamental geometric conservation laws. The parity symmetry means that the equations of particle physics are invariant under mirror inversion, that is the mirror image of a reaction occurs at the same rate as the original reaction.

In 1956 the theoretical physicists Lee and Yang stated that while parity conservation had been verified in decays via the strong or electromagnetic interactions, it was untested in the weak interaction and they proposed several possible direct experimental tests [23].

The first test based on beta decay of  $^{60}\text{Co}$  nuclei, carried out in 1956 by a group led by Chien Shiung Wu, demonstrated that weak interactions violate the P-symmetry (Parity symmetry) as some reactions did not occur as often as their mirror image [16].

However, parity symmetry still appears to be valid for all reactions involving electromagnetism and strong interactions.

The symmetry of a quantum mechanical system can be restored if another symmetry  $\mathcal{S}$  can be found such that the combined symmetry  $\mathcal{PS}$  remains unbroken. After the discovery of  $\mathcal{P}$  violation, it was proposed that charge conjugation,  $\mathcal{C}$ , which transforms a particle into its antiparticle, was the suitable symmetry to restore order. According to the  $\mathcal{CP}$ -



symmetry (a composition of  $\mathcal{C}$ -symmetry, charge conjugation and  $\mathcal{P}$ -symmetry, parity), there is no change in the physics laws if a particle is interchanged with its antiparticle ( $\mathcal{C}$ -symmetry) and the particle's coordinates are inverted ( $\mathcal{P}$ -symmetry).

James Cronin and Val Fitch were awarded the Nobel Prize in Physics in 1980 for their discovery of CP violation in 1964 in neutral kaon decays [24].

The CP violation discovered shows that neutral kaons can transform into their antiparticles (in which each quark is replaced with its antiquark) and vice versa, but such transformation does not occur with exactly the same probability in both direction (indirect  $\mathcal{CP}$  violation).

In 1990 the NA31 Experiment at CERN suggested evidence for CP violation in the decay process of the very same neutral kaons (direct CP violation). The observation was controversial, but the final proof came in 1999 from the KTeV experiment at Fermilab [25] and the NA48 Experiment at CERN [26].

In 2013 Large Hadron Collider beauty (LHCb) experiment announced discovery of CP violation in strange B meson decays [27]; in March 2019, LHCb announced discovery of CP violation in charmed  $D^0$  decays with a deviation from zero of 5.3 standard deviations [28]. In 2020, the Tokai to Kamioka (T2K) Collaboration reported some indications of CP violation in leptons neutrinos for the first time [29].

The matter-antimatter asymmetry of the Universe provides one of the primary motivations to study extensions of the Standard Model of Particle Physics.

The Universe is made up of matter, rather than of equal parts of matter and antimatter as might be expected. It can be demonstrated that to explain the baryons versus antibaryons asymmetry observed in the Universe, the Sakharov conditions must be satisfied. After the Big Bang, physical laws must have acted differently for matter and antimatter, i.e. violating CP-symmetry.

Various ideas have been proposed [30] [31] that satisfy the Sakharov conditions [32] to generate a net surplus of matter over anti-matter in the Universe, one of which is the existence of CP violation during the extreme conditions in the first seconds after the Big Bang.

Charge-Parity (CP) symmetry has to be violated to have different reaction rates for baryons and anti-baryons:

$$\Gamma(N \xrightarrow{\mathcal{L}(\Delta n_{Bar} \neq 0)} f) \neq \Gamma(\bar{N} \xrightarrow{\mathcal{L}(\Delta n_{Bar} \neq 0)} \bar{f}) \quad (1.4)$$

In the SM, CP violation is encoded in the Cabibbo–Kobayashi–Maskawa (CKM) matrix for the quarks and in the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix for leptons. A source of CP violation only appears in the charged current couplings, an effect too small to generate the observed matter-antimatter asymmetry. In extensions of the SM, the required CP violation is often obtained by introducing complex phases into the scalar

sector [33] or into Higgs-fermion interactions [34, 35]. Precision studies of the Higgs boson interactions, in particular CP-violating interactions, can help us to explain the basic mechanism of matter-antimatter asymmetry.

In the SM the Yukawa coupling matrix is real and thus contains no source of CP violation. However, the Higgs that we have observed may actually be a superposition of a CP-even and CP-odd Higgs. This allows for mixing of CP eigenstates like it is observed for K and B mesons. Just as in the meson case, this mixing makes a CP violating phase possible. This is approximated by the introduction of the  $\phi_{CP}^*$  mixing angle.

The measured level of CP violation in these processes and similar ones, however, is insufficient to explain the matter-antimatter asymmetry by itself. If CP violation also existed in the Higgs sector, it may have caused a little excess of matter over antimatter to form in the early cosmos. This may happen as a result of Higgs boson interactions or decay processes. Physics researchers are currently looking at the precise ways through which this may occur.

It is hypothesized that the baryogenesis, the creation of baryons such as protons and neutrons, is what led to the creation of the matter-antimatter imbalance. The formation of a matter excess may be enhanced by CP violation in the Higgs sector, which may offer a second source of CP violation to complement that seen in other particle interactions.

The Standard Model predicts that the Higgs boson has a spin-parity ( $J^P$ ) of  $0^+$ . During Run-1 of the Large Hadron Collider (LHC), the first period of operation of the collider, different studies on the nature of the Higgs have led to the exclusion of the spin-parity states  $0^-$ ,  $1^\pm$  and  $2^\pm$  at more than 99% Confidence Level (CL). The studies conducted using data from A Toroidal LHC ApparatuS (ATLAS) and Compact Muon Solenoid (CMS) detectors with a centre-of-mass energy,  $\sqrt{s}$ , of 7 and 8 TeV focused on the decays of the Higgs in  $\gamma\gamma$ ,  $ZZ$ ,  $WW$  [36] [37] [38] [39]. Using the same data, a first test for CP-odd couplings was also performed in the Vector Boson Fusion (VBF) channel of the Higgs decay in  $\tau\tau$  channel [40].

During Run-2 data collection, with a  $\sqrt{s}$  of 13 TeV, the number of Higgs collected was an order of magnitude more than Run-1, and as a result CP-invariance constraints were tightened for the Higgs couplings to vector bosons and improved by studying the Higgs Yukawa couplings to fermions in the di- $\tau$  decays as well as Higgs boson top pair production ( $t\bar{t}H$ ) [41] [42] [43] [44].

The current studies on Higgs boson interactions provided by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) are too loosely constrained to provide a fine-grained picture of electroweak symmetry breaking.

The Higgs found within the Standard Model is even under a CP transformation. Therefore, it would be a clear indication of physics Beyond the Standard Model if there were a

measurable deviation from the pure CP-even interaction of the Higgs with any particle of the SM.

Analysing the detailed properties of the Higgs boson therefore remains at the heart of the LHC research programme.

### 1.3 Brout-Englert-Higgs mechanism

One of the key problems in modern particle physics is to explain the breaking of electroweak symmetries. The BEH is a simple mechanism that makes use of a complex iso-doublet scalar field, while accomodating all observed phenomena, to explain the symmetry breaking [13]. When interacting with itself, the scalar field acquires a non-zero vacuum expectation value which is responsible for the spontaneous electroweak symmetry breaking of  $SU(2)_L \times U(1)_Y$  down to  $U(1)_{EM}$ . The general expression of the scalar field doublet can be written as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (1.5)$$

with a weak hypercharge  $Y = \frac{1}{2}$  relative to  $U(1)_Y$ , where the Higgs doublet  $\phi$  consists of a complex charged component, described by  $\phi^+$  and a neutral complex component  $\phi^0$  that can be decomposed into its CP-even component  $\phi_3$  and a CP-odd component  $\phi_4$ . For such a scalar field, show in equation 1.5, the covariant derivative takes the form:

$$D_\mu \phi = \left( \frac{\partial}{\partial \mu} + \frac{i}{2} g \sigma^a W_\mu^a + \frac{i}{2} g' B_\mu \right) \phi \quad (1.6)$$

where  $\sigma^a$  (with a=1,2,3) is the Pauli matrices term,  $W^a$  represents the  $SU(2)_L$  gauge field while B represent the  $U(1)_Y$  one and g and g' are the respective gauge couplings.

With those considerations in mind, the scalar part of the Higgs Lagrangian can be expressed as:

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi) \quad (1.7)$$

where, in order to maintain the  $SU(2)_L \times U(1)_Y$  invariance, the  $V(\phi)$  term in equation 1.7 must take the form of the most general renormalisable scalar potential with constants  $\lambda$  and  $\mu^2$ . Those constants can be respectively associated to a self-coupling field term and a scalar boson mass term:

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.8)$$

The solutions of the potential in equation 1.8 are therefore dependent on the  $\lambda$  and  $\mu$  values and this leads to different scenarios:

- $\lambda < 0$  The potential is devoid of a stable minimum.

- $-\mu^2 \phi^\dagger \phi, \lambda > 0$  In this scenario the minimum is located at  $\phi = 0$  corresponding to the minimum energy and the symmetries of the Lagrangian are also respected.
- $-\mu^2 < 0, \lambda > 0$  This configuration of the potential has the minimum energy located at  $\phi$  values different from  $\phi = 0$ .

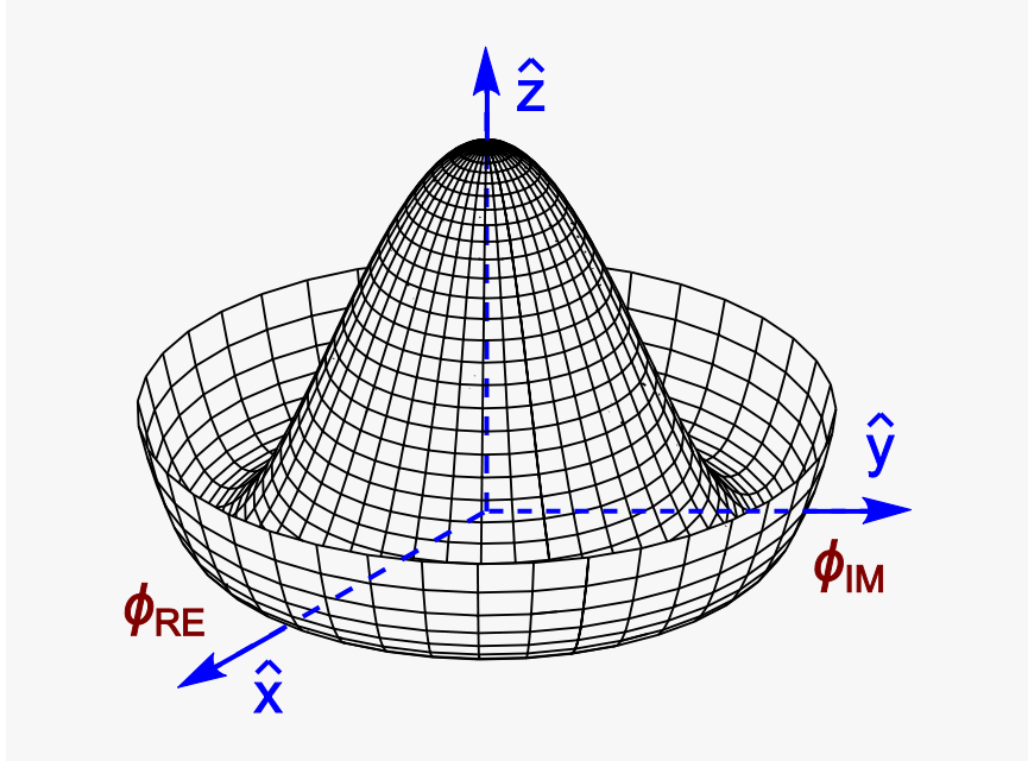


Figure 1.2: Representation of the scalar field potential  $V(\phi)$ , commonly known as the Mexican hat potential.

Under the conditions of  $-\mu^2 < 0, \lambda > 0$  the potential assumes the shape shown in figure 1.2 where the minimum energy of the Higgs Potential is not located at  $\phi = 0$ . The SM gauge symmetry is therefore spontaneously broken by the neutral component of the scalar field, in equation 1.5, which acquires a non-zero Vacuum expectation value (VEV)  $v$ . Such a potential can be satisfied by an infinite number of degenerate states but one arbitrary choice is to consider the ground state to be:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \text{with} \quad v = \sqrt{\frac{-\mu^2}{\lambda}} \quad (1.9)$$

By applying a small perturbation around the ground state by using a unitary gauge configuration and considering the neutral component of the scalar field in equation 1.5 as an expansion of the vacuum expectation value, the scalar doublet take the following form:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \quad (1.10)$$

The  $h$  term in the neutral component represents a real scalar field.

If we consider the Higgs potential in equation 1.8 with a scalar field as given by equation 1.10, the corresponding Lagrangian shown in equation 1.7 will acquire a term depending on  $\lambda v^2 h^2$  which represents the physical Higgs boson with a mass

$$m_\lambda = v\sqrt{2\lambda} \quad (1.11)$$

The derivative, shown in equation 1.6, allows the W and B gauge field associated with the  $SU(2)_L \otimes U(1)_Y$  symmetry to couple with the Higgs field, and the choice of the unitary gauge configuration allows the values of  $\phi_1 = \phi_2 = \phi_4$  to disappear from the scalar field of equation 1.5. Thus  $\phi_1, \phi_2$  and  $\phi_4$  represent three Goldstone degrees of freedom which become the longitudinal components of the W and Z physical gauge bosons. The mass of the W and Z gauge bosons represents the extra degree of freedom that comes out through the covariant derivative of the kinetic term in the Higgs Lagrangian, equation 1.7.

To demonstrate the generation of the gauge boson masses, the general scalar field  $\phi$  of the covariant derivative, equation 1.6, has to be replaced with the modified scalar field, equation 1.5:

$$D_\mu \phi = \left( \frac{\partial}{\partial \mu} + \frac{i}{2} g \sigma^a W_\mu^a + \frac{i}{2} g' B_\mu \right) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.12)$$

$$D_\mu \phi = \frac{v^2}{8} \begin{pmatrix} gW_\mu^1 - igW_\mu^2 \\ -gW_\mu^3 + g'B_\mu \end{pmatrix}. \quad (1.13)$$

By considering the scalar field in absence of perturbation (i.e. without any  $h$ -mixed terms), the kinetic term of the Lagrangian, equation 1.7, takes the form

$$(D^\mu \phi)^\dagger (D_\mu \phi) = \frac{v^2 g^2}{8} (W_1^\mu - iW_2^\mu)(W_\mu^1 + iW_\mu^2) + \frac{v^2}{8} (g'B^\mu - gW_3^\mu)(g'B_\mu - gW_\mu^3) \quad (1.14)$$

where this term of the Lagrangian density can be interpreted as new fields corresponding to massive particles. By defining the physical gauge fields as

$$W_\mu^+ = \frac{1}{2}(W_\mu^1 - iW_\mu^2) \quad (1.15)$$

and

$$W_\mu^- = \frac{1}{2}(W_\mu^1 + iW_\mu^2) \quad (1.16)$$

the Lagrangian term becomes

$$\frac{1}{2} \left( \frac{gv}{2} \right)^2 = W_\mu^\dagger W^\mu \quad (1.17)$$

and the W mass can be expressed as

$$m_W = \frac{gv}{2} \quad (1.18)$$

By using the same convention, the Z and A remaining gauge bosons can be defined as

$$Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 - g'B_\mu) \quad (1.19)$$

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(g'W_\mu^3 + gB_\mu) \quad (1.20)$$

with their respective mass

$$m_Z = \frac{v}{g}\sqrt{g^2 + g'^2}, \quad (1.21)$$

$$m_A = 0. \quad (1.22)$$

The  $Z_\mu$  represent the neutral boson of the weak force, while the  $A_\mu$  describe the massless vector field associated to the electromagnetic (EM) photon field. The nonzero vacuum expectation value yields the breaking scheme  $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{EM}$  which remain conserved in Electroweak symmetry breaking (EWSSB).

### 1.3.1 Lepton Higgs Coupling

The Yukawa coupling between the Higgs field and the Yukawa term produces an additional term interpreted as the fermion mass. The coupling ensures either the invariance of the symmetry  $SU(2)_L \otimes U(1)_Y$  or a single state with a coupling constant  $g_f$ .

$$\mathcal{L}_{\text{Yukawa}} = -g_f(\bar{\chi}_{L,f}\phi\psi_{R,f} + \bar{\psi}_{R,f}\bar{\phi}\chi_{L,f}). \quad (1.23)$$

where  $\chi$  and  $\psi$  represent, respectively, the left-handed fermions isospin doublet and the right-handed fermions singlet. By applying the scalar field of equation 1.5 in the Yukawa Lagrangian, equation 1.23, the lepton Yukawa Lagrangian will take the form:

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}}^{\text{lepton}} &= -\frac{g_l}{\sqrt{2}} \left( (\bar{\nu}, \bar{l})_L \begin{pmatrix} 0 \\ v+h \end{pmatrix} l_R + \bar{l}_R(0, v+h) \begin{pmatrix} v \\ l \end{pmatrix}_L \right) \\ &= -\frac{g_l(v+h)}{\sqrt{2}} (\bar{l}_L l_R + \bar{l}_R l_L) \\ &= \frac{g_l}{\sqrt{2}} v \bar{l} l - \frac{g_l}{\sqrt{2}} h \bar{l} l. \end{aligned} \quad (1.24)$$

The results shown in equation 1.24 prove that the masses of the leptons appear in the Lagrangian when considering the non-zero vacuum expectation values of the Higgs field:

$$m_l = \frac{g_l v}{\sqrt{2}}. \quad (1.25)$$

Even though the calculation for the quark masses is slightly more complicated it can be proven that the result is similar to the fermion case and the mass term appearing in equation 1.25 is valid for all fermions.

## 1.4 Higgs Production

A general description of the main processes involved in the Higgs production will be given in this section. The Standard Model Higgs, as demonstrated in section 1.3.1, can be coupled to either fermions, top and bottom quarks as well as  $\tau$  lepton, or weak bosons such as  $W^\pm$  and  $Z$ .

The Higgs boson is predominantly produced via gluon-gluon fusion. The second dominant process is the vector boson fusion, while the associated production mode, where the Higgs is produced via  $qq \rightarrow HW, HZ, t\bar{t}H, b\bar{b}H$ , has smaller cross sections.

Table 1.3 shows the differential cross section values for the most important Higgs production mechanisms available at the LHC with a center-of-mass energy of  $\sqrt{s} = 13$  TeV.

### 1.4.1 Gluon-gluon fusion

The dominant Higgs production mechanism at the LHC involves gluon fusion via an intermediate top-quark loop:

$$gg \rightarrow X + H \quad (ggF) \tag{1.26}$$

The leading order (LO) process proceeds through a top quark loop between the gluons and the Higgs particle. Although in principle all quarks should be included in the loop, in practice the restriction to just the top quark suffices because the Higgs couples about 35 times more strongly to the top than to the next-heaviest fermion, the bottom quark, leading to a relative suppression of the bottom contribution by a factor  $35^2$ . The calculation of the matrix element in lowest order is finite, even though the diagrams feature a loop, which is a common source of infinities. The reason for that is that there is no fundamental  $ggH$  coupling in the Standard Model which could absorb such infinities. To calculate the cross section one has to consider either the Parton Distribution Function (PDF) or the next-to-leading order (NLO) corrections in  $\alpha_s$  [46, 47].

The cross section measurements can be improved by considering also the emission of soft, virtual and collinear gluons which contribute to the soft limit where the relative center of mass edges towards the Higgs mass. Other corrections involving soft-gluon radiation have been done at next-to-next-leading (NNLO) order and NNNLO order [48, 49].

To evaluate the cross section of the gluon-gluon fusion Higgs production mode, the leading top quark has to be considered in the limit of the top mass approaching to infinity

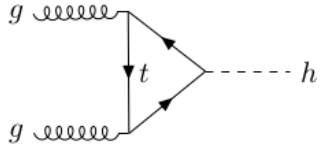
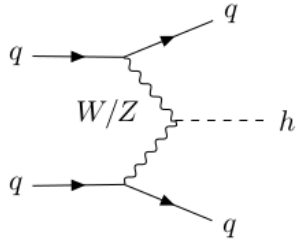
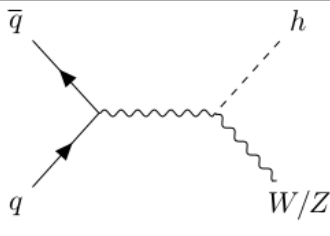
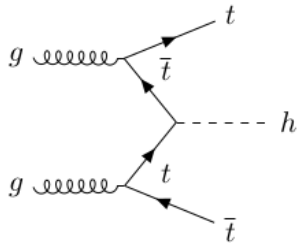
Production Mechanism	Cross-section	Diagram
Gluon-gluon fusion	43.92 pb	
Vector boson fusion	3.748 pb	
V-associated	$Wh$ : 1.380 pb $Zh$ : 0.8696 pb	
$t\bar{t}$ -associated	0.5085 pb	

Table 1.3: The main mechanisms for Standard Model Higgs boson production at  $\sqrt{s} = 13$  TeV, for a Higgs mass  $m_h = 125.0$  GeV, and the corresponding cross sections. [45].

$m_t \rightarrow \infty$ . In this approximation the  $ggH$  coupling becomes point-like; the corresponding Feynman diagram results with one less loop to calculate and the approximation at higher order can be computed for N3LO [50], NNLO [51, 51, 52], NLO [53]. The  $ggF$  cross section value shown in table 1.3 is calculated combining the N3LO corrections with the soft-gluon resummation at different orders of accuracy [54, 55].

### 1.4.2 Vector boson fusion

Corresponding to the second largest contributor to the Higgs boson production in the SM, the VBF is important for its power to discriminate signal from background in the proton-proton (pp) collision, due to its peculiar final state. The VBF channel is therefore of great interest even if its cross-section is an order of magnitude lower than  $ggF$ 's, the main Higgs production mode.

The production of the Higgs comes through the scattering of two quarks that interact via



the exchange of a W or Z boson:

$$q + q \rightarrow q + q + H \quad (VBF) \quad (1.27)$$

where the Higgs is radiated from the weak propagator.

The two quarks in the final state produce hadronic jets which are detected at ATLAS. Since the momentum transfer in the weak propagator which produces the Higgs Boson is  $\ell^+\ell^-\sqrt{s}$ , the two hadronic jets are at high pseudo-rapidity  $\eta$ , defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \quad (1.28)$$

where  $\theta$  is the angle between the jet direction and the beam axis. In most cases of parton-parton scattering a coloured object is exchanged with an associated colour flow between the outgoing partons and the remnant particles. This can be modelled by a colour string which is stretched across the central rapidity interval. This string then fragments into particles which occupy the region between the two jets.

However, it is also possible for the exchanged object to be a colour singlet. In this case the colour strings connect each outgoing parton with the remnant jet closest to it in rapidity. This leads to a suppression of particle production in the rapidity region between the two jets.

Higgs is colour neutral and if it decays into quarks, the decay products must be colour neutral as well, so the colour flow is between jets coming from the Higgs. Thus there is no colour connection between the forward VBF jets and the Higgs related jets. Gluon emission is typically soft and in the direction of the jets (collinear), hence the rapidity gap. The cross section is computed with strong and electroweak corrections to an accuracy of NLO state [56, 57]. Additional NNLO corrections are also applied [58].

### 1.4.3 Vector boson associated production $Vh$

The process involved in the production of the Higgs boson with a weak vector boson increases the cross section, compared to the VBF channel, when a Z boson is involved in the process. If the Z boson is not present, the  $Vh$  process, also known as *Higgs-strahlung*, is strongly suppressed by a quark-quark subprocess corresponding to  $q\bar{q}Vh$ .

The cross section is computed with QCD correction up to NNLO order [59].

### $t\bar{t}$ associated production

Even though the Higgs production in association with a  $t\bar{t}$  pair occurs with a very small cross section, it is very important for the Higgs sector for its resolution power of the Yukawa coupling between the top quark and the Higgs, a feature that allows this process to be unique in this regard. NLO corrections are shown in Refs. [60, 61, 62, 63].

## 1.5 Higgs Boson Decay

The boson known as the Standard Model Higgs Boson has a tiny lifetime equal to  $1.56 \times 10^{-22}$  and is not directly observable. What we actually measure in the detector are the decay products that can be summarised in three categories, as shown in figure 1.3.

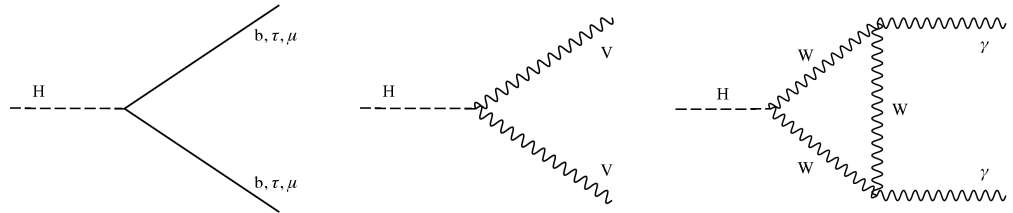


Figure 1.3: Leading order Feynman diagram for Higgs Boson decay to two fermions (left), two bosons (center) and two  $\gamma$  (right).

For the Higgs boson discovered at CERN, with a mass equal to 125 GeV, the dominant decays correspond to the production of two fermions through the Yukawa interaction as well as the production of a pair of W or Z bosons. Fig. 1.3 shows the leading order Feynman diagrams for those processes but also the Higgs to  $\gamma\gamma$  diagram where the decay products are obtained through a fermion loop similarly to the gluon pair decay mode.

Fig. 1.4 shows the Higgs decay branching ratios as a function of the Higgs mass. As it is clear from the figure, the main decay mode for the SM Higgs is into a pair of b-quarks due to their heaviest mass that can be produced in the collider.

It is also worth to mention the Higgs decay mode in  $\gamma\gamma$  mainly for two reasons. Despite having a smaller branching ratio compared to the fermions and bosons decays, it has a very clean signature, due to the two very energetic photons that are produced, but also the di-photon decay mode was one of the first decay channels in which the Higgs boson was discovered by both ATLAS [64] and CMS [65]. The other discovery mode was Higgs to a pair of Z bosons where both particles decay to a pair of leptons ( $\mu^+\mu^-$ ) each, providing a very clean signature to select, as well as the decay of the Higgs in  $W^+W^- \rightarrow e^+\nu_e e^-\bar{\nu}_e$ .

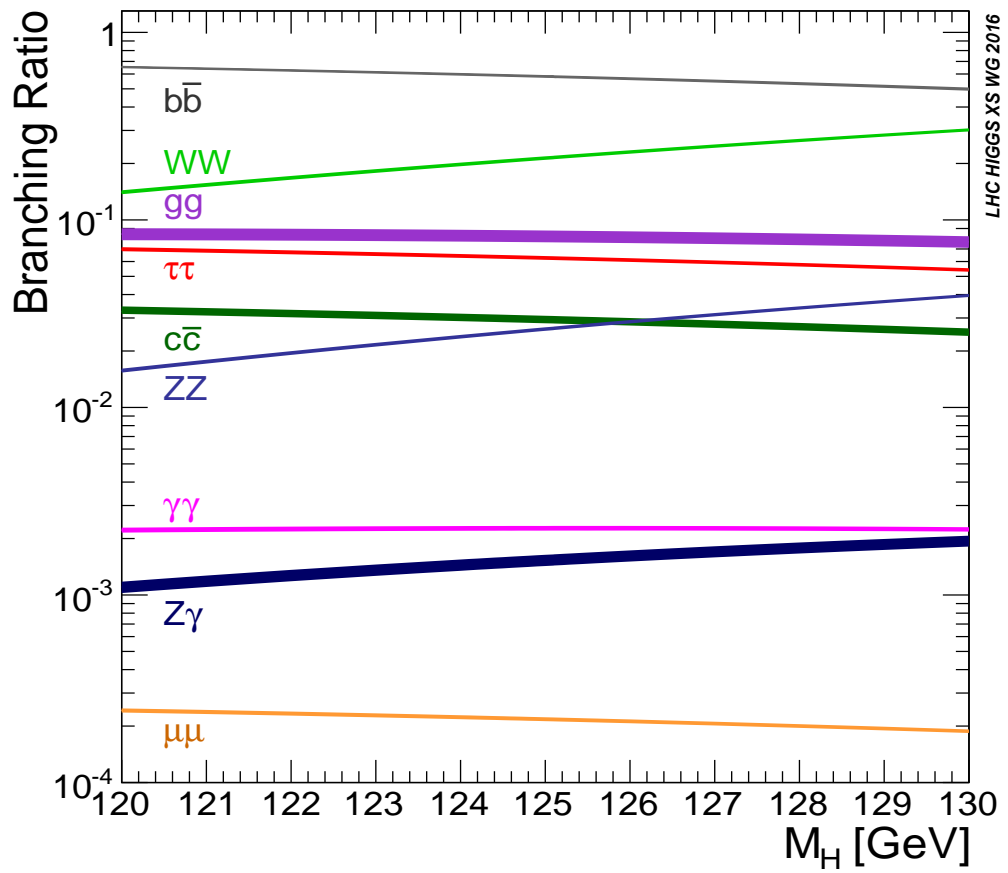


Figure 1.4: Standard Model Higgs decay branching ratio as a function of the Higgs mass.

# THE ATLAS EXPERIMENT AT THE LHC

2

**ATLAS** (A Toroidal LHC ApparatuS) is one of the four main LHC experiments, the others being CMS, A Large Ion Collider Experiment (ALICE) and LHCb, taking data at a centre-of-mass energy of 13 TeV using beams delivered by the LHC. In this chapter an overview of the LHC will be given in Section 2.1, then the ATLAS detector will be described in Section 2.2, and finally the Trigger system, used to select the data, will be described in Section 2.3.

## 2.1 The Large Hadron Collider

As of today, the LHC [66] is the world's largest and most powerful particle accelerator. It was designed to explore some of the fundamental open questions in particle physics by colliding protons at energy and luminosity never explored before. It is located at CERN, in the Geneva area, at a depth ranging from 50 meters up to 175 meters underground. It consists of a 27-kilometers ring made of superconducting magnets divided in two separate beam pipes to let the particle travel in opposite directions.

The beams are guided around the ring by a strong magnetic field generated by coils - made of special electric cables - that can operate in a superconducting regime. More than thousand superconducting dipole and more than three hundred quadrupole magnets, with an average magnetic field of 8.3 Tesla, are employed and kept at a temperature below 1.7 Kelvin, in order to preserve their superconducting properties.

The former are used to bend the beams and the latter to keep them focused while they get accelerated. The beams are accelerated by radiofrequency (RF) cavities. An RF cavity is a metallic chamber that contains an electromagnetic field. Charged particles that pass through the cavity are affected by the electromagnetic field, which transfers energy pushing them forwards along the beam line. To achieve the confinement of beams in "bunches",

protons are forced to stay close in energy by accelerating or decelerating them accordingly to the energy they have when arriving.

The collider first went live on September 2008 but, due to a magnet quench incident, it has been fully operational since November 2009, when low-energy beams circulated in the tunnel for the first time since the incident. This also marked the start of the main research programme and the beginning of the so-called Run 1: first operational run (2009 - 2013) at 7/8 TeV centre-of-mass energy.

## Performance of the LHC

After a long shutdown In June 2015 the LHC restarted delivering physics data, the so-called Run 2 - second operational run (2015 – 2018) - during which the LHC collided up to  $10^{11}$  bunches of protons every 25ns at 13TeV and the design luminosity, the highest luminosity the detector was designed to cope with, of  $2 \cdot 10^{34} \text{cm}^{-2} \text{s}^{-1}$ . The definition of the luminosity is [67]:

$$\mathcal{L} = f \frac{n_b N_1 N_2}{4\pi\sigma_x\sigma_y} \quad (2.1)$$

where  $n_b$  is the number of bunches,  $N_1$  and  $N_2$  represent the numbers of protons per bunch in each of the colliding beams,  $f$  is the revolution frequency of the bunches, and  $4\pi\sigma_x\sigma_y$  is the transverse area of the bunches at the interaction point, described by the Gaussian widths  $\sigma_x$  and  $\sigma_y$  (horizontal and vertical dimensions of the beam).

The relation between the number of collision during a specific experiment and the luminosity is given by the following expression:

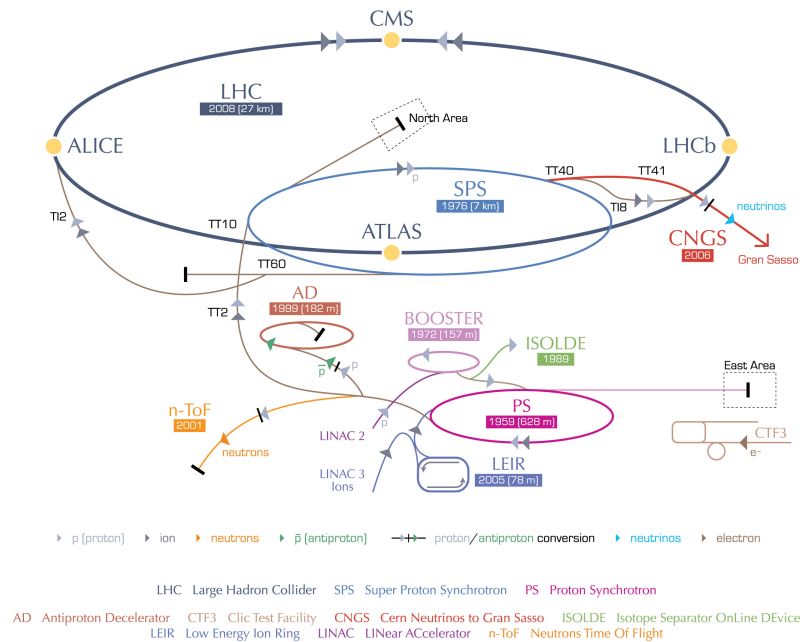
$$\mathcal{N}_{\text{event}} = \mathcal{L}\sigma_{\text{event}} \quad (2.2)$$

where  $\sigma_{\text{event}}$  is the cross section of the process under investigation. The LHC has not only collided protons but also heavy ions, in particular lead nuclei at  $\sqrt{s_{NN}} = 5.02$  TeV, at a luminosity of  $10^{27} \text{cm}^{-2} \text{s}^{-1}$  [68].

## Acceleration stages

The acceleration stage occurs at different stages where the proton beams are slightly accelerated each time before reaching the maximum energy. Figure 2.1 shows a sketch of the CERN's accelerator complex. The first stage take place at Linear Accelerator 2 (LINAC2). Here protons are accelerated up to 50 MeV, and then injected in the PSB where they reach 1.4 GeV. The Proton Synchrotron corresponds to the second stage, where the beams are boosted up to 25 GeV and then Super Proton Synchrotron (SPS) makes them reach energies up to 450 GeV. Eventually, the beams are injected in bunches with a 25 ns spacing

## CERN's accelerator complex



European Organization for Nuclear Research | Organisation européenne pour la recherche nucléaire

© CERN 2008

Figure 2.1: CERN Accelerator complex. The LHC is the last ring (dark gray line). Smaller machines are used for early-stage acceleration and also to provide beams for other experiments [1].

into the LHC, where they travel in opposite directions, while they are accelerated up to a centre-of-mass energy of 13 TeV. Once the bunches reach the maximum energy, they are made collide at four different points, corresponding to the origin point of the collision, inside four experiments around the ring [66].

The four large detectors at the collision points are the multi-purpose detectors ATLAS [3] and CMS [69], the LHCb detector [70], which focuses on flavour physics, and ALICE [71] which specialises in heavy ion physics. The *big four* are not the only experiments at the CERN's accelerator complex. There are also smaller experiments based at the four caverns about the collision points e.g. TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC (TOTEM) [72], Large Hadron Collider forward (LHCf) [73] and Monopole & Exotics Detector At the LHC (MoEDAL) [74], but these will not be discussed any further.

## 2.2 The ATLAS Detector

ATLAS is a general-purpose detector designed to collect data with the highest luminosity provided by the LHC. It measures about 45 m in length and 25 m in diameter. It has a

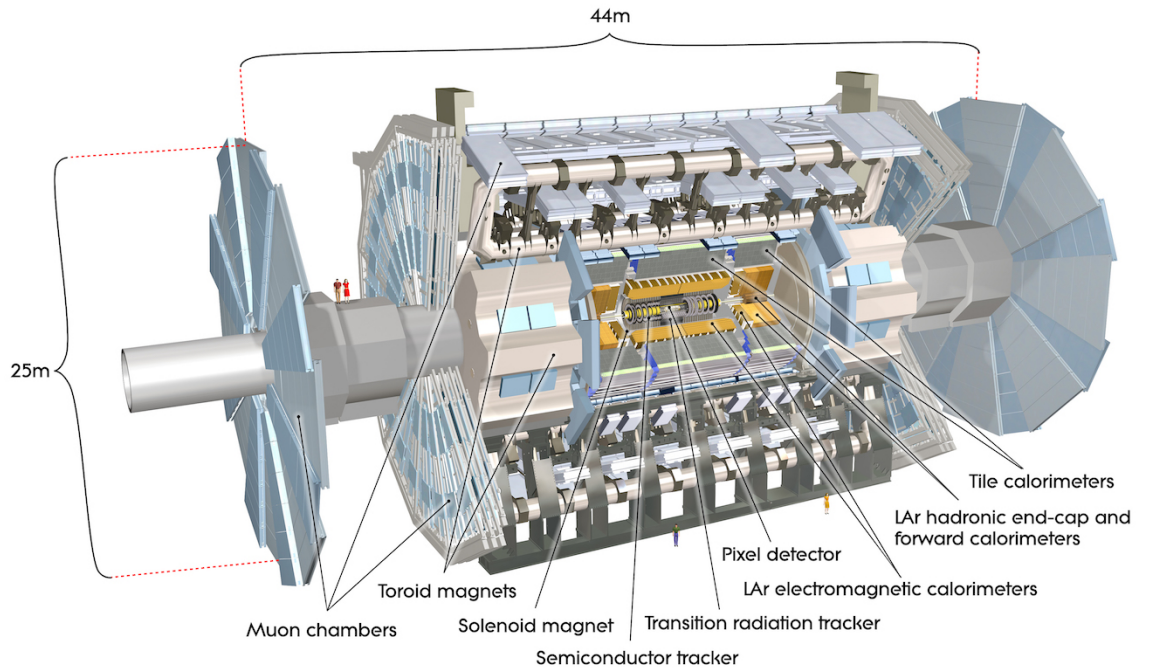


Figure 2.2: Cut-away view of the ATLAS detector. The dimensions of the detector are 25 m in height and 44 m in length. The overall weight of the detector is approximately 7000 tonnes [1].

forward-backward symmetric cylindrical geometry with respect to the interaction point and is designed to reconstruct and measure physics objects such as electrons, muons, photons hadrons and hadronic jets. Its design was optimised to be as sensitive as possible to the discovery of the Higgs boson and Beyond Standard Model (BSM) physics. For this reason, the sum of all the main detectors with the help of several sub-systems give the possibility to observe all decay products by covering almost  $4\pi$  steradians of solid angle.

In Figure 2.2 a cut-away view of ATLAS with all its components is shown. The innermost layer is the Inner Detector (ID) which is the core of the tracking system and consists of a Pixel, a SemiConductor Tracker (SCT), and a Transition Radiation Tracker (TRT) detectors. It is submerged in a 2 T magnetic field, generated by a thin superconducting solenoid, which bends all the charged particle's trajectories allowing transverse momentum measurement. The electromagnetic and hadronic calorimeters form the next layer and they are both used to perform precise energy measurements of photons, electrons, and hadronic jets. Located on both sides of the detector, finally, there are the two muon tracking chambers which perform measurements of the muon momenta.

## The ATLAS coordinate system

ATLAS uses a right-handed coordinate system with its origin at the nominal Impact Parameter (IP) in the centre of the detector and the  $z$ -axis along the beam pipe. The  $x$ -axis points from the IP to the centre of the LHC ring, and the  $y$ -axis points upward. Cylindrical coordinates  $(r, \phi)$  are used in the transverse plane,  $\phi$  being the azimuthal angle around the beam pipe.

The pseudorapidity is defined in terms of the polar angle  $\theta$  as:

$$\eta \equiv -\ln\left(\tan(\theta/2)\right) \quad (2.3)$$

Rapidity is defined as:

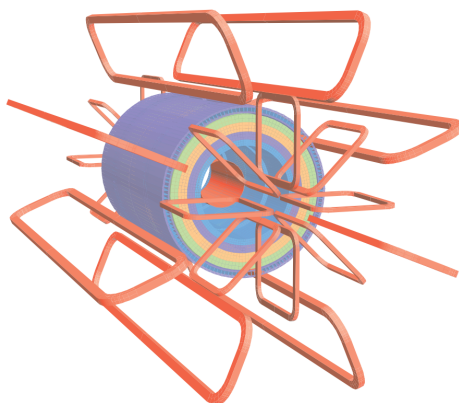
$$y = 0.5 \ln \frac{E + p_z}{E - p_z} \quad (2.4)$$

where  $E$  denotes the energy and  $p_z$  is the component of the momentum along the beam direction. In the  $(\eta, \phi)$  space a distance  $\Delta R$  can be therefore defined as:

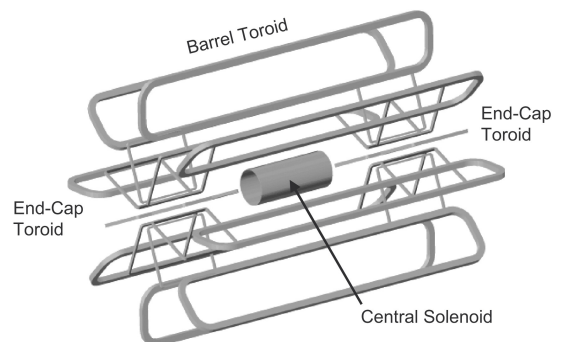
$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (2.5)$$

where  $\Delta\eta$  and  $\Delta\phi$  are the differences in pseudo-rapidity and azimuthal angle between any two objects. By considering  $\Delta\eta$  and  $\Delta\phi$  different regions can be defined such as a central region and a forward region.

### 2.2.1 The Magnet System



a Geometry of magnet windings and tile calorimeter steel. The eight barrel toroid coils, with the end-cap coils interleaved are visible. The solenoid winding lies inside the calorimeter volume [3].



b Schematic view of the superconducting magnets [75].

Figure 2.3: The ATLAS magnet system.



The ATLAS magnet system, with a length of 26 m and a diameter of 22 m, is of fundamental importance to bend the trajectories of charged particle and therefore to measure the relative momentum. Figures 2.3a and 2.3b show the geometry of the system and its components, which are made of Alloy of niobium and titanium (NbTi) - superconducting material - and will be described in the following paragraphs.

### **The Central Solenoid**

The dimensions of the central solenoid correspond to an axial length of 5.8 m, an inner radius of 2.46 m, and an outer radius of 2.56 m. It is located between the ID and the Electromagnetic Calorimeter (ECAL) and its main purpose is to bend the charged particles that go through the ID. It is aligned to the beam axis providing a 2 T axial magnetic field that allows accurate momentum measurement up to 100 GeV [75].

### **The Barrel and the End-cap Toroids**

Figure 2.3b displays the toroid magnetic system that surrounds the calorimeters. With its cylindrical shape this component consists of a barrel and two end-caps toroids. The barrel toroid is made of eight coils and produces an approximately 0.5 T toroidal magnetic field for the central muon detectors.

The end-cap toroids, made of eight coils each, produce an approximately 1 T toroidal magnetic field, able to improve bending power for the muon spectrometer in the end-cap regions.

#### **2.2.2 The Inner Detector**

The innermost component of the ATLAS detector is the ID [76], used to reconstruct the track momenta and charges for a wide range of particle momenta (from 100MeV to several TeV) and within  $\eta$  within  $|\eta| < 2.5$ . The tracks are then used to find primary and secondary vertices with the best possible resolution.

Figure 2.4 shows an overview of the ATLAS ID while in figure 2.5 is summarized the ID inclusive of independent sub-systems, the Insertable B-Layer (IBL), the Pixel detectors, the SCT tracker and the TRT tracker.

These sub-detectors will be discussed in the following paragraphs.

#### **Insertable B-Layer IBL**

The IBL [77] is the innermost Pixel Detector layer as shown in Figure 2.5. It was added during ATLAS Run 2 upgrade (2013/2014), the innermost Pixel Detector layer improves

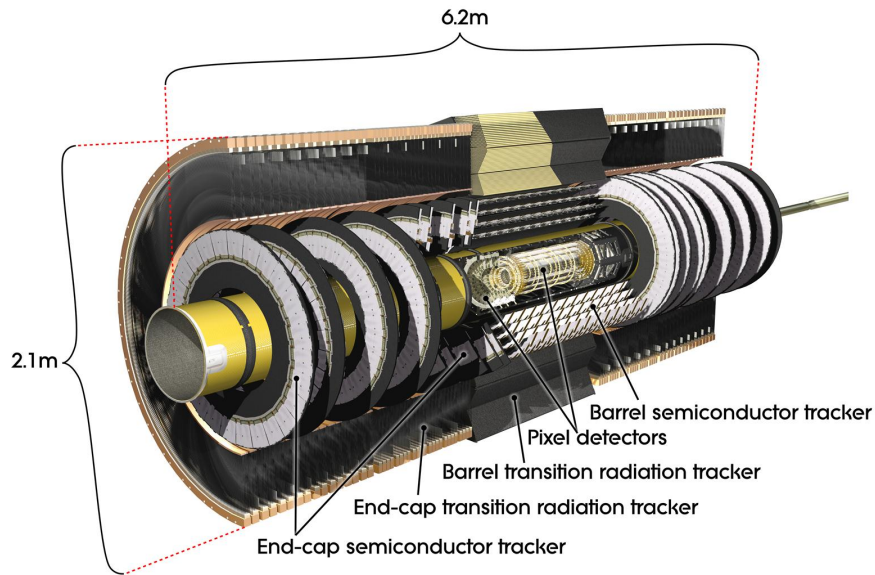


Figure 2.4: Overview of the ATLAS ID with labels and dimensions. [2].

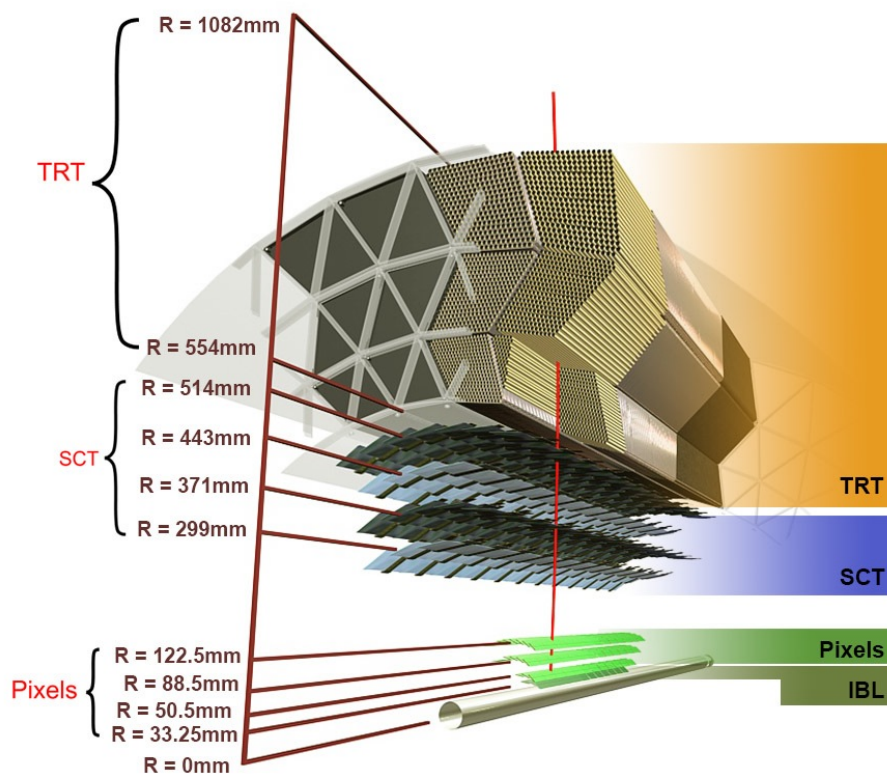


Figure 2.5: Diagram of the ATLAS ID and its sub-detectors.

the vertex reconstruction by a factor  $\sim 1.4$  as well as tracking precision and the impact parameter resolution by a factor of 2. It is comprised of 6M channels and each pixel measures  $50 \times 250 \mu\text{m}$ . Its resolution is  $8 \times 40 \mu\text{m}$ . The IBL improves the impact parameter reconstruction of tracks by a factor 2, as well as improving the Primary Vertex (PV) reconstruction by a factor 1.4, of primary importance e. g. for several aspect of the physics

object reconstruction.

### Pixel detector

The Pixel detector is made of 1750 identical sensorchip-hybrid modules, and each of them can cover an active area of  $16.4 \times 60.8$  mm. The total number of modules correspond to roughly 80 million semiconductor silicon pixels. Each pixel covers an area of  $50 \mu\text{m}$  in the  $\phi$  direction and  $400 \mu\text{m}$  along the  $z$ -axis (beam axis) [78]. The silicon pixel detector measures 48.4 cm in diameter and 6.2 m in length providing a pseudo-rapidity coverage of  $|\eta| < 2.5$ . Three concentric barrel layers placed at 50.5 mm, 88.5 mm and 122.5 mm respectively are shown in Figure 2.5. The Pixel detector also has six disk layers, three for each forward region, which provide enough precision hits for tracks with  $|\eta| > 2.0$ . The fine granularity of this detector allows accurate measurement and precise vertex reconstruction. It has a resolution of  $10 \times 115 \mu\text{m}$ .

### The silicon microstrip detector system SCT

The SCT comprised of silicon micro-strip layers. Thanks to its resolution ( $17 \times 580 \mu\text{m}$ ) it can accurately measure particle momenta. It is made of 4088 modules of silicon micro-strip detectors arranged in four concentric barrel layers form the SCT detector. Similarly to the ID, it is mainly used for precise momentum reconstruction within the range  $|\eta| < 2.5$ . The structure of the SCT is shown in figure 2.5, made of four concentric barrel layers and two end-cap layers. The resolution for each module is  $17 \mu\text{m}$  in the  $R - \phi$  direction and  $580 \mu\text{m}$  in the  $z$  direction. The fact that the SCT covers a larger area and with more measurements means that the SCT with less granularity maintain the same level of performance of the Pixel detector: SCT uses  $\sim 6.3$  million read-out channels.

### The Transition Radiation Tracker TRT

The TRT, as shown in figure 2.5, is the last and outermost sub-system of the Inner Detector. It consists of straw tubes (4 mm diameter) wound from a multilayer film reinforced with carbon fibers and a  $30 \mu\text{m}$  gold plated tungsten wire is located in its centre. The straw, full of a gas mixture of 70% Xe, 27% CO<sub>2</sub> and 3% O<sub>2</sub> [79], contains three concentric layers with different radii, each of which has about 50 thousands of straws located in 32 modules.

When a charged particle passes through the gas, ionisations occur allowing the electrons to be collected at the anode and the ions at the cathode, and therefore, using the time it takes for electrons to drift to the wire, we can calculate the distance from the wire tube and the resulting position.

The TRT provides up to 36 point per track - the large the number and the larger average radius compensate for the lower resolution ( $170\mu m$  per straw).

Another important aspect of the TRT is that it is capable detecting the transition radiation photons emitted when a relativistic particle passes through the edge of two media with different dielectric constants. The amount of radiation is proportional to the specific particle and therefore thresholds are set to classify different particles.

With a resolution of  $130\mu m$ , on average almost 35 hits are observed when a particle crosses the gas mixture.

## Performance of the ID

The reconstruction of the tracks is of fundamental importance to measure well the properties of the physical objects expected in the final state of this analysis such as  $\pi$  and leptons. Derived from the contribution of the three ID sub-systems, the overall performance expressed in terms of momentum resolution was calculated using cosmic muons [80]:

$$\frac{\sigma_{p_T}}{p_T} = 1.6 \pm 0.1\% \oplus \frac{(53 \pm 2) \cdot 10^{-5}}{\text{GeV}} \times p_T \quad (2.6)$$

### 2.2.3 The Calorimeters

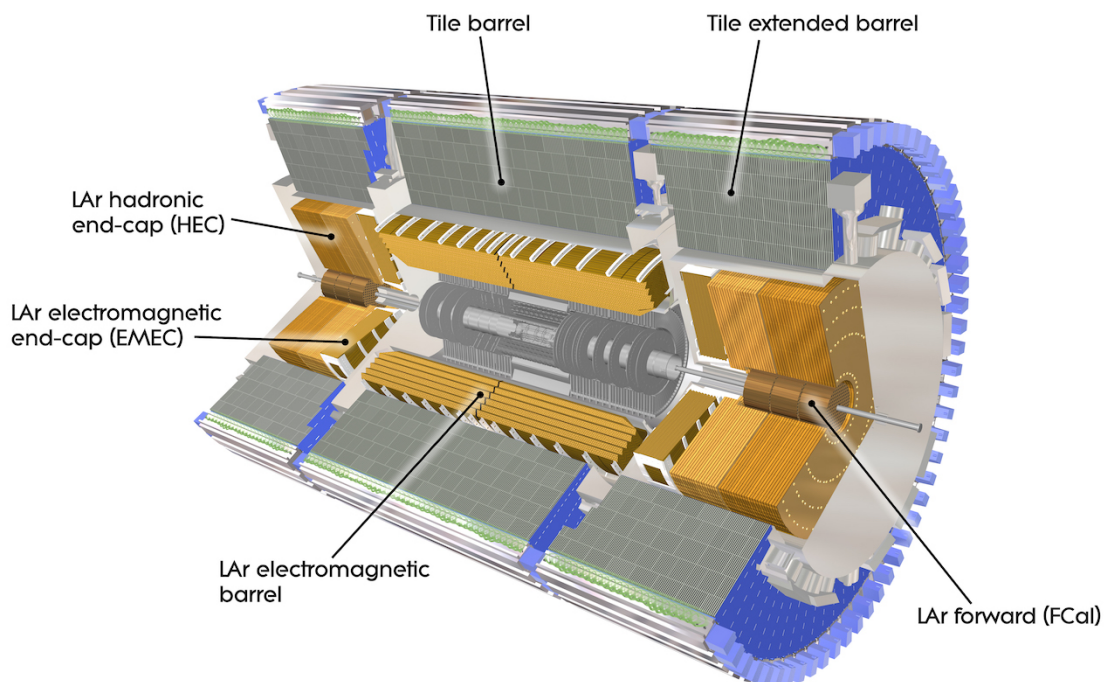


Figure 2.6: A computer generated image of the full calorimeter.

The ATLAS Calorimeter system, shown in Figure 2.6, is comprised of two main sub-systems. The ECAL and the Hadronic Calorimeter (HCAL) are the places where particles that interact via the electromagnetic force and hadronic particles respectively are measured and identified. A cover angle in  $\phi$  and  $|\eta| < 4.95$  is reached when combining both detectors. ECAL is comprised of one barrel and two end-cap sectors. When particles cross the layers of the ECAL they slow down and lose energy while generating showers as electron pairs that are collected to measure the energy. The pair production process occurs above a few MeV regime, below it photoelectric and Compton scattering are the most dominant effects. Photons are also emitted via bremsstrahlung by electrons and positrons, and this effect, together with the pair production is maintained until the energy of the photon is lower than the pair production threshold.

The choice to design the ECAL as multiple layers of Liquid Argon (LAr) samplers and lead absorbers brings two advantages, a full  $\phi$  coverage with no non-interactive regions and a fast extraction of the electric signals from the readout. ECAL is made of two half-barrel wheels providing a pseudo-rapidity coverage up to  $|\eta| < 1.475$  and two end-cap providing  $1.375 \leq |\eta| \leq 3.20$  coverage. The barrel is divided in three layers with different granularities. Layer 1, used to separate  $\gamma$  and  $\tau$ , is segmented into strips of  $\Delta\phi \times \Delta\eta = 0.0031 \times 0.098$ . Layer 2, that correspond to the biggest layer of the calorimeter and therefore the responsible of the energy collection, has a granularity of  $\Delta\phi \times \Delta\eta = 0.025 \times 0.0245$ . The third layer helps to collect shower tail with a granularity  $\Delta\phi \times \Delta\eta = 0.0031 \times 0.098$ .

The total thickness of the ECAL is 22 radiation length  $X_0$  in the barrel and 24 radiation length  $X_0$  in the end-caps. Similarly to ECAL, HCAL is comprised of one barrel and two end-cap sectors. The barrel is made of tiles of scintillating material (Tile Calorimeter) whereas the Liquid Argon Calorimeter (LAr) is employed as an end-cap calorimeter. For best performance, all hadronic particles interacting with the detectors have to deposit their whole energy within the calorimeter.

Measuring the energy of the hadronic showers is the main purpose of HCAL, which is built using steel and scintillating tiles [81]. It is made up of three barrels, a central one that covers the region  $|\eta| < 1.0$ , and two extended barrels that cover the region  $0.8 < |\eta| < 1.7$ . Each barrel is made up of 64 modules and each module is in turn made up of three layers. Ultimately, the smallest section of the calorimeter module is a cell with a  $\Delta\phi \times \Delta\eta = 0.1 \times 0.1$  granularity for the two innermost layers and  $\Delta\phi \times \Delta\eta = 0.2 \times 0.1$  for the outermost one.

An additional  $3.1 < |\eta| < 4.9$  coverage is provided by the LAr forward calorimeter (FCal). It uses LAr as active material and copper as absorber [81].

## Performance of the Calorimeter

The performance of the calorimeter is important to measure the properties of the  $\tau$  leptons used in the analyses presented in this thesis. Experimental measurements have been performed using test beam data [82]. The energy resolution is then fitted using equation 2.7

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E[\text{GeV}]}} \oplus b, \quad (2.7)$$

where  $a$  is the stochastic term and  $b$  is a constant that includes local non-uniformities in the calorimeter response. The fitted energy resolution for ECAL is

$$\frac{\sigma(E)}{E} = \frac{(10 \pm 0.4)\%}{\sqrt{E}} \oplus (0.4 \pm 0.1)\% \quad (2.8)$$

with a variation of no more than 0.7% for the entire coverage of the calorimeter. The measurement was performed firing an electron beam at a module that is identical to those in ATLAS.

The fitted energy resolution for HCAL is shown in equation 2.9.

$$\frac{\sigma(E)}{E} = \frac{(52 \pm 1.0)\%}{\sqrt{E}} \oplus (3.0 \pm 0.1)\% \oplus (1.6 \pm 0.1)/E, \quad (2.9)$$

where the additional term is added to account for electronic noise.

### 2.2.4 The Muon Spectrometer

The outermost sub-system of the whole ATLAS detector is the Muon Spectrometer (MS) [83], shown in Figure 2.7. The main purpose of the MS is to perform precision measurement of muon momenta. Such a precision measurement is performed with big superconducting toroid magnets that bend the muon tracks as well as high-precision tracking chambers.

One large barrel toroid and two end-cap toroids made up the MS, covering the  $|\eta| \leq 1.4$  and  $1.6 < |\eta| \leq 2.7$  regions respectively, and both together act over charged muons by deflecting the track accordingly to the magnetic field, which has a magnitude ranging between 0.5 to 2T generated by eight large superconducting coils. To measure tracks, three cylindrical layers are arranged parallel to the beam axis in the barrel. In the transition and endcap regions three chambers are placed in planes perpendicular to the beam.

The Monitored Drift Tube (MDT), a set of Aluminium tubes containing a *W-Re* (Tungsten-Rhenium) wire fully covered by non-flammable Ar-CH<sub>4</sub>-N<sub>2</sub> mixture, are used to measure precisely the track coordinates in the bended direction. They cover most of the pseudo-rapidity range and the particle position resolution is 80  $\mu\text{m}$ .

The Cathode Strip Chamber (CSC), covering a pseudo-rapidity range of  $2 < |\eta| < 2.7$ , is similar to the MDT with the only difference of having a cathode strips rather than tubes

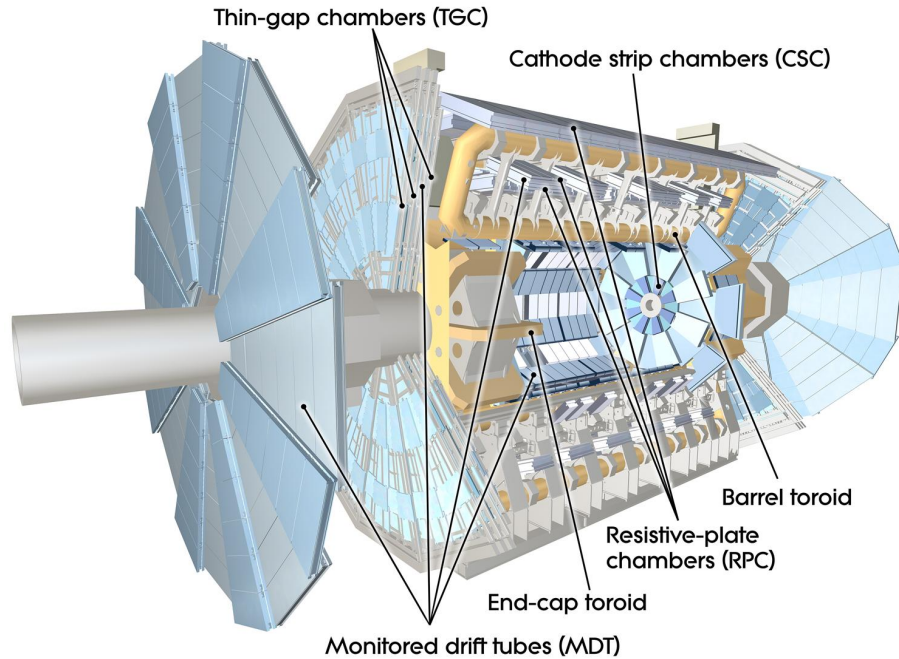


Figure 2.7: Cut-away view of the ATLAS muon system [3].

placed above and below the anode wires. The cathode wires are placed both orthogonal and parallel to the wires enabling measurement of the transverse coordinate. CSC is filled with non-flammable gas mixture of Ar-CO<sub>2</sub>-CF<sub>2</sub>.

Similarly to the CSV, the Thin-Gap Chamber (TGC) is a technology which is used in the end-cap region of the MS for triggering muons by providing large signals in a very tiny time window.

Lastly, the Resistive-Plate Chamber (RPC), covering the pseudo-rapidity in the range  $|\eta| < 2.7$  provide a measurement of the track coordinate in the non-bending plane. They are also gas-based detectors, made of two resistive parallel plates. It makes use of an electric field to generate avalanche multiplication having as a center the first ionisation electron. A set of Aluminium strips detect the avalanche signal.

## 2.3 The ATLAS Trigger System

The core of data taking is the ATLAS Trigger System, a crucial component of any high energy physics experiment able to decide whether or not to store an event. The principal function is to reduce the event rate from  $\sim 40$  MHz bunch-crossing to  $\sim 1$  kHz.

The Trigger System is made of a software trigger and a hardware trigger, the High Level Trigger (HLT) and Level-1 (L1) respectively. L1 is capable do analyzing information from the Muon Spectrometer and from the calorimeter to determine the Regions of Interest, an  $\eta - \phi$  region where the event are located. L1 tau trigger is entirely based on calorimeter

---

information. Topological clusters are calibrated locally at cell level and added up vectorially to form a proto-jet. This allows to set a  $p_T$  requirement of 20 or 25 GeV and isolation criteria on single taus. Also two isolated taus with a  $p_T$  of at least 12 GeV are accepted. Thresholds depend on instantaneous luminosity conditions to limit the rate into the trigger system. For high luminosities a jet with  $p_T > 25$ , GeV and  $\eta < 3.2$  (L1 J25) has been required.

Additional inspection is performed with the HLT software running on a computer cluster known as the HLT farm.

A further investigation is performed by the Fast TracKer (FTK) system [84], that use the information from the inner detector to improve the selection. It greatly enhances the identification of particles with improved energy deposition and track reconstruction.



# OBJECT RECONSTRUCTION

## 3

This chapter describes how the ATLAS objects used in the analysis are reconstructed, with a particular emphasis on primary vertices, important for the construction of the main observable  $\varphi_{CP}^*$ , as well as either charged pions or neutral pions, of primary importance for the decay plane reconstruction. Section 3.2.3 will describe in detail the hadronic  $\tau$  decays, used to reconstruct the mass of di-tau system  $m_{\tau\tau}^{MMC}$  and sections 3.2.4 and 3.2.5 will show, respectively, the jets and missing transverse energy reconstruction, both used to select the events as defined in table 6.3 and table 6.4 for the semi-leptonic and fully-hadronic case, respectively.

Moreover, it is of fundamental importance to reconstruct the 4-vector of each particle present in the individual decay of the  $\tau$ 's to construct the observable  $\varphi_{CP}^*$  through the  $\tau_{had-vis}$ , a visible hadronic decay of the  $\tau$ , to get the highest possible sensitivity.

### 3.1 Primary vertex and track parameters

The detector responsible for the reconstruction of charged particles is the ID. When a proton-proton collision occurs at LHC, each charged particle deposits hits in the Inner Detector and this is used to reconstruct charged tracks. The tracks are then used in the vertex fit to find the point where they originate, the initial point, corresponding to the vertex of the collision. For each p-p collision a bunch of vertices can be reconstructed, but only the one having the highest  $p_T^2$ , in which  $p_T$  is defined as the component of transverse momentum (i.e. perpendicular) to the beam line of the reconstructed tracks, is selected as the primary vertex (PV) of the interaction, while all the others are classified as pile-up vertices. A later algorithm can assign a different vertex as the primary vertex of the interaction if the reconstructed  $\tau$  objects, of interest of this analysis, are associated to another vertex.

A set of Perigee parameters, defined as  $d_0, z_0, \phi, \theta, q/p$ , is used in ATLAS to describe the reconstructed tracks. The first two,  $d_0$  and  $z_0$ , are related to the impact parameter ( $b$ ), which is defined as the distance of closest approach to the vertex, and correspond to the longitudinal  $z_0$  and the perpendicular  $d_0$  components. The two variables  $\phi$  and  $\theta$ , plus the distance of the point to the origin in the  $z$  direction, are the cylindrical coordinates.  $\theta$  is the polar angle, measured between the  $z$ -axis and particle, with  $0 \leq \theta \leq \pi$  and  $\phi$  is the azimuthal angle in the  $x$ - $y$  plane with  $0 \leq \phi \leq 2\pi$ . The relation between the spherical coordinates  $(r, \theta, \phi)$  and the Cartesian coordinates  $(x, y, z)$  is given by the following:

$$\begin{aligned} x &= r \sin(\theta) \cos(\phi), \\ y &= r \sin(\theta) \sin(\phi), \\ z &= r \cos(\theta). \end{aligned} \tag{3.1}$$

$q/p$  is the ratio of the charge and the momentum of the particle.

## 3.2 Object Reconstruction

In this section will be given a brief description of the reconstruction of the main objects involved in this analysis. Electron reconstruction is given in section 3.2.1, muon reconstruction in section 3.2.2, hadronic  $\tau$  reconstruction in section 3.2.3, jets reconstruction in section 3.2.4 and the missing transverse energy in section 3.2.5.

### 3.2.1 Electron Reconstruction

The ATLAS Detector that is responsible for the electron reconstruction is the ECAL. When electrons pass through the calorimeter, they interact with an absorber layer that has the task to stop the particle in order to measure both position and magnitude of the deposited energy. The electrons are also required to match a track which is reconstructed in the ID detector [85].

Before reaching the calorimeter, the electron passes via the pixels, SCT, TRT, and presampler, where it is subjected to a series of interactions. When interacting with other detectors it produces photons that also enter the calorimeter and deposit energy near the electron.

There are two processes in identifying an electron. The first step is pattern recognition, where the candidate's energy loss is examined using the hypothesis of being a  $\pi$  particle. If the likelihood of the candidate particle of being a  $\pi$  is low, the signal is fitted using the electron hypothesis, which takes into consideration the potential energy loss from bremsstrahlung of up to 30% at each intersection. Subsequently, the Gaussian Sum Filter

algorithm is employed to consider the effects of non-linear bremsstrahlung on the electron candidate [86].

In the leptonic channel of this analysis, the criterion for the electron transverse momentum is tightened even more to the recommended values that are advised for electrons matched to the single-electron trigger.

### 3.2.2 Muon Reconstruction

The ATLAS detector dedicated to the muon's reconstruction is the Muon Spectrometer. The reconstruction of the muon is performed by matching the signal of the Muon Spectrometer with tracks detected by the ID. The muons must satisfy  $p_T > 10\text{GeV}$  and  $|\eta| < 2.47$ . Muons that satisfy this criteria are then used as an input to the overlap removal procedure (OLR).

An algorithm is used when signals are found in the muon chamber to compare them to similar hits. If the hit happens in the MDT then the algorithm goes through each layer to find a segment, which represents a collection of linked positions. If the hit happen in the CSC, the segment is reconstructed in the  $\eta$ - $\phi$  plane. RPC and TGC are capable of measuring the perpendicular vector to the bending plane.

A muon track can be created using two segments or a high-quality section. Reconstructed muon tracks may be divided into four categories: combined muons, segment-tagged muons, calorimeter-tagged muons, and extrapolated muons [87]. In order to rebuild the muon tracks, the combined muons category uses information from both the ID and the MS. This strategy uses a global fitting technique to merge data from diverse sources. A muon candidate is typically reconstructed by first taking into account the outermost hit and then comparing it to hits from the tracks in the inner detector. Muons that are reconstructed utilizing at least one segment of the track in the MDT or CSC chambers, along with the data from the MS, are referred to as segment-tagged category muons. The calorimeter-tagged muons are those reconstructed by using the energy deposit in the calorimeter while muons hitting only the MS are categorised as extrapolated muons, of primary interest when recovering muons exceeding the ID range and therefore in the  $2.5 < |\eta| < 2.7$  region.

In the leptonic channel, the criterion for the electron transverse momentum is tightened even more to the recommended values that are advised for muons matched to the single-muon trigger [6].

### 3.2.3 Tau Hadronic Reconstruction

The main reconstructed analysis object is the  $\tau$  lepton. The  $\tau$  hadronic decay mode, as shown in figure 3.1, with the exception of the neutrino, has only decay products that are measurable in the hadronic calorimeters, usually referred to as visible decay product. These decays can be identified and consequently separated from jets originated by quarks or gluons by looking at their properties, such as the energy deposit, as well as the track multiplicity. They can also be separated from electrons with the additional information of the Transition Radiation Tracker. The first algorithm used to reconstruct  $\tau$  particles is

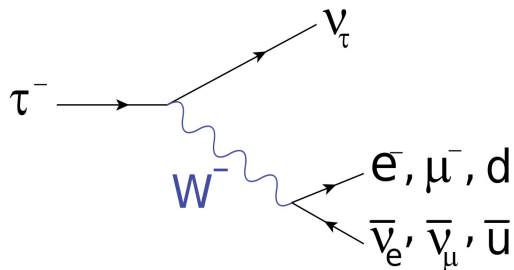


Figure 3.1: Feynman diagram of the  $\tau$  hadronic and leptonic decays.

the anti- $k_T$  algorithm [88], topo-cluster based jet, that uses information from the tracks and the clusters associated to either the core of the hadronic shower ( $R=0.2$ ) or the outer part ( $R=0.4$ ), where  $R$  is the jet radius. It is set to be  $R=0.4$  to distinguish between  $\tau$ 's or jets, with the additional requirements on the jet seeding of a visible hadronic  $\tau$  ( $\tau_{had-vis}$ ) candidate to have  $p_T > 10\text{GeV}$  and  $|\eta| < 2.5$  [89] [90].

The choice of the  $\tau$ -lepton vertex is determined by the amount of momentum provided by tracks associated to a jet. The transverse momentum ( $p_T$ ) of the tracks must be larger than 1 GeV as well as the closest distance in the transverse plane ( $d_0$ ) between the track and the candidate  $\tau$ -lepton vertex must be less than 1 mm in order to identify the vertex. Similar to this, the closest distance in the longitudinal plane has to fulfill  $|\Delta z_0 \sin(\theta)| < 1.5$  mm, where  $\Delta z_0$  is the closest approach along the longitudinal axis, and  $\theta$  is the polar angle of the track.

The energy deposited in the calorimeter cells is used to calculate the energy of the candidate  $\tau$ -lepton [90]. The direction of the candidate  $\tau$ -lepton is identified by considering the vectors of the clusters of calorimeter cells within a cone, which is constrained by an angular difference of  $\Delta R < 0.2$  from the direction of the seed jet that emerges from the candidate  $\tau$ -lepton vertex.

The reconstructed  $\tau$  identification, and separation from jets, is then performed by an RNN algorithm [89] which uses as input information from the reconstructed tracks and the energy deposited in the hadronic calorimeter by the  $\tau$  decay products.

To discriminate the  $\tau$  candidates from the electron fakes ( $\tau_{had-vis}$ ), mainly produced in background processes such as  $Z \rightarrow ee+jets$ , a multivariate discriminant (Electronic Boosted Decision Tree (eBDT)) has been developed [90]. Similarly to the RNN algorithm, it provides discrimination based on tracks and shower information, with additional information from the TRT.

The number of tracks associated to the  $\tau$  candidate equal to 1 or 3, separates the 1-prong case from the 3-prong case. A transverse momentum cut of  $p_T > 20$  GeV and a pseudorapidity cut  $|\eta| < 2.47$ , except for the crack region ( $1.37 < |\eta| < 1.52$ ), are then applied.

The RNN Tau ID score is a number between  $[0,1]$  where 0 represent a Fake  $\tau$  while 1 is a True  $\tau$ . The RNN score threshold is set to 0.01 and, to avoid muons reconstructed as  $\tau$ 's, a veto criteria is applied (MUONORL). The choice to apply a RNN Tau ID score of 0.01 was made to minimize the loss of signal events at the level of 2.5% for the ggH process and 4% for the VBF process while populating of the anti-ID region, a region with orthogonal selection criteria of the identification and isolation algorithms, essential for the data-driven estimation of the hadronic visible  $\tau$  misidentified as jet.

Table 3.1 shows the recommended transverse momentum requirements for the  $\tau_{had-vis}$  in both channels.

Trigger chain	Data period	HLT $p_T$ 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	$p_T(\mu) > 20$ GeV	$p_T(\mu) > 21$ GeV
	2016-2018	$p_T(\mu) > 26$ GeV	$p_T(\mu) > 27.3$ GeV
Ditau	2015-2018	$p_T(\text{leading } \tau_{had-vis}) > 35$ GeV	$p_T(\text{leading } \tau_{had-vis}) > 40$ GeV
		$p_T(\text{subleading } \tau_{had-vis}) > 25$ GeV	$p_T(\text{subleading } \tau_{had-vis}) > 30$ GeV

Table 3.1: The  $p_T$  thresholds applied at trigger and off-line stage for the selected electrons, muons and  $\tau_{had-vis}$ .

### 3.2.4 Jet Reconstruction

To reconstruct jets in ATLAS different detectors are involved, in particular the calorimeters which provide the principal signal for jet measurement and the particle trajectories detected by the ID.

A growing volume algorithm is used to expand the calorimeter cells from the calorimeters seed cell defining what is known as the topo-cluster, corresponding to the linked calorimeter cells [88].

The jet calibration is performed in several steps. Initially, the vertex with the highest  $\sum p_T^2$  is selected as primary vertex and cleaned by the unwanted pile-up effect. Following that, a simulation-based calibration is performed to correct the jets four-momentum. Then, jet quality is improved by consider tracks variables, muon segment and calorimeter readout. Finally, calibration of the jet energy is performed taking into account real data [9].

The algorithm able to distinguish between hard-scatter and pile-up vertices is called Jet vertex Tagger (JVT). It is based on a multivariate algorithm trained on simulated samples and applied to jets within the range  $20 < p_T < 50$  GeV and  $|\eta| < 2.4$ . For each jet, the algorithm returns a score that can be a number between 1, corresponding to the maximum likelihood of being an hard-scatter vertex, and 0. Jets with a transverse momentum higher the 50 GeV are consider hard-scattering jets.

It is also used the DL1r b-tagging algorithm [91] to identify the b-quark jets and the algorithm result is applied to every reconstructed jet with  $p_T > 20$ GeV and  $|\eta| < 2.5$ .

### 3.2.5 Missing Transverse Energy

Missing Transverse Energy (MET) in ATLAS is reconstructed from calorimeters clusters and from reconstructed muons, electrons, and  $\tau_{had-vis}$  as well as the whole jet collection.

The missing transverse energy is an estimate of the imbalance of the transverse momentum momentum in the detector,  $p_T^{MISS}$  [10]. To calculate the vector, firstly the transverse momenta of all reconstructed final-state objects, such as electrons, muons and  $\tau$ 's, are considered, then it is obtained by considering the negative sum of them 3.2.

The MET used in this analysis is obtained with the official ATLAS tool, named ATLAS Missing Transverse Energy Tool, that has been set to use the default TIGHT criteria requiring a transverse momentum greater than 30GeV. The Missing Transverse Energy Tool runs its own dedicated OLR procedure in order to prevent track and calorimeter clusters from being double-counted.

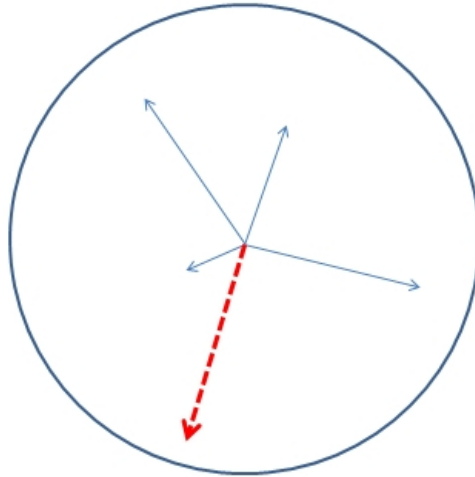


Figure 3.2: A transverse view of the detector. The blue arrows represent observed particles and their momentum while the red arrow represent the missing energy of the particle(s) that did not interact with the detector but due to the conservation laws can be calculated.

### 3.2.6 Overlap Removal

Physics objects are reconstructed in ATLAS from a set of tracks plus the calorimeter information, either electromagnetic or hadronic. The information can be shared between different reconstructed objects and therefore an ambiguity can arise. The procedure applied in ATLAS to remove this ambiguity is called overlap removal procedure (OLR) and it is applied to the baseline object through the so-called AssociationUtils package [92].

## 3.3 Tau Decay Plane Reconstruction

This analysis depends on the specific  $\tau$  decay mode for the reconstruction of the  $\tau$  decay plane, so it is crucial to get a precise measurement of the  $\tau$  decay products and their vectors. The Tau Particle Flow method [4], developed during ATLAS Run-2 data, satisfies both requirements. The algorithm provides a more detailed classification of tau decay modes than the reconstruction algorithm used in Run 1 and referred to as Baseline.

By combining tracking detector measurements and calorimeter measurements, the particle flow algorithm determines the charge and momentum of the charged and neutral pions from  $\tau$  decay. Utilizing a Boosted Decisions Tree (BDT) and counting the number of charged and neutral pions, the decay mode is determined using the properties of the  $\tau$  decay products and the number of reconstructed pions. The analysis takes advantage of three different BDTs, developed to distinguish between the different decay modes of the  $\tau$  with the aim of improving the neutral pions identification.

The possibility of misidentified tracks and  $\tau$  decays, in which 1 prong or 3 prong (1p/3p)

are involved leads to a small misidentification.

The Tau Particle Flow efficiency that correspond to the efficiency of correctly identifying a certain decay mode, calculated using simulated  $Z \rightarrow \tau\tau$  Run-1 events [4], is given in figure 3.3, which shows for each generated decay mode, listed on the horizontal axis, the probability of being classified as a certain reconstructed decay mode, as listed on the vertical axis. Overall, the 74.7% of tau candidates are correctly classified. The algorithm used in Run 1 was able to only distinguish between 1-prong and 2-prong while the Tau Particle Flow algorithm uses new categories involving charged hadrons such as  $\pi^+$  and  $\pi^-$  as well as  $\pi^0$ s. Taking advantage of the superior resolution provided by the  $\pi^\pm$  tracks, the Tau Particle Flow provides significantly improved angular and energy resolution compared to Baseline algorithm. This analysis uses samples from Run-2 MC and data. Similar efficiency matrices are also calculated using SHERPA 2.2.1, for the Run-2  $Z \rightarrow \tau\tau$  simulated samples in figure 3.4, and using Powheg +Pythia8 for the signal ggH sample, figure 3.5.

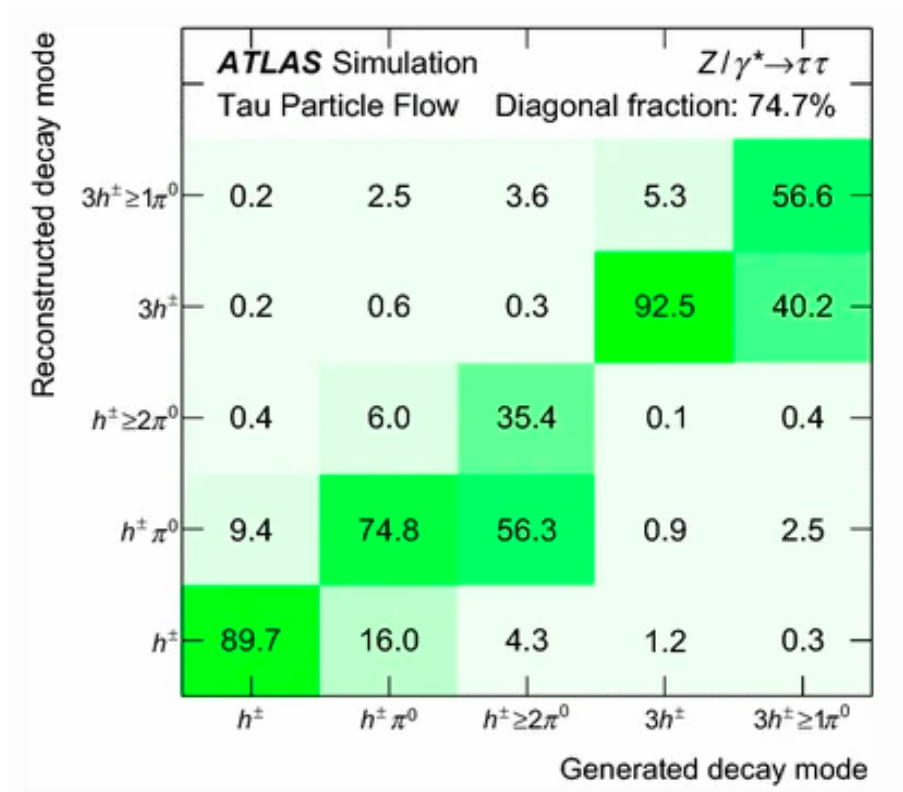


Figure 3.3: Efficiency matrix of the  $\tau$  particle flow decay mode classification [4]. Each element in the efficiency matrix is normalized by the total number of events in its corresponding generated decay mode (column), while an element in the purity matrix is normalized by the total events in the corresponding reconstructed decay mode (row).



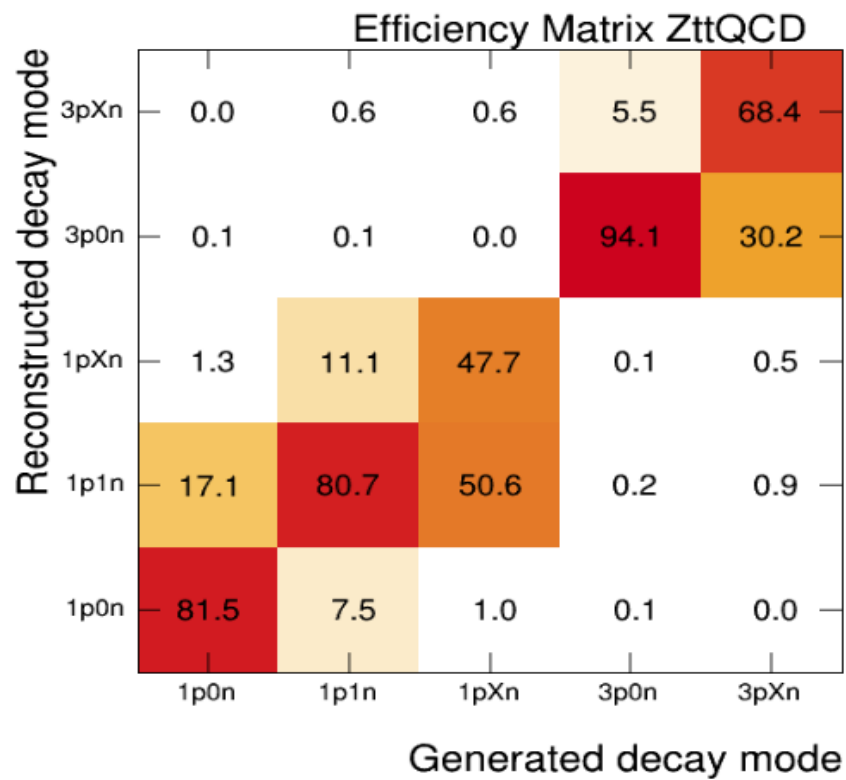


Figure 3.4: Efficiency matrix of the Sherpa 2.2.1  $Z \rightarrow \tau\tau$  MC sample used in the analysis. Each element in the efficiency matrix is normalized by the total number of events in its corresponding generated decay mode (column), while an element in the purity matrix is normalized by the total events in the corresponding reconstructed decay mode (row).

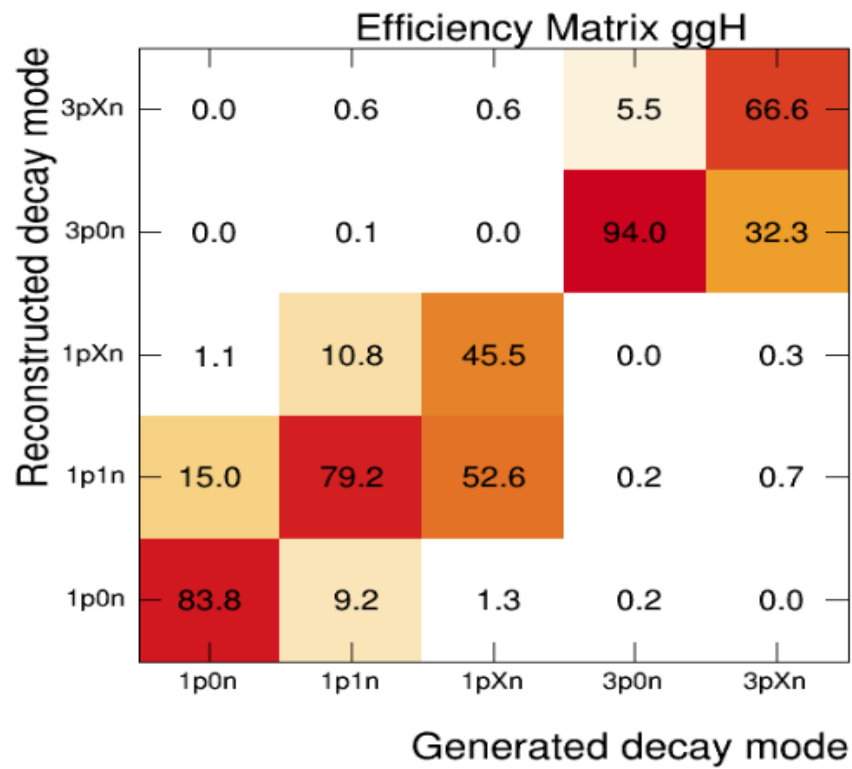


Figure 3.5: Efficiency matrix of the Powheg + Pythia8  $H \rightarrow \tau\tau$  ggH signal sample used in the analysis. Each element in the efficiency matrix is normalized by the total number of events in its corresponding generated decay mode (column), while an element in the purity matrix is normalized by the total events in the corresponding reconstructed decay mode (row).

# ANALYSIS STRATEGY

4

Considered as a topic of major interest in Higgs physics, the measurement of a slight but measurable deviation of the SM Higgs from a pure CP-even, where CP is a composition of Charge symmetry and Parity symmetry state, would be a signature of physics beyond the standard model. This strong motivation, since the birth of the Higgs physics, was the aim of different studies conducted to evaluate the Higgs coupling to other SM particles. A list of Higgs boson production and decay modes is shown in figure 4.1.

A coupling or interaction is said to be CP-even (scalar) if it remains unchanged under the combined CP transformation. If however the coupling sign changes under the combined CP transformation it is said to be CP-odd (pseudoscalar).

Different motivations bring the attention to the coupling with fermions. This is mainly due to the fact that the CP-odd contributions are introduced in the Higgs to vector gauge boson coupling as high order operators that yield a small contribution to the coupling itself. On the other side, the Higgs to fermion coupling could occur at tree level, moving the LHC attention to two channels,  $H \rightarrow \tau\tau$  and  $H \rightarrow bb$ , bearing in mind that the  $bb$  decay mode suffers from significant QCD jet background at the LHC, while the  $\tau\tau$  decay mode has a more favourable signal to background ratio.

Recently both ATLAS [42] and CMS [93] studied the Higgs CP properties in the Higgs to top quark channel and both experiments rejected the pure CP-odd hypothesis with a significance respectively of  $3.9\sigma$  and  $3.2\sigma$ .

The studies shown in the following is the first attempt for ATLAS to reject the CP-odd hypothesis in the coupling of Higgs to  $\tau$  particles while the first attempt for CMS [94] rejected the pure CP-hypothesis with a significance of  $3.0\sigma$ .

For the purpose of this analysis, both hadronic and leptonic  $\tau$  decays are used to determine the CP-sensitive observable, the mixing angle  $\varphi_{CP}^*$ . With a relevant branching fraction of 65%, the hadronic  $\tau$  decays are a factory of pion mesons, both charged  $\pi^\pm$  or

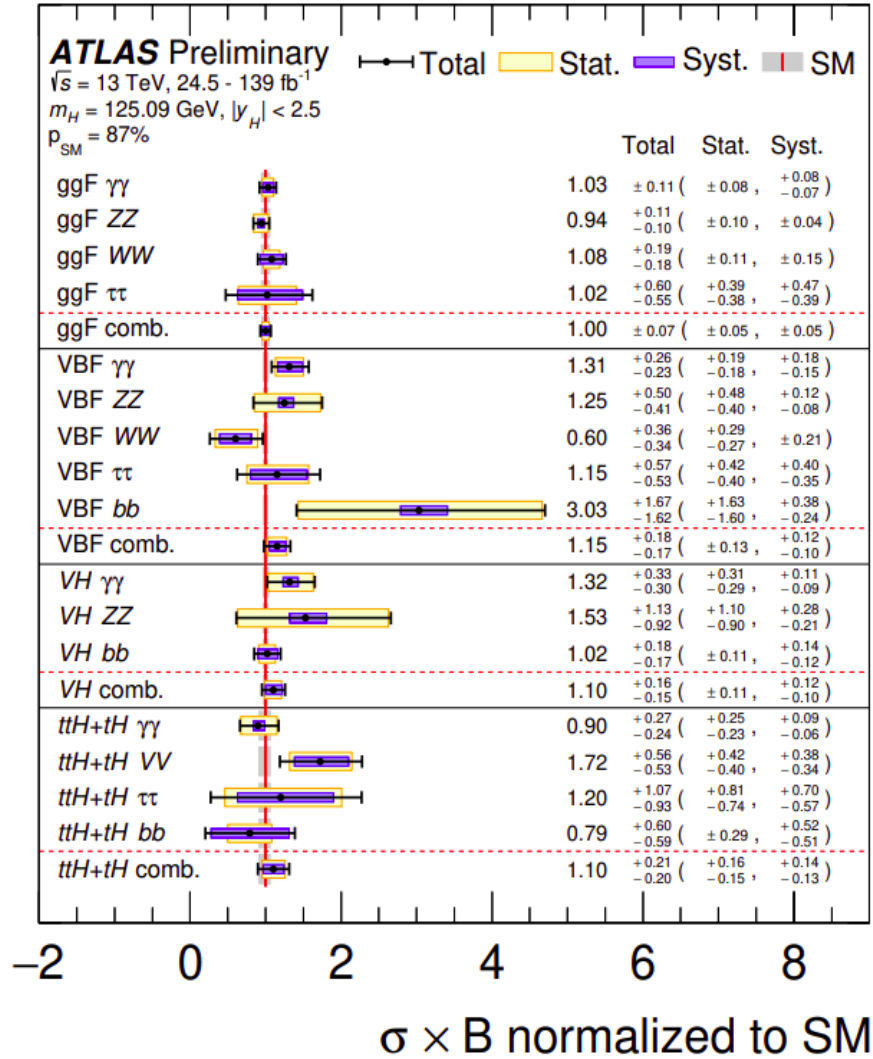


Figure 4.1: Summary of Higgs boson production and decay modes by the ATLAS collaboration [5].

neutral  $\pi^0$ , that are produced with an additional  $\tau$  neutrino.

Leptonic  $\tau$  decays produce muons and electrons, corresponding to the visible part of the decays, whereas electron neutrino and muon neutrino energy is calculated using the missing energy information. Both hadronic and leptonic  $\tau$  branching fraction are summarised in table 4.1:

The nomenclature,  $XPYN$ , or similarly  $XpYn$ , used in this work to classify different hadronic decay modes looks at the decay products and counts the number of X charged particles (P) and the number of Y neutral particles (N). The leptonic decays are simple represented with an "l".

Decay mode	Branching ratio	Nomenclature
$\pi^- \nu_\tau$	10.8 %	1P0N
$\pi^- \pi^0 \nu_\tau$	25.4 %	1P1N
$\pi^- \geq 2\pi^0 \nu_\tau$	10.4 %	1PXN
$3\pi^- \nu_\tau$	9.3 %	3P0N
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.7 %	
$e^- \bar{\nu}_e \nu_\tau$	17.4 %	

Table 4.1: Branching ratio of the  $\tau^-$  hadronic and leptonic decay modes [95], calculated using PYTHIA8 [96].

## 4.1 CP-measurements via $\phi_\tau$ mixing angle.

It is possible to access the different CP-scenarios by looking at the spin correlation in the di- $\tau$  system, that is carried by the taus' decay products. Both longitudinal and transverse components can be accessed, following the notations used in reference [8], through the model-independent Yukawa interaction between the Higgs and the taus:

$$\mathcal{L}_{H\tau\tau} = -\frac{m_\tau}{\nu} k_\tau (\cos \phi_\tau \bar{\tau}\tau + \sin \phi_\tau \bar{\tau}i\gamma_5\tau)H \quad (4.1)$$

where  $\nu = 246$  GeV is the vacuum expectation value of the Higgs field,  $k_\tau > 0$  is the reduced Yukawa coupling strength, and  $\phi_\tau$  (where  $\phi_\tau$  is defined in the range  $[-\pi/2, \pi/2]$ ) is the ‘‘CP-mixing’’ angle, that parametrizes the CP-even and CP-odd components contribution to the coupling between Higgs and the two  $\tau$  particles. Equation 1.25 sets the SM Yukawa coupling as

$$g_l = \frac{m_l \sqrt{2}}{\nu} \quad (4.2)$$

In order to allow a CP odd/even mixture signature, the coupling strength modifier  $k_\tau$  and the mixing angle  $\phi_\tau$  are introduced in equation 4.1. The Higgs boson parity is therefore determined as follows:

**scalar**  $\phi_\tau = 0$ , P = + 1,

**pseudoscalar**  $\phi_\tau = \pi/2$ , P = -1

where P is the parity of the Higgs boson. Any other value of  $\phi_\tau$  is a signature of CP-violating scenarios.

The differential decay width can be written, if  $\beta_\tau = \sqrt{1 - \frac{4m_\tau^2}{m_h^2}}$ , in terms of the tau spin components in the Higgs rest frame as [8] [97]:

$$d\Gamma_{H \rightarrow \tau^+ \tau^-} \propto 1 - s_z^- s_z^+ + \cos(2\phi_\tau)(s_\perp^- \cdot s_\perp^+) + \sin(2\phi_\tau)[(s_\perp^- \times s_\perp^+) \cdot \vec{k}^-] \quad (4.3)$$

The normalised  $\tau^-$  spatial momentum in the Higgs boson rest frame is  $\vec{k}^-$ ,  $\vec{s}^\pm$  are the unity spin vectors of  $\tau^\pm$  in their respective rest frames and the transverse and longitudinal components of  $\vec{s}^\pm$  with respect to  $\vec{k}^-$  are represented, by respectively,  $s_\perp^\pm$  and  $s_\parallel^\pm$ . As is clearly evident in equation 4.3, the only component sensitive to the  $\varphi_{CP}^*$  mixing angle is the transverse spin correlation of the two  $\tau$ 's, that can be obtained from the angular distribution of the  $\tau$  decay products. Specifically, the acoplanarity angle  $\varphi_{CP}^*$ , defined in the interval  $[0, 2\pi]$  is the angle, in the Zero Momentum Frame (ZMF), constructed between the  $\tau$  decay planes spanned by the  $\tau$  decay products, as is schematically shown in figure 4.2.

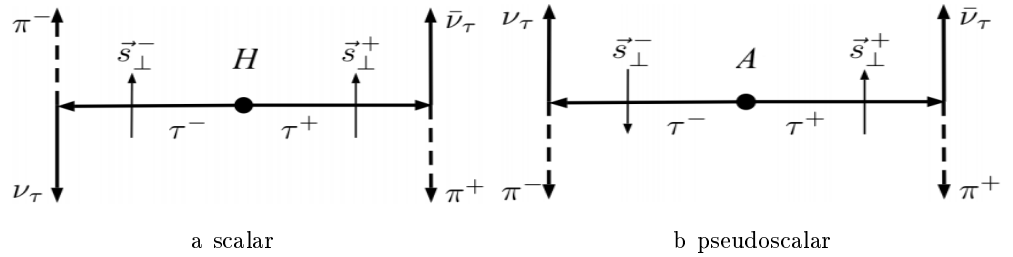


Figure 4.2: Angular correlations of the tau decay products from  $H/A \rightarrow \tau^+\tau^-$  decays. H represents the scalar Higgs boson while A is a pseudoscalar Higgs boson.

A scalar Higgs boson, denoting a SM CP-even case, can be reconstructed if the transverse spin components are parallel and thus the charged pions, the decay products of the  $\tau$ , are emitted back-to-back, whereas antiparallel spin components as well as charged pions emitted into the same direction are a signature of a CP-odd case, and therefore a pseudoscalar Higgs boson. In figure 4.3 the Higgs rest frame is approximated by the visible di-tau in the ZMF; in this case the pions are emitted in opposite directions and the  $\varphi_{CP}^*$  peak is therefore located at  $\pi$  while for the pseudoscalar boson the pions are emitted in the same direction and the  $\varphi_{CP}^*$  peaks are located at 0 and  $2\pi$ , as illustrated in figure 4.4.

The differential decay width, shown in equation 4.3, can be more accurately written in terms of  $\varphi_{CP}^*$  and  $\phi_\tau$  as follows:

$$d\Gamma_{H \rightarrow \tau^+\tau^-} \approx 1 - b(E_+)b(E_-) \frac{\pi^2}{16} \cos(\varphi_{CP}^* - \phi_\tau) \quad (4.4)$$

where the spin analysing power of the particular decay mode is parametrized by the energy-dependent terms  $b(E_\pm)$ , which are equal to 1 for decays where no neutral particles are involved and for the other decays depend on the energy.

The normalised  $\varphi_{CP}^*$  distribution is plotted against the differential decay width in figure 4.4, showing that the SM-like hypothesis of a scalar Higgs boson follows a cosine function while, for the pseudoscalar Higgs boson it is shifted by a factor of  $\pi$ . Any other non-integer

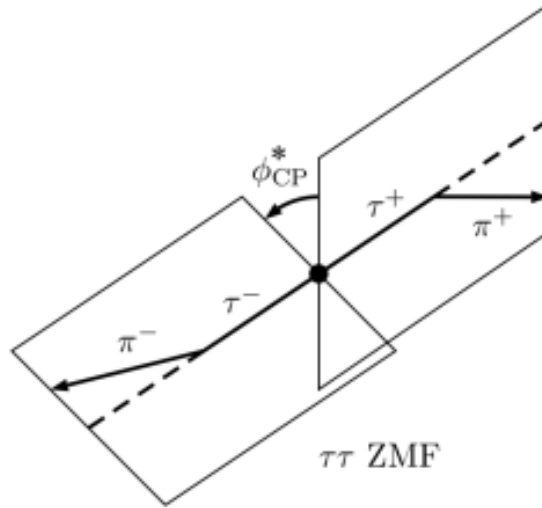


Figure 4.3: Illustration of a decay  $H \rightarrow \tau^+\tau^- \rightarrow \pi^+\pi^- + 2\nu$  in the Zero Momentum Frame. The decay planes spanned by the momenta of the tau leptons and the pions are shown, together with the angle they form [6].

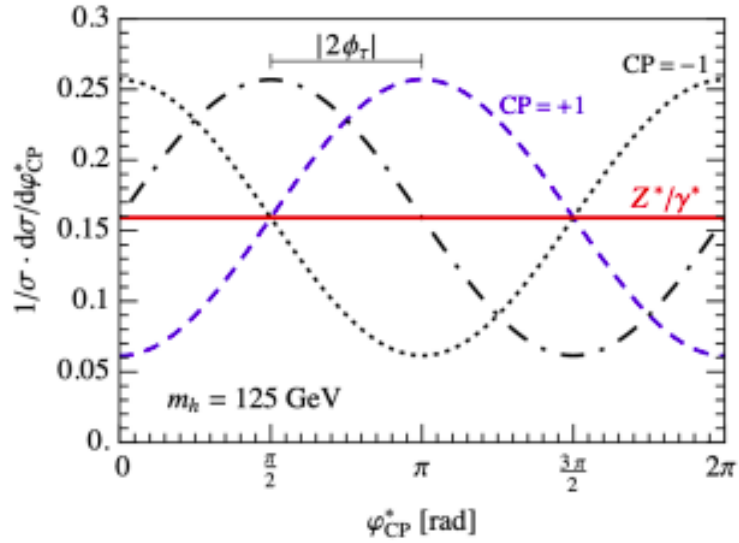


Figure 4.4: Normalised  $\phi_{CP}^*$  distribution for  $pp \rightarrow H/Z^*/\gamma^* \rightarrow \tau^+\tau^-$  production [6].

phase represents a CP-mixing signature, and therefore a BSM. The red line in figure 4.4 shows a crucial feature of  $\phi_{CP}^*$ , a flat and uniform distribution of the background processes thus no sensitivity bias to  $\phi_\tau$ .

Accordingly to the decay mode, two methods are developed in [8] and [97] to build the  $\phi_{CP}^*$ , the IP-method, described in section 4.1.1, and the  $\rho$ -method, described in 4.1.2 which can be used to calculate the CP-mixing angle in presence of  $\pi_0$  particles, through the production of a  $\rho$  meson. The combined IP-Rho method is then described in section 4.1.3 and an a1-method, used for 3-prong decays, in section 4.1.4.

### 4.1.1 IP method

The IP method described in this section takes its name from the impact parameter vector of the charged decay product ( $\pi^\pm, l^\pm$ ) used to reconstruct the decay plane. The direct hadronic and leptonic decays are the best candidate for this method because there is only one visible charged particle, but overall it can be used on any decay mode. Here, this method is applied for the hadronic decay where a  $\tau^\pm$  particle produces a  $\pi^\pm$  and a  $\nu_\tau$  and the leptonic decay where a lepton, either  $\mu$  or  $e$ , is produced with its relative neutrino as well as the tau neutrino.

The impact parameter vector, as shown in figure 4.5, is constructed from the  $\tau$  production vertex towards the pion track. In particular, the spatial vector  $q^\pm$  of the charged particle and its three-dimensional impact parameter are used to produce the decay plane, where the 3D-IP is the distance of closest approach of a charged particle's track to the production vertex of the  $\tau$  leptons, that also corresponds to the Higgs production vertex (Higgs PV), assuming the Higgs boson does not travel. Specifically, the impact parameter vector is defined by the 2D closest approach, the minimal distance to the pion in the transverse plane ( $d_0$ ) and in the longitudinal plane ( $z_0$ ). The relation between the 3D closest approach and the 2D closest approach is given by the assumption that the pion track is a straight line and therefore it is obtained as the minimum distance between the track momentum and the primary vertex.

For kinematic description reasons,  $\hat{n}_+$  and  $\hat{n}_-$  define the normalised impact parameter vectors in the laboratory frame. A boost in the di- $\tau$  rest frame allows one to separate the boosted impact parameter vector  $\hat{n}^{*\pm}$  into its parallel  $\hat{n}_{\parallel}^{*\pm}$  and perpendicular  $\hat{n}_{\perp}^{*\pm}$  normalised components to the spatial vector  $\hat{q}^{*\pm}$ , where  $\hat{q}^*$  represents the momentum of the pion. The angle between the decay planes  $\varphi^*$  is defined by the following:

$$\varphi^* = \arccos(\hat{n}^{*+} \cdot \hat{n}^{*-}) \quad (4.5)$$

The equation 4.5 is only valid in the range  $0 \leq \varphi^* \leq \pi$  and therefore any mixing between scalar and pseudoscalar Higgs coupling components would not be covered. To extend the measurement toward  $2\pi$  and to check the  $\tau^\pm$  decay plane side as well a triple correlation is performed as follows:

$$O_{CP}^* = \hat{q}^{*-} \cdot (\hat{n}^{*+} \times \hat{n}^{*-}) \quad (4.6)$$

The sign obtained by equation 4.6 allows one to define a CP-sensitive angle  $\varphi_{CP}^*$  as



follow:

$$\varphi_{CP}^* = \begin{cases} \varphi^* & O_{CP}^* \geq 0 \\ 2\pi - \varphi^* & O_{CP}^* < 0 \end{cases} \quad (4.7)$$

Figures 4.5 and 4.6 illustrate the definition of the impact parameter vector and the 3-vectors for constructing  $\varphi^*$  in the IP method.

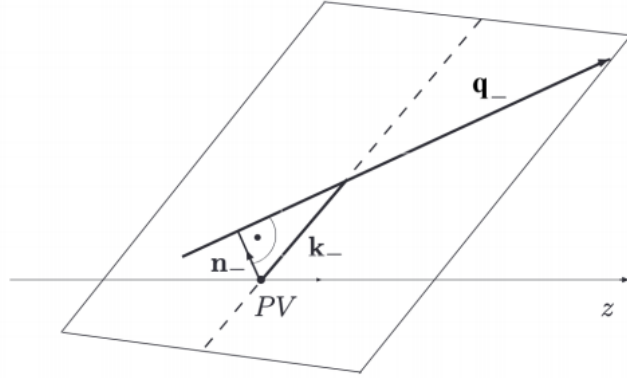


Figure 4.5: Definition of the Impact Parameter vector [7]  $n^\pm$  in the decay plane of either  $\tau^\pm \rightarrow \pi^\pm \nu_\tau$  or  $\tau^\pm \rightarrow l^\pm \nu_\tau$  [6].

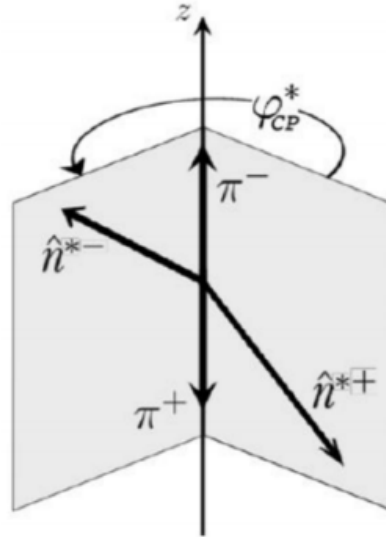


Figure 4.6: Geometrical view of the  $\varphi^*$  angle construction in the IP-method [8].

The comparison between the primary vertex, corresponding to the Higgs production vertex, uncertainties and the magnitude of the impact parameter vectors carries information about the performance of this method. In particular, poor performance is achieved

when the impact parameter vectors are smaller or similar to the primary vertex uncertainties. Among those uncertainties is the significance of the track impact parameter in the transverse plane  $d_0^{sig}$  that can be obtained by dividing the  $d_0$  vector by its uncertainty. Defining a  $d_0^{sig}$  threshold is therefore crucial to better split the events with 1P0N and leptonic decays in two regions based on this value, the high and low sensitivity regions. Several  $d_0^{sig}$  threshold values were tested and it was decided to use  $d_0^{sig} = 2.1$  for the hadronic case and  $d_0^{sig} = 2$  for the leptonic case. The  $\phi_{CP}^*$  shape distributions are shown for the 1P0N-1P0N channel in figures 4.7.

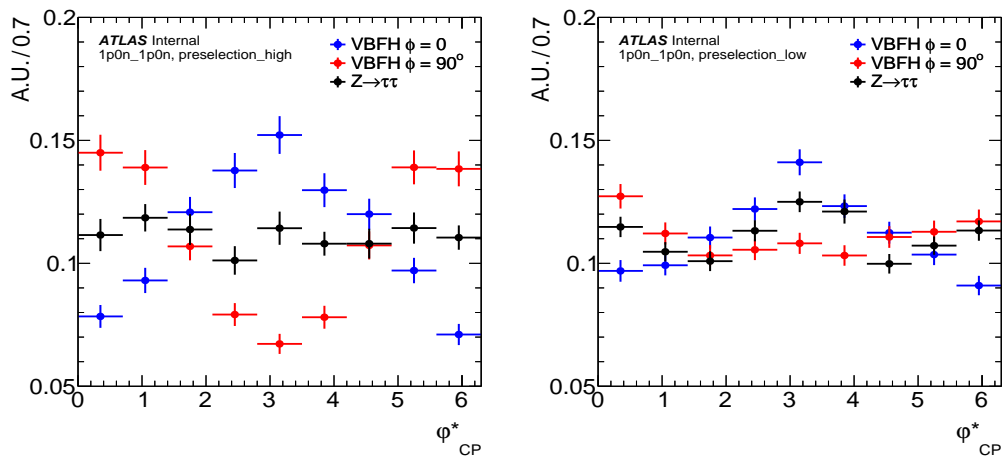


Figure 4.7:  $\phi_{CP}^*$  shape distribution for 1p0n-1p0n VBF preselection high(left) and low(right).

#### 4.1.2 $\rho$ method

The  $\rho$  method, as its name suggests, it is based on the  $\rho$  meson decay, equation 4.8, where in addition to a charged pion a neutral pion is emitted and they can both be used to calculate the decay plane and subsequently the  $\varphi_{CP}^*$  observable. In this analysis, this method is applied to the 1p1n-1p1n and 1p1n-1pXn channels, where for the 1p1n-1pXn case the 4-momentum sum of the whole set of  $\pi^0$ s is considered.

$$\tau^\pm \rightarrow \rho^\pm \nu, \quad \rho^\pm \rightarrow \pi^\pm \pi^0 \quad (4.8)$$

Similarly to the IP method, to define the  $\varphi^*$  we need to decompose the  $\rho$  4-momentum vector into its components, thus  $q^{*-}$  ( $q^{*+}$ ) and  $q^{*0-}$  ( $q^{*0+}$ ) vectors identify the charged and neutral components of  $\rho^-$  ( $\rho^+$ ). By boosting both charged and neutral components in the di- $\rho$  rest frame the angle  $\varphi^+$  and the triple-odd correlation  $O_{CP}^*$  can be obtained, and, as a consequence, an angle  $\varphi^*$  can be defined:

$$\varphi^* = \begin{cases} \varphi^* & O_{CP}^* \geq 0 \\ 2\pi - \varphi^* & O_{CP}^* < 0 \end{cases} \quad (4.9)$$

To make the  $\varphi^*$  variable sensitive to the  $\phi_\tau$  angle the epsilon variables, equation 4.10, should be taken into account. They are constructed to be sensitive to the sign of the  $\tau$  spin-analysing functions.

$$y_- = \frac{E_{\pi^-} - E_{\pi^0}}{E_{\pi^-} + E_{\pi^0}}, \quad y_+ = \frac{E_{\pi^+} - E_{\pi^0}}{E_{\pi^+} + E_{\pi^0}} \quad (4.10)$$

where  $E_{\pi^{\pm,0}}$  represent the pions energy in the lab frame. All of those considerations allow to define  $\varphi_{CP}^*$  as follows:

$$\varphi_{CP}^* = \begin{cases} \varphi^* & y_+ y_- \geq 0 \\ \varphi^* + \pi & y_+ y_- < 0 \end{cases} \quad (4.11)$$

Different values of the  $|y_+ y_-|$  epsilon product were tested and it was found that larger absolute values are more sensitive to  $\phi_\tau$ . To account for this effect a threshold is used to categorize the events into high and low sensitive regions.

The  $\phi_{CP}^*$  shape distributions are shown for the 1p1n-1p1n channel in figure 4.8 and for the 1p1n-1pXn/1pXn-1p1n channel in figure 4.9.

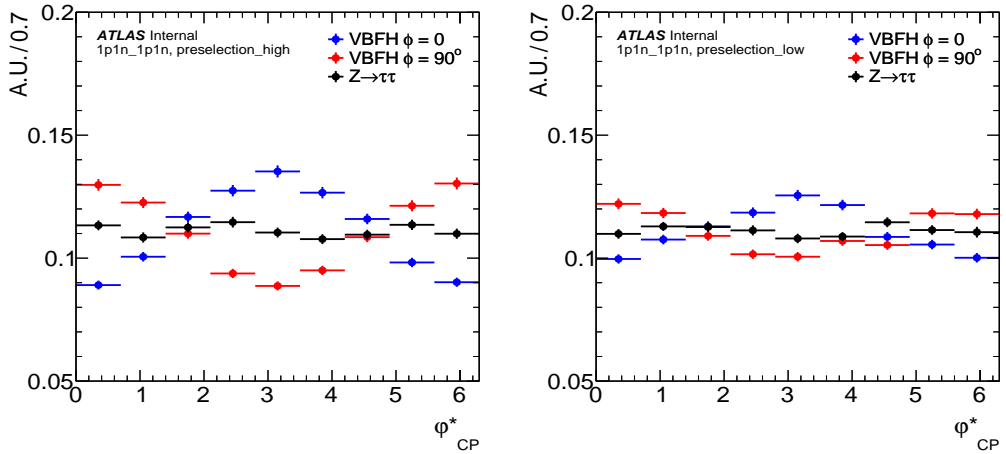


Figure 4.8: The  $\phi_{CP}^*$  shape distributions for 1p1n-1p1n VBF preselection high(left) and low(right).

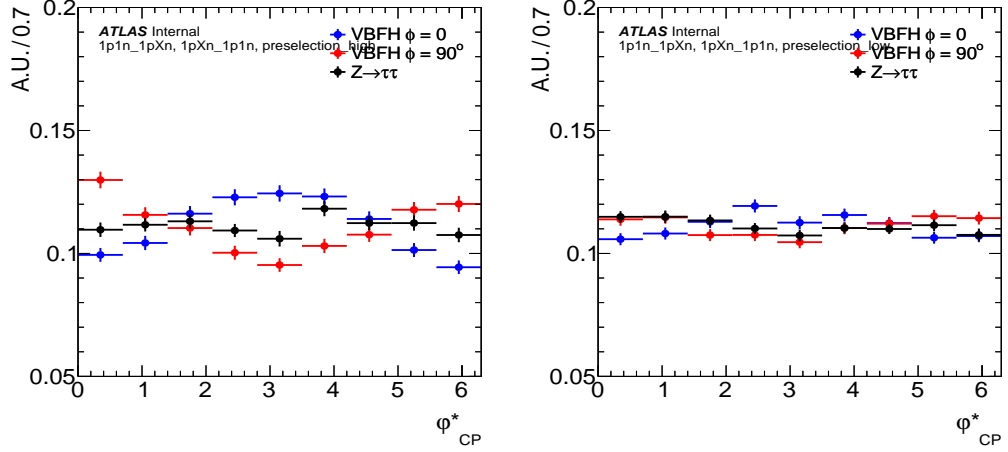


Figure 4.9: The  $\phi_{CP}^*$  shape distributions 1p1n-1pXn 1pXn-1p1n VBF preselection high(left) and low(right).

### 4.1.3 IP- $\rho$ method

The IP method in section 4.1.1 and the  $\rho$  method in section 4.1.2 apply to the cases where both taus decay to the same final state. For the cases when one  $\tau$  decays to a charged pion while the other through a  $\rho$  meson the IP- $\rho$  method is used. For a given decay  $H \rightarrow \tau^- \tau^+ \rightarrow \pi^- \rho^+ + 2\nu_\tau$ , the  $\varphi_{CP}^*$  is defined in the  $\pi^- \rho^+$  rest frame and all relative momentum vectors are boosted in it. In the case of the  $\tau^- \rightarrow \pi^- \nu_\tau$  the three-momentum of the neutral pion is identified by the vector  $\hat{q}^{*0+}$ , similarly the three-momentum component of the impact parameter vector of the  $\pi^-$  is represented by  $\hat{n}^{*-}$ . Thus the  $\varphi^*$  is defined by:

$$\varphi^* = \arccos(\hat{q}^{*0+} \cdot \hat{n}_\perp^{*-}) \quad (4.12)$$

Similarly to the already described methods, a triple-odd correlation with the three momentum vector of the negative pion is helpful to compute the angle sensitive to the Higgs CP properties, equation 4.13. Using again the upsilon function in equation 4.10 the  $\varphi^*$  takes the form:

$$\varphi^* = \begin{cases} \varphi^* & O_{CP}^* \geq 0 \\ 2\pi - \varphi^* & O_{CP}^* < 0 \end{cases} \quad (4.13)$$

$$\varphi_{CP}^* = \begin{cases} \varphi^* & y_+ \geq 0 \\ \varphi^* + \pi & y_+ < 0 \end{cases} \quad (4.14)$$

Similarly to the previous methods, the separation in  $\varphi_{CP}^*$  can be enhanced by categorising the event based on the  $d_0^{sig}$  value, as well as the upsilon  $y$  function, equation 4.10.

This method is applied for the  $\phi_{CP}^*$  construction of the following channels: 1p0n-1p1n in figure 4.10, 1p0n-1pXn in figure 4.11, l-1p1n in figure 4.12 and l-1pXn in figure 4.13.

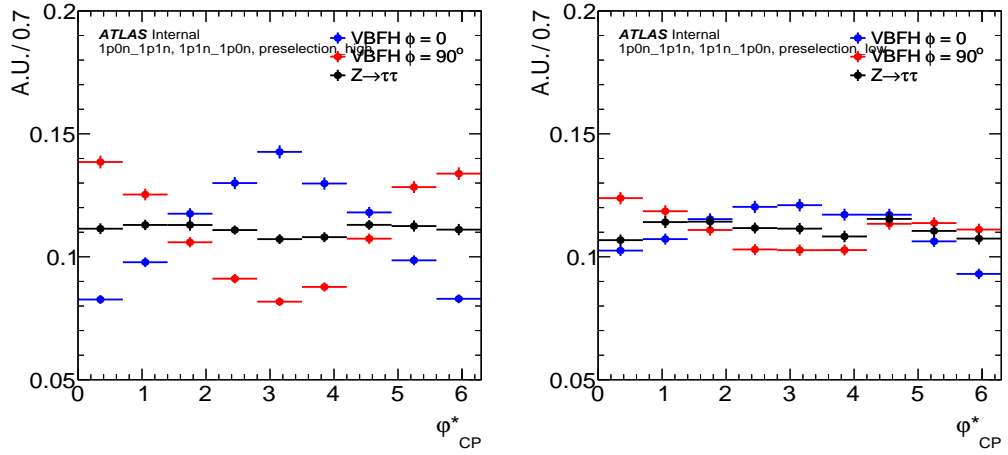


Figure 4.10:  $\phi_{CP}^*$  shape distribution for 1p0n-1p1n 1p1n-1p0n VBF preselection high(left) and low(right).

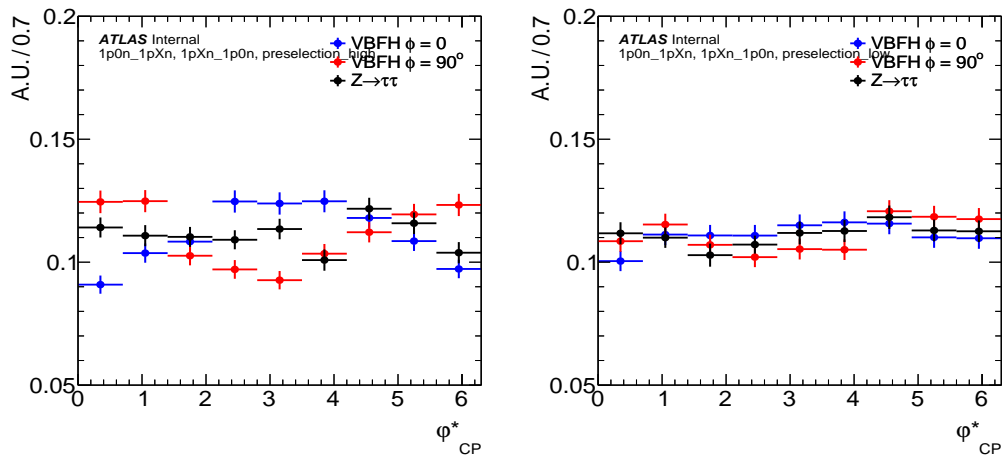


Figure 4.11:  $\phi_{CP}^*$  shape distribution for 1p0n-1pXn 1pXn-1p0n VBF preselection high(left) and low(right).

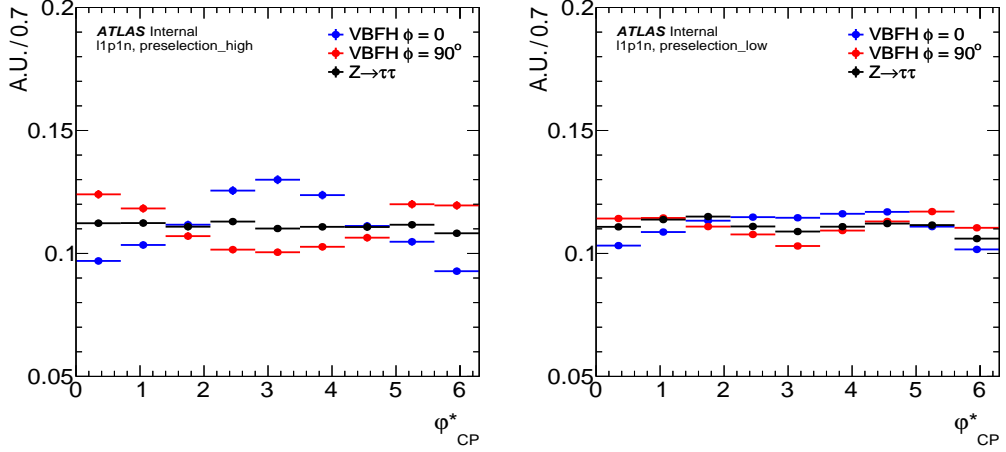


Figure 4.12:  $\phi_{CP}^*$  shape distribution for 1-1p1n VBF preselection high(left) and low(right).

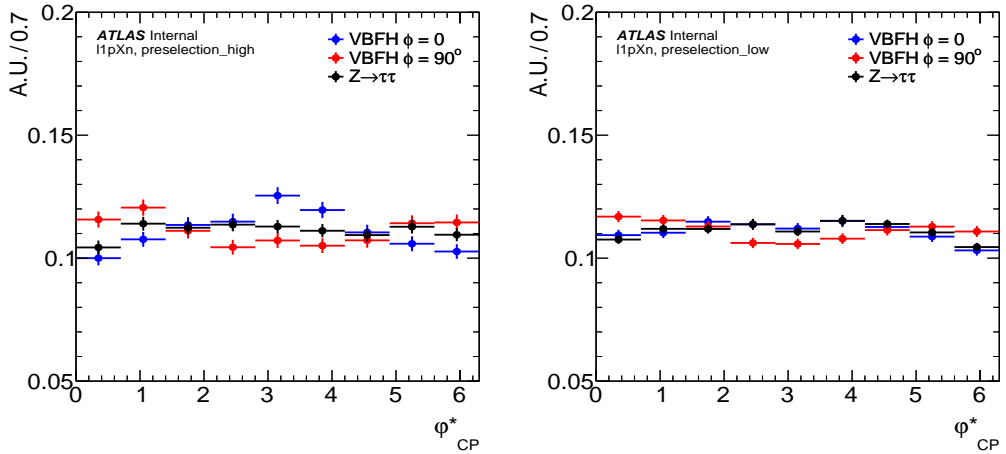


Figure 4.13:  $\phi_{CP}^*$  shape distribution for 1-1pXn VBF preselection high(left) and low(right).

#### 4.1.4 $a_1$ method

The  $a_1$  method, similarly to the  $\rho$ -method, is mainly developed to account for decays where 3-prongs are involved, it is shown in equation 4.15.

$$\tau \rightarrow a_1 \nu_\tau, \quad \text{where } a_1^\pm \rightarrow \rho^0 \pi^\pm \text{ and } \rho^0 \rightarrow \pi^+ \pi^- \quad (4.15)$$

To construct the  $\varphi^*$  observable two of the three prongs needs to be grouped together to form the  $\rho^0$  resonance. The  $\varphi^*$  observable can be calculated by applying the rules of the  $\rho$ -method, with the condition that the  $\pi^0$  is now represented by the  $\rho^0$ .

As already tested in [92] to have a slightly optimal CP-sensitivity, the way to group together the charged  $\pi$ 's is by requiring the opposite charges for the two pions forming the  $\rho$  and to take into account the mass effects to calculate the correct upsilon variable, but this approach did not provide the best sensitivity when evaluating the  $a_1$  method performance on MC events and therefore a different approach is considered.

Specifically, in this analysis two important points are taken into account. The first is

related to the intermediate  $\rho^0$  meson that is formed by an arbitrary  $\pi^+\pi^-$  pair, related to the broad resonance peak of  $\rho^0$  at the reconstruction level. The second point, as already stated, is related to the mass effect in the upsilon variable used in this method which needs to be defined considering those effects. The mass effect is dependent on the decay channel and even in the 1p1n case the mass terms are negligible due to the very close mass values between the  $\pi^0$  and  $\pi^\pm$ , but differs considerably for the  $a_1$  case where the  $\rho^0$  has a mass of  $m_{\rho^0} = 770$  MeV while  $m_{\pi^\pm} = 139$  MeV and therefore the associated mass terms in the upsilon calculation have a bigger effect in the construction of  $\varphi_{CP}^*$ .

The upsilon calculation for this method, as already defined in [92], is:

$$y_{a_1}^\pm = \frac{E_{\rho^0} - E_{\pi^\pm}}{E_{\rho^0} + E_{\pi^\pm}} - \frac{m_{a_1}^2 - m_{\pi^\pm}^2 + m_{\rho^0}^2}{2m_{a_1}^2} \quad (4.16)$$

Only events with three charged pions and zero neutral pions are considered (3p0n). In particular, the events considered in the  $\tau_{hadhad}$  channel is the 1p1n-3p0n while, in the  $\tau_{lephad}$  channel is l-3p0n. The corresponding  $\phi_{CP}^*$  are shown in figures 4.14 and 4.15.

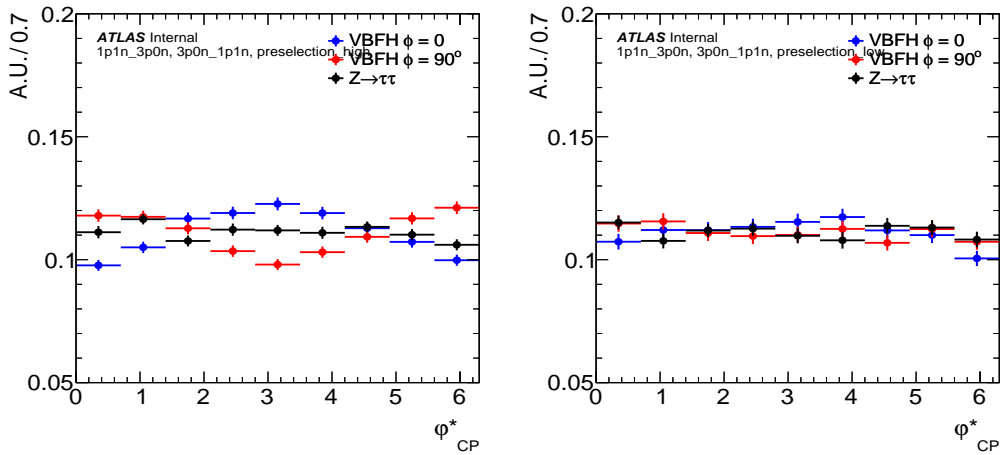


Figure 4.14:  $\phi_{CP}^*$  shape distribution for 1p0n-3p0n VBF preselection high(left) and low(right).

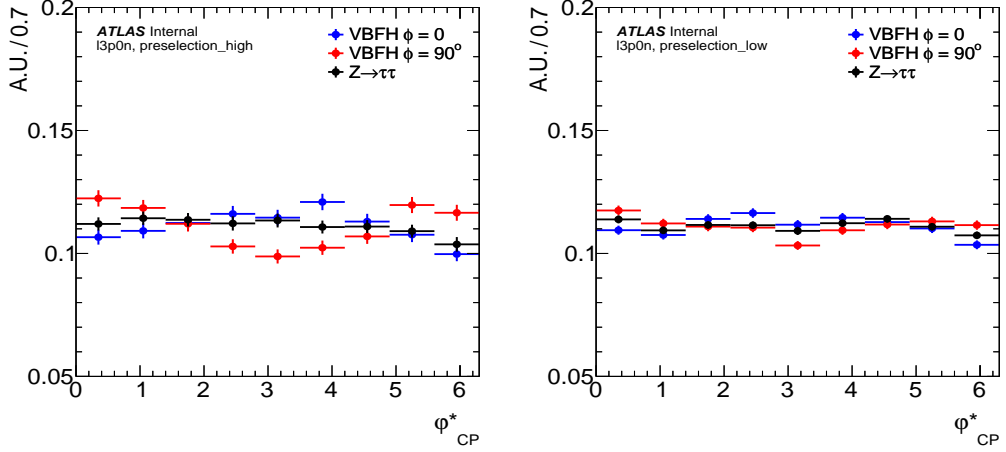


Figure 4.15:  $\phi_{CP}^*$  shape distribution for 1-3p0n VBF preselection high(left) and low(right).

## 4.2 Analysis Decay mode

As stated above, three different methods are implemented to construct the observable decay plane. Section 4.1.1 describes the IP method used for decays containing only one charged particle, the rho method, described in section 4.1.2, extends the calculation of the decay plane to the case where there is an additional neutral particle, while, the  $a_1$  method, described in section 4.1.4, is used in the 3-prong decays case.

Those methods need to be combined to accommodate both  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  decay channels used in the analysis. As a result, 10 different  $di - \tau$  decay modes are identified, four of them for the  $\tau_{\text{lep}}\tau_{\text{had}}$  case, 1-1p0n, 1-1p1n, 1-1pXn, 1-3p0n, and six of them related to the  $\tau_{\text{had}}\tau_{\text{had}}$  case, 1p0n-1p0n, 1p0n-1p1n, 1p1n-1p1n, 1p0n-1pXn, 1p1n-1pXn, 1p1n-3p0n. They are summarised in table 4.2, where for each decay mode the  $\phi_{CP}^*$  construction method used to calculate it and the relative branching fraction of the  $di - \tau$  decays are shown. The general Feynman diagram of the decay of the Higgs in two  $\tau$  is given in figure 4.16, whereas the Feynman diagram of the leptonic  $l$  decay mode in figure 4.17a, the 1p0n decay mode in figure 4.17b and the 1p1n decay mode in figure 4.17c.



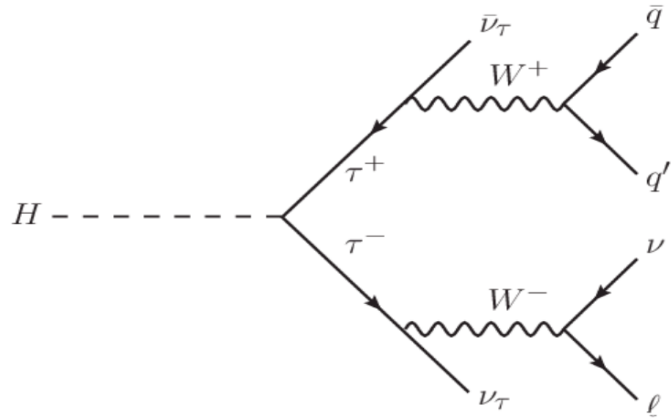
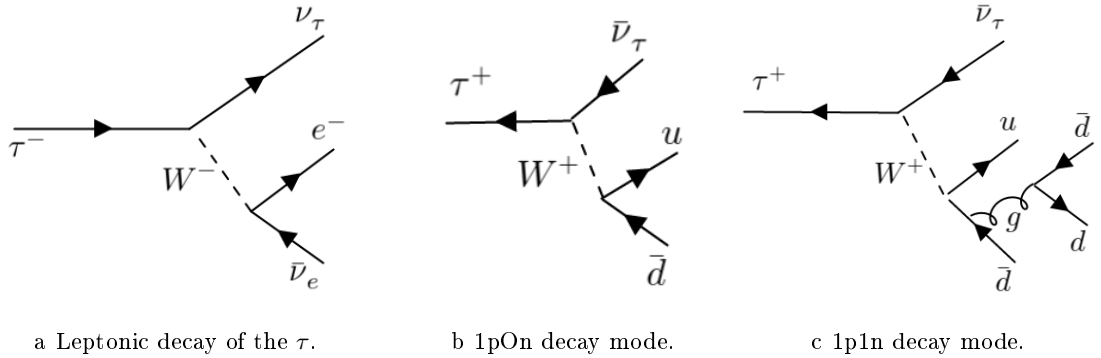


Figure 4.16: Higgs decay in a  $\tau^+\tau^-$  couple. The  $\tau^+$  decays hadronically producing a quark-antiquark while the  $\tau^-$  decays leptonically producing an electron and electron-antineutrino.



a Leptonic decay of the  $\tau$ .

b 1p0n decay mode.

c 1p1n decay mode.

Figure 4.17: Figure 4.17a shows the Feynman diagram of a  $\tau$  decays into an electron, an electron-antineutrino and a tau-antineutrino. Figure 4.17b shows the Feynman diagram of the hadronic  $\tau$  decay in a couple quark-antiquark plus a tau-antineutrino and a  $\pi^+$  particle (1p0n decay mode) while figure 4.17c shows the hadronic  $\tau$  decays into a  $\pi^+$  and a  $\pi^0$  (1p1n decay mode).

Decay Channel	Decay mode combination	Method	Fraction in all di-tau decay
$\tau_{lep}\tau_{had}$	l-1p0n	IP-IP	7.6%
	l-1p1n	IP- $\rho$	17.8%
	l-1pXn	IP- $\rho$	7.3%
	l-3p0n	IP- $a_1$	6.5%
$\tau_{had}\tau_{had}$	1p0n-1p0n	IP-IP	1.2%
	1p0n-1p1n	IP- $\rho$	5.5%
	1p1n-1p1n	$\rho$ - $\rho$	6.4%
	1p0n-1pXn	IP- $\rho$	2.2%
	1p1n-1pXn	$\rho$ - $\rho$	5.3%
	1p1n-3p0n	$\rho$ - $a_1$	4.7%

Table 4.2: Decay mode combinations used in the analysis with the relative fraction.

### 4.3 Background estimation

In any field of particle physics research it is of fundamental importance to evaluate and identify all the background processes having a contribution to the signal in order to get valid conclusions and to produce sensible results. With this in mind, the main background contributions to this analysis come from events where two  $\tau$  particles are emitted by the  $Z$  and  $\gamma$  bosons, as well as events, such as the  $W$  boson decaying to a real  $\tau$  plus jets, having as a final decay product a misidentified QCD jet. In addition to those, a small contribution is also given by  $t\bar{t}$  and di-boson decays.

The majority of the processes considered in the background are taken from the Monte Carlo simulation such as  $Z \rightarrow \tau\tau + \text{jets}$ ,  $W \rightarrow \tau\nu + \text{jets}$  as well as di-boson events and  $t\bar{t}$ , while data-driven methods are employed when a hadronic  $\tau$  comes from a mis-reconstructed jet. Therefore, control regions are deeply linked with the background modeling, constructed to make the best validation of the background processes.

The main background process, corresponding to  $Z \rightarrow \tau\tau + \text{jets}$  events, is modeled and estimated by using SHERPA simulations. Two normalisation factors are defined for  $VBF\_0$  and  $VBF\_1$   $Z$  control regions (ZCR) respectively to drive the corresponding normalisation factors of the  $Z \rightarrow \tau\tau$  events obtained from data.

The second dominant source of background, comes from jets which are wrongly identified as  $\tau$  leptons. They are mainly due to  $W + \text{jets}$  events, as well as QCD multijet and top production. The estimation of this background is performed using a data driven method called Fake-Factor method. The fake-factor, by using either data and MC simulated samples, can estimate the contribution of fake events in distinct regions. An anti- $\tau$  region, defined as the region where the  $\tau$  candidate satisfy all analysis requirements with the exception of the  $\tau$  identification requirement. The estimations of observables are estimated using the events occurring within this range. Equation 4.17 is used to calculate the expected yield of the fake events in the signal region.

( $N_{Data}^{anti-\tau}$  represent the number of events which fall in the corresponding anti- $\tau$  region,  $N_{MC,nojet}^{anti-\tau}$  is the number of non-fake events falling in the same anti- $\tau$  region of Monte Carlo simulated samples. The difference between those two quantities is multiplied by a fake-factor  $\mathcal{F}$ , defined in equation 4.18.

$$N_{FAKES}^{SR} = (N_{Data}^{anti-\tau} - N_{MC,nojet}^{anti-\tau})\mathcal{F} \quad (4.17)$$

The combined fake factor in equation 4.17 is constructed from the individual fake-factor of the most relevant processes, weighted by the relative expected fractional contribution in the anti- $\tau$  region:

$$\mathcal{F} = R_{QCD}F_{QCD} + R_W F_W + R_{Top} F_{Top} \quad (4.18)$$

The top production fake-factor is not expected to play a dominant role in the analysis, mostly due to the very low level, almost 3%, of the MC prediction over the analysis, which lead to its embedding in the W fake-factor and therefore the combined fake-factor take the form of:

$$\mathcal{F} = R_{QCD}F_{QCD} + R_W F_W \quad (4.19)$$

where  $R_{QCD}$ , corresponding to the fraction of QCD multijet events, is retrieved from data, while the fraction of W events is  $R_W = 1 - R_{QCD}$ . The fake-factor  $F_i$ , either  $F_{QCD}$  or  $F_W$ , are calculated separately in the control regions. In order to enrich the region with fake events, the control regions for these background sources are defined by inverting the selection criteria [98] [99]. The Fake Factor calculation is shown in figure 4.20.

$$F_i = \frac{N_{Data}^{pass-\tau, CR_i} - N_{MC, nojets}^{pass-\tau, CR_i}}{N_{Data}^{anti-\tau, CR_i} - N_{MC, nojets}^{anti-\tau, CR_i}} \quad (4.20)$$

$N_{Data}^{pass-\tau, CR_i} - N_{MC, nojets}^{pass-\tau, CR_i}$  represent the difference between the number the number of events passing the  $\tau$  identification in the data and MC samples, while the difference  $N_{Data}^{anti-\tau, CR_i} - N_{MC, nojets}^{anti-\tau, CR_i}$  correspond to the number of events failing the  $\tau$  identification in the data ad MC.

The fake-factors are then summed up with the expected fractional contribution of that process,  $R_i$ .

The fake-factor method has the following properties. Fake-factors related to different data periods can be combined as well as for different decay channels and the fake-factor calculated in Z control region can be used in the signal regions.

# DATA AND SIMULATED SAMPLES

## 5

This chapter will describe the data in section 5.1 and the Monte Carlo (MC) simulated samples in section 5.2, where both data and MC samples are shared with the  $H \rightarrow \tau\tau$  coupling analysis [92]. Moreover, in order to study the different CP hypotheses, CP-even CP-odd and CP-mixed  $H \rightarrow \tau\tau$  dedicated samples are used.

### 5.1 Data

The data used in this analysis correspond to the full  $pp$  Run 2 dataset. The proton-proton collisions occur every 25 ns and the total integrated luminosity is equal to  $139 fb^{-1}$ . Table 5.1 shows, for each year, the integrated luminosity ( $\int \mathcal{L} dt$ ) and the related Good Run List (GRL) files.

Year	$\int \mathcal{L} dt \text{ (fb}^{-1}\text{)}$	Good Run List XML file
2015	3.22	data15_13TeV_periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good_25ns.xml
2016	32.99	data16_13TeV_periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml
2017	44.31	data17_13TeV_periodAllYear_DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_All_Good_25ns-Triggerno17e33prim.xml
2018	58.45	data18_13TeV_periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_All_Good_25ns-Triggerno17e33prim.xml

Table 5.1: GRL files used in the analysis.

## 5.2 Monte Carlo simulations

To generate all the CP-hypotheses simulated samples a set of tools, each one with a specific goal, are being used.

The three important steps to simulate are the Higgs production, which is carried out by the POWHEG tool by simulating the di- $\tau$  system, the Higgs decay and hadronisation, carried out by PYTHIA 8 and eventually the TauSpinner [100, 101, 102] tool which takes care of the tau polarization. The shared procedure of generation of the simulated samples initially provides the simulation of the di- $\tau$  system without the spin correlations, also known as the unpolarized samples. Then, the TauSpinner tool, by applying event weights to the unpolarized samples, introduces the spin correlations and creates the templates for each of the CP-hypotheses.

The need to consider the spin effect in the  $\tau$ -decays pushed the development of a tool able to do it, and indeed TauSpinner can therefore add or remove the spin-effects in these interactions. The way in which the tool does that, is by firstly estimating the  $\tau$ -decay polarization by checking the  $\tau$  decay kinematics stored in the MC truth block. Then, TauSpinner provides the derivation of the event weight, which can be used to either add or remove the spin effects of the process, by considering the Born-level matrix element of the process itself.

This procedure gives the possibility to avoid the event generation for each specific CP template, because the same simulated events can be used, by applying different weights, to produce different CP-templates.

Table 5.2 show an overview of the signal MC generators. The simulated Higgs boson signal samples include the principal Higgs boson production modes at LHC: Gluon Gluon Fusion (ggF), VBF, and the two associated production processes,  $VH$  and  $t\bar{t}H$ . Gluon-gluon fusion is the dominant process with an expected cross-section of 48.58 pb, the second dominant process is the VBF with 3.781 pb,  $VH$  with 2.2496 pb and  $t\bar{t}H$  with 0.507 pb. They are initially simulated, as already stated, without any  $\tau$  spin correlations for each of the production modes and subsequently weights are applied to obtain the different CP hypotheses.

### 5.2.1 Signal generator

**ggF** The gluon gluon fusion accuracy is calculated at the Next-to-Next-Leading Order (NNLO) QCD with the POWHEG NNLOPS program [103, 104, 105, 106, 107]. This precision is achieved, for arbitrary  $gg \rightarrow H$  observables, by reweighting the Higgs boson rapidity spectrum MiNLO [108, 109, 110], where MiNLO represent the Multiscale Improved

Next-to-Leading Order (NLO). The Pythia8 [95] setup uses the The PDF4LHC15 NLO Parton Distribution Function (PDF) [111] and AZNLO tune [112]. The ggF prediction coming out from the MC sample is then normalised to the Next-to-Next-to-Next-Leading Order (N3LO) cross-section and, in addition to that, QCD corrections plus electroweak corrections at next-to-leading NLO order are also applied [113, 50, 114, 115, 116, 117, 118, 119, 120, 121, 122]. The decay branching ratios are calculated with HDECAY [123, 124, 125] and PROPHECY4F [126, 127, 128].

**Vector Boson Fusion** POWHEG [105, 106, 107, 129] is the tool used to generate the Higgs boson production via the VBF channel while non-perturbative effects and parton showers are simulated with Pythia8 [95]. POWHEG brings with it a next-to-leading order (NLO) accuracy. To take into account effects due to either finite heavy-quark masses or soft-gluon resummation it is also tuned to the next-to-next-to-leading order accuracy (NNLO). Similarly to the gluon-gluon fusion case, the PDF4LHC15 PDF set [111] and the AZNLO tune [112] of Pythia8 are used. The MC prediction is then normalised to an approximate-NNLO QCD cross section plus NLO electroweak corrections [56]. The decay branching ratios are calculated with HDECAY [123, 124, 125] and PROPHECY4F [126, 127, 128].

**VH** POWHEG is the tool for the the production of the Higgs boson with the associated vector boson, responsible for the generation of the samples while non-perturbative effects and partons showers are simulated with Pythia8. The POWHEG prediction accuracy is calculated at the next-to-leading order for the VH bosons plus one jet production. The accuracy of the loop responsible for the  $gg \rightarrow VH$  process is at leading order and it is also simulated separately. The PDF4LHC15 PDF set and the AZNLO tune of Pythia8 are used. The MC prediction for this process is then normalised to an NNLO QCD accuracy plus the NLO EW corrections for the proton-proton interaction producing VH final state at NLO, while, for the  $gg \rightarrow VH$  process [130, 131, 132, 133, 134, 59, 135] the QCD accuracy is calculated at the next-to-leading-logarithm. The decay branching ratios are calculated with HDECAY and PROPHECY4F.

**$t\bar{t}H$**   $t\bar{t}H$  events are simulated using the PowhegBox v2 generator [105, 106, 107, 136, 137] with a next-to-leading order accuracy plus the NNPDF3.0nlo [138] PDF set. Events are then interfaced to Pythia8 using the A14 tune [139] and the NNPDF2.3lo [138] PDF set. EvtGen v1.6.0 is used to perform the bottom and charm decays.

Process	Generator		PDF set		Tune	Order
	ME	PS	ME	PS		
$H \rightarrow \tau\tau/WW$						
ggH	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	NNLO + NNLL
VBF	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	(N)NLO
VH	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	NNLO
$t\bar{t}H$	Powheg	Pythia8	NNPDF30		A14	NLO

Table 5.2: Overview of the data generators used for the main signal samples.



## 5.2.2 Background generator

**V+jets** Sherpa v2.2.1 [140] is the generator used to simulate the V+jets process. Different accuracy is used according to how many partons are involved in the process. The NLO Matrix Element (ME) is the accuracy used for up to two partons, while, when considering more partons in the process, Leading Order (LO) matrix elements accuracy is used. Both of them are calculated using Comix [141] and OpenLoops 1 [142, 143, 144] libraries. Both cases are then matched with SHERPA parton showers [145] using either MEPS@NLO [146, 147, 148, 149] prescription or the set of tuned parameters suggested by the SHERPA team. The set of PDF used in this channel is NNPDF3.0nnlo [150] and the samples are normalised to a NNLO order prediction [151].

**Electroweak production** Final states involve either  $lj\bar{j}$ ,  $l\nu j\bar{j}$  or  $\nu\nu j\bar{j}$ , where  $l$  is a lepton,  $\nu$  a neutrino and  $j$  is an hadron, are generated with Sherpa v2.2.1. The accuracy for this process is at leading order matrix elements with up to two additional parton emissions. Following the MEPS@LO prescription [147], the matrix elements are merged with SHERPA parton showers using the set of default tuned parameters suggested by the SHERPA authors. The NNPDF3.0nnlo set of PDF is used. In order to avoid the overlap with the semi-leptonic diboson topologies, VBF approximation is applied to the simulated samples.

**tt** tt samples are simulated using PowhegBox v2 [152] generator with an NLO accuracy using the NNPDF3.0NLO PDF set with the  $h_{damp}$  parameter set to  $1.5 m_{top}$  [153], where  $h_{damp}$  is a resummation damping factor and one of the parameters that controls the matching of POWHEG matrix elements to the parton shower and  $m_{top}$  is the quark top mass. To simulate the parton shower, hadronisation and underlying event, Pythia8.230 [154] generator is used with parameter set to A14 tune [139] plus the NNPDF2.3lo PDF set [155]. EvtGen v1.6.0 [156] is utilised for bottom and charm decays.

**Single-top s-channel** PowhegBox v2 is the generator responsible for the single-top s-channel production with an NLO QCD accuracy using the NNPDF3.0nlo parton distribution function set. Pythia8.230 with the A14 tune and the NNPDF2.3lo PDF set is then used to interface the events.

**Single-top t-channel** Similarly to the single-top s-channel, PowhegBox v2 [157] is the generator of this channel with a NLO accuracy in the four-flavour, rather than five-flavour, scheme with NNPDF3.0nlo PDF set. Again Pythia8.230 with the A14 tune and

the NNPDF2.3lo PDF set is then used to interface the events.

**$t\mathbf{W}$**  The process involving the associated production of a top quark with W bosons ( $t\mathbf{W}$ ) is simulated using the PowhegBox v2 [158] generator with next-to-leading order accuracy in QCD using the five-flavour scheme at the NNPDF3.0nlo parton distribution function set. In order to remove any interference or overlap with the  $t\bar{t}$  production, the diagram removal scheme is used [159]. Pythia8.230 with the A14 tune and the NNPDF2.3lo PDF set is then used to interface the events.

**diboson** Sherpa v2.2.1 is the generator, depending on the specific process, responsible for the simulation of processes involving samples of dibosons final states (VV), including off-shell effects and Higgs-boson contributions, when needed. Either fully leptonic final states, where both bosons decay leptonically, or semileptonic final states, where one of the two bosons decays hadronically, are generated using matrix elements at NLO order QCD accuracy when only one additional parton emission is involved and at leading order accuracy when three additional parton emissions are involved. The matrix element accuracy of the loop-induced processes  $gg \rightarrow VV$ , for both semileptonic and fully leptonic case, is calculated at leading order for at maximum one parton additional emission.

The matrix elements are then matched and merged with the SHERPA parton shower based on Catani-Seymour dipole factorisation [160, 161] using the MEPS@NLO prescription. The OpenLoops 1 library is then used to calculate the virtual QCD corrections. The NNPDF3.0nnlo set of PDF is used plus a set of tuned parton-shower parameters developed by the SHERPA authors.

Finally, to consider also the pile-up corrections, Pythia8, with the NNPDF2.3lo set of PDF and the A3 tune [162], is used to model the effect of multiple interactions in the same and neighbouring bunch crossing on top of the original hard-scatter event.

Table 5.3 shows an overview of the background MC generators.

Process	Generator		PDF set		Tune	Order
	ME	PS	ME	PS		
Background						
V+jets	SHERPA 2.2.1		NNPDF30		SHERPA	NNLO
$t\bar{t}$	Powheg	Pythia8	NNPDF2.3		A14	NLO
single top	Powheg	Pythia8	NNPDF2.3		A14	NLO
di-boson	SHERPA 2.2.1		NNPDF30		SHERPA	NNLO

Table 5.3: Overview of the MC generators used for the main background samples.

# EVENT SELECTION

## 6

The events are selected following a recipe that includes three main steps. In the first one, events are classified by looking at the proper decay channel they belong to and for each channel the associated trigger requirements are applied. The aim of the second step is to define the signal regions, and to do so events are categorised to extract the ones that contain Higgs events, identify the production modes and all possible topologies. Lastly, to optimise the observable's sensitivity, Higgs's events are divided by hadronic tau decay mode and by kinematic regions.

This analysis shares the same data and Monte Carlo (MC) simulated samples with the  $H \rightarrow \tau\tau$  coupling analysis. In addition, this analysis uses dedicated signal  $H \rightarrow \tau\tau$  samples with different CP hypotheses. The trigger requirements are kept identical to the previous full Run 2  $H \rightarrow \tau\tau$  coupling analysis [92], and in turn the lep-had and had-had baseline preselection is kept similar to the  $36fb^{-1}$  Run-2 data  $H \rightarrow \tau\tau$  coupling analysis [99].

### 6.1 Trigger selection

According to the channel different triggers are used. In particular, the di-hadronic  $\tau$  triggers are used for the  $\tau_{\text{had}}\tau_{\text{had}}$  channel, while the single lepton triggers are used for  $\tau_{\text{lep}}\tau_{\text{had}}$ . Table 6.1 shows the lowest unrescaled triggers while table 6.2 shows the  $p_t$  thresholds requirements for the trigger to match the leptonic candidates in the used datasets, 2015-2018. In addition, for the di-hadronic case, the individual single  $\tau$ 's trigger objects are used to match the two  $\tau_{\text{had-vis}}$ , see chapter 3, candidates to the relative legs of the di- $\tau$  object.

An additional level-1 calorimeter trigger jet with  $p_t$  higher than 25 GeV and  $\eta$  lower than 3.2 is taken into account in the trigger used in the dataset 2016-2018 to mitigate the increasing of instantaneous luminosity. In addition to that, the leading jet is required to be

located within  $R \leq 4$ , see section 3.2.4 for the  $R$  definition, with the jet Region of Interest (ROI), that satisfies the jet part of the trigger criteria. The two motivations for that are to guarantee a consistent kinematics selection in the 2015 and 2016-2018 dataset and to avoid possible biases due to the turn-on effects of the trigger's jet part.

Trigger	Data period	HLT Chain name (in the menu)
Single electron	2015	e24_lhmedium_L1EM20VH, e60_lhmedium or e120_lhloose
	2016-2018	e26_lhtight_nod0_ivarloose, e60_lhmedium_nod0 or e140_lhloose_nod0
Single muon	2015	mu20_iloose_L1MU15 or mu50
	2016-2018	mu26_ivarmedium or mu50
Ditau	2015	tau35_medium1_tracktwo_tau25_medium1_tracktwo_L1TAU20IM_2TAU12IM
	2016	tau35_medium1_tracktwo_tau25_medium1_tracktwo
	2017	tau35_medium1_tracktwo_tau25_medium1_tracktwo_03dR30_L1DR_TAU20ITAU12I_J25
	2018	tau35_medium1_tracktwoEF_tau25_medium1_tracktwoEF_03dR30_L1DR_TAU20ITAU12I_J25

Table 6.1: Trigger items used in this analysis. The single lepton triggers are used in  $\tau_{\text{lep}}\tau_{\text{had}}$  channel. The di-hadronic  $\tau$  trigger is used in the  $\tau_{\text{had}}\tau_{\text{had}}$  channel.

Trigger chain	Data period	HLT $p_T$ threshold	Offline $p_T$ threshold
Single electron	2015	$p_T(e) > 24 \text{ GeV}$	$p_T(e) > 25 \text{ GeV}$
	2016-2018	$p_T(e) > 26 \text{ GeV}$	$p_T(e) > 27 \text{ GeV}$
Single muon	2015	$p_T(\mu) > 20 \text{ GeV}$	$p_T(\mu) > 21 \text{ GeV}$
	2016-2018	$p_T(\mu) > 26 \text{ GeV}$	$p_T(\mu) > 27.3 \text{ GeV}$
Ditau	2015-2018	$p_T(\text{leading } \tau_{\text{had-vis}}) > 35 \text{ GeV}$ $p_T(\text{subleading } \tau_{\text{had-vis}}) > 25 \text{ GeV}$	$p_T(\text{leading } \tau_{\text{had-vis}}) > 40 \text{ GeV}$ $p_T(\text{subleading } \tau_{\text{had-vis}}) > 30 \text{ GeV}$

Table 6.2: The  $p_T$  thresholds applied at trigger and offline stage for the selected electrons, muons and  $\tau_{\text{had-vis}}$ .

## 6.2 Preselection

On top of the trigger requirements, specific sets of preselection cuts are applied for both  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  channel, summarised in table 6.3 and 6.4.

Specifically, the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel is firstly required to have a single reconstructed  $\tau_{\text{had-vis}}$  lepton and a single reconstructed lighter lepton such as an electron or muon, and, in addition, both  $\tau$  candidates must have opposite charges with an absolute value equal to 1, with the condition that the  $\tau_{\text{had}}$  has to pass the  $\tau$  jets medium identification criterion. A transverse momentum higher than 40 GeV is required for the leading jet of the event as a consequence of the level 1 calorimeter trigger being in use since 2016 data-taking [163]. Additional conditions, due to trigger requirements, are applied on both  $\tau$  transverse momentum such as a thresholds of 21 GeV for the single electron trigger chain and 27.3 GeV for the single muon trigger chain, while a 30 GeV  $p_T$  cut is applied in the di- $\tau$  trigger chain. In the semileptonic  $\tau_{\text{lep}}\tau_{\text{had}}$  channel a set of additional cuts are applied, such as a  $\Delta R$  cut (defined as the angular distance between the  $\tau$  candidates), a pseudorapidity

cut between both candidates  $\Delta\eta$  to reject non-resonant events, a number of track cuts to identify the number of prongs and therefore classify the events, a transverse missing energy  $E_T^{miss} > 20$  GeV cut and, finally, a transverse mass cut on the lepton to reject  $W + \text{jets}$  processes.

Category and Event Selection			
<b>Preselection</b>			
Leading jet with $p_T > 40$ GeV			
One lepton (e or $\mu$ ) as $\tau_{\text{lep}}$ candidate			
One $\tau_{\text{had-vis}}$ candidate, classified as 1p0n, 1p1n, 1pXn and 3p0n			
Opposite charge between $\tau_{\text{lep}}$ and $\tau_{\text{had-vis}}$ candidates			
$p_T(\tau_{\text{lep}}) > 21$ to $27.3$ GeV, $p_T(\tau_{\text{had}}) > 30$ GeV			
$\tau_{\text{had-vis}}$ passing <i>medium</i> tau identification with RNN			
Electron and muon passing <i>medium</i> identification			
Electron passing the FCLoose isolation and muon passing FCTightTrackOnly isolation			
$\Delta R_{\tau\tau} < 2.5$			
$\Delta\eta_{\tau\tau} < 1.5$			
Collinear approx.: $0.1 < x_1 < 1.4$ , $0.1 < x_2 < 1.2$			
$E_T^{miss} > 20$ GeV			
$m_T < 70$ GeV			
<b>Signal regions</b>			
$110 < m_{\tau\tau}^{\text{MMC}} < 150$ GeV			
<b>VBF</b>		<b>Boost</b>	
Second jet $p_T > 30$ GeV		Failed VBF selection	
$m_{jj} > 400$ GeV		$p_T(\text{H}) > 100$ GeV	
$ \Delta\eta_{jj}  > 3.0$ ; $\eta(j_0) \times \eta(j_1) < 0$			
Both $\tau_{\text{had}}$ candidates must lie between the two leading jets in pseudorapidity			
<b>VBF 1</b>	<b>VBF 0</b>	<b>Boost Tight</b>	<b>Boost Loose</b>
BDT(VBF) $> 0.0$	BDT(VBF) $< 0.0$	$\Delta R \leq 1.5$ and $p_T(\text{H}) \geq 140$ GeV	Not tight
<b>Signal optimisation regions</b>			
<b>High</b>			<b>Low</b>
$\ell$ -1p0n,	$(d_0^{\text{sig}}(e) > 2.5$ or $d_0^{\text{sig}}(\mu) > 2.0)$ and $d_0^{\text{sig}}(\tau_{1p0n}) > 1.5$		Not high
$\ell$ -1p1n,	$(d_0^{\text{sig}}(e) > 2.5$ or $d_0^{\text{sig}}(\mu) > 2.0)$ and $ \gamma(\tau_{1p1n})  > 0.1$		
$\ell$ -1pXn,	$(d_0^{\text{sig}}(e) > 2.5$ or $d_0^{\text{sig}}(\mu) > 2.0)$ and $ \gamma(\tau_{1pXn})  > 0.1$		
$\ell$ -3p0n,	$(d_0^{\text{sig}}(e) > 2.5$ or $d_0^{\text{sig}}(\mu) > 2.0)$ and $ \gamma(\tau_{3p0n})  > 0.6$		
<b>Z <math>\rightarrow</math> <math>\tau\tau</math> Control regions</b>			
Same Preselection except with $60 < m_{\tau\tau}^{\text{MMC}} < 110$ GeV			
<b>VBF Z CR</b>		<b>Boost Z CR</b>	
Same VBF selection		Same Boost selection	

Table 6.3: Summary of event selection requirements in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel.

The full hadronic final state  $\tau_{\text{had}}\tau_{\text{had}}$  is characterized by two  $\tau_{\text{had-vis}}$  reconstructed objects. Compared to the semi leptonic channel, the event preselection uses a slightly higher leading jet transverse momentum corresponding to a threshold of 70 GeV. Similarly to the  $\tau_{\text{lep}}\tau_{\text{had}}$  case, additional level-1 calorimeter trigger is applied by including a  $|\eta|$  cut lower than 3.2 [163]. The main preselection requirement is to have exactly two different  $\tau_{\text{had-vis}}$  candidates, both associated to the same primary vertex, with two additional criteria cor-

responding to, an opposite charge and an absolute value equal to one. On top of charge criteria, to pass the preselection, both  $\tau$ s have to pass the medium identification criteria for rejection against non- $\tau$  jets. To safely reject uninteresting leptons such as electrons and muons, a corresponding veto is applied. The two  $\tau$ s have to satisfy a cut on the angular distance  $\Delta R$  between them corresponding to  $0.6 < \Delta R < 2.5$ , a relative pseudorapidity  $\Delta\eta$  lower than 1.5 and also an odd number of tracks to differentiate between 1-prong case and 3-prong case. cuts on transverse momentum, 40 GeV for the leading  $\tau$  and 30 GeV for the sub-leading  $\tau$ , and missing energy,  $E_{miss}^T$  higher than 20 GeV, are applied to improve the invariant mass estimation.

Category and Event Selection			
<b>Preselection</b>			
Leading jet with $p_T > 70$ GeV, $ \eta  < 3.2$			
Two $\tau_{had}$ candidates with opposite charges, classified as 1p0n, 1p1n, 1pXn and 3p0n			
Both $\tau_{had}$ candidates associated to same reconstructed primary vertex			
$p_T(\tau_1) > 40$ GeV, $p_T(\tau_2) > 30$ GeV			
Both passing <i>medium</i> tau identification with RNN			
No electron or muons			
$0.6 < \Delta R_{\tau\tau} < 2.5$			
$\Delta\eta_{\tau\tau} < 1.5$			
$E_{miss}^T > 20$ GeV			
Collinear approx.: $0.1 < x_1 < 1.4$ , $0.1 < x_2 < 1.4$			
<b>Signal regions</b>			
$110 < m_{\tau\tau}^{MMC} < 150$ GeV			
<b>VBF</b>		<b>Boost</b>	
Second jet $p_T > 30$ GeV		Failed VBF selection	
$m_{jj} > 400$ GeV		$p_T(H) > 100$ GeV	
$ \Delta\eta_{jj}  > 3.0$ ; $\eta(j_0) \times \eta(j_1) < 0$			
Both $\tau_{had}$ candidates must lie between the two leading jets in pseudorapidity			
<b>VBF 1</b>	<b>VBF 0</b>	<b>Boost Tight</b>	<b>Boost Loose</b>
BDT(VBF) > 0.0	BDT(VBF) < 0.0	$\Delta R \leq 1.5$ and $p_T(H) \geq 140$ GeV	Not tight
<b>Signal optimisation regions</b>			
<b>High</b>			<b>Low</b>
1p0n-1p0n,	$d_0^{sig}(\tau_0) > 1.5$ and $d_0^{sig}(\tau_1) > 1.5$		
1p0n-1p1n,	$d_0^{sig}(\tau_{1p0n}) > 1.5$ and $ y(\tau_{1p1n})  > 0.1$		
1p0n-1pXn,	$d_0^{sig}(\tau_{1p0n}) > 1.5$ and $ y(\tau_{1pXn})  > 0.1$		
1p1n-1p1n,	$ y_0 y_1  > 0.2$		
1p1n-1pXn,	$ y_0 y_1  > 0.2$		
1p1n-3p0n,	$ y(\tau_{1p1n})  > 0.1$ and $ y(\tau_{3p0n})  > 0.6$		Not high
<b>Z <math>\rightarrow</math> <math>\tau\tau</math> Control regions</b>			
Same Preselection except with $60 < m_{\tau\tau}^{MMC} < 110$ GeV			
<b>VBF Z CR</b>		<b>Boost Z CR</b>	
Same VBF selection		Same Boost selection	

Table 6.4: Summary of event selection requirements in the  $\tau_{had}\tau_{had}$  channel.

### 6.3 Signal regions

Once both  $\tau$ s pass the preselection criteria, signal regions can be constructed by the invariant mass of the system. The algorithm which is responsible to calculate it is the Missing Mass Calculator  $m_{\tau\tau}^{MMC}$ , an advanced likelihood-based algorithm [164]. The signal region, and thus the Higgs boson candidate, is therefore located in the following window  $110 \text{ GeV} < m_{\tau\tau}^{MMC} < 150 \text{ GeV}$  by using the four-vectors of the visible  $\tau$  decay products and the missing transverse energy  $E_T^{miss}$ . The first definition of the  $m_{\tau\tau}^{MMC}$ , defined as the di- $\tau$  invariant mass, appears in the Run-2 data  $H \rightarrow \tau\tau$  coupling cross section measurements [99].

#### Event categorisation regions

For the purpose of this analysis the events falling in the MMC mass range are afterwards categorised in two production modes, vector boson fusion (VBF) and gluon gluon fusion (ggF), where the event has to pass a set of cuts to be classified as VBF, while if it fails to pass those requirements is classified as ggF.

To be classified as VBF event, it has to firstly contain a minimum number of two jets with a corresponding threshold of 70 GeV for the first one and 30 GeV for the second jet, with a relative pseudorapidity higher than 3.0 ( $|\Delta\eta_{jj}|$ ) and an invariant mass higher than 400 GeV. On top of that, the two reconstructed candidate  $\tau$ s must lie between the two pseudorapidity leading jets. The VBF signal region is further split in two categories, named *VBF\_0* and *VBF\_1*, constructed from the output of the VBF BDT, see table 6.3 and table 6.4. The reason behind this choice is to enhance the vector boson fusion production mode over the ggF production mode and the main background known as  $Z \rightarrow \tau\tau + \text{jets}$ . That further split is based on the Multivariate analysis (MVA)-BDT discriminant output coming from the VBF tagger [111]. In order to either get enough event yields in each category or having higher signal significance, defined according to the probability that such an observation is due to statistical fluctuation of background events, to have a better observable binning and therefore a more stable fit, the BDT cut point, that needs to be optimised, is chosen at  $\text{BDT}=0.1$  [92].

As previously said, if the events fail to satisfy the VBF selection conditions they are categorised in the boost region. The major requirement is related to the Higgs Boson candidate, in particular to its transverse momentum ( $p_t(H)$ ) that has to be larger than 100 GeV; The Higgs transverse momentum is computed from the transverse momenta of all the visible decay products of both  $\tau$ -leptons plus the missing transverse momentum. The boost region is in turn split in boost *tight* and boost *loose*, where the boost *tight* includes events with a  $\Delta R \leq 1.5$  and  $p_t(H) > 140 \text{ GeV}$  while the boost *loose* events with  $\Delta R >$

1.5 and  $p_t(H) < 140$  GeV.

### 6.3.1 Decay mode and signal optimisation regions

A further split of production categories, VBF and ggF, is performed by selecting the tau decay modes. In particular, only the 1-prong case is considered in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel plus the combination of the corresponding di-tau decay modes such as l-1p0n, l-1p1n, l-1pXn, while 1-prong and 3-prong cases are taken into account in the full hadronic channel  $\tau_{\text{had}}\tau_{\text{had}}$  plus di-tau system decay mode combinations such as 1p0n-1p0n, 1p0n-1p1n, 1p1n-1p1n, 1p0n-1pXn, 1p1n-1pXn and 1p1n-3p0n.

An additional splitting in *high* and *low*, defined in table 6.3 and table 6.4, performed to get the highest maximum  $\Delta\text{NLL}$  value in the stat-only fit for the observable  $\phi_\tau$ , is computed by using kinematic observables correlated with the  $\phi_{CP}^*$  amplitude. In section 10.1 are shown the  $\phi_{CP}^*$  distributions, in high and low optimisation regions, calculated at preselection level, related to the  $0^\circ/90^\circ$  and  $40^\circ/140^\circ$  templates.

The analysis aim is to perform a shape-only measurement and therefore checking the distortion of the  $\phi_{CP}^*$  shape at various level gives more insight than calculating the efficiency. Given that the  $\phi_{CP}^*$  shape of the CP-even/odd signal and  $Z \rightarrow \tau\tau$  background are well determined at generator level. Z control region  $\phi_{CP}^*$  distributions, shown in section 10.2, give an hint on the distortion of the  $\phi_{CP}^*$  shape per decay mode combination in the preselection level. The  $\phi_{CP}^*$  distribution are then produced for the Signal Region (SR) and for the Z control region in section 10.3.

### 6.3.2 Fit Signal Regions

An important aspect to take into account while performing the fit with a huge number of signal regions, of which there are 48 in the  $\tau_{\text{had}}\tau_{\text{had}}$  channel and 24 in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel, is to avoid statistical fluctuations that could come up in many of the background templates of low statistics regions. Those issues are overcome during the fit by merging together all the high sensitivity regions in a single *high* region and similarly all the low sensitivity regions in a common *low* region. Merging the regions per sensitivity, corresponding to a total number of 16 signal regions, 8 related to the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel and 8 to the  $\tau_{\text{had}}\tau_{\text{had}}$  channel, could potentially minimize the fit instability and allow the use of a finer binned  $\phi_{CP}^*$  distribution with a higher sensitivity.



## 6.4 Control regions for $Z\tau\tau$ and multijet background

The Z control region for this analysis is defined as the  $Z \rightarrow \tau\tau$  region that corresponds to the di-tau system mass range between 60 GeV and 110 GeV. The mass of the binary system is, similarly to the Higgs to  $\tau\tau$  case, calculated by the missing mass calculator algorithm. The Z Control Region (ZCR) is divided in  $ZCR\_0$  and  $ZCR\_1$  but each event can fall randomly in any of the two regions according to a randomly assigned number, between 0 and 1. This trick is helpful because  $ZCR\_0$  is used to get the  $Z \rightarrow \tau\tau$  normalisation while  $ZCR\_1$  for estimating the  $\phi_{CP}^*$  shape uncertainty in the principal background events.

# SYSTEMATIC UNCERTAINTIES

7

The systematic uncertainties play an important role in the fit results and, for this reason, they have to be taken into account with a proper evaluation. This section will give an overview over all the systematic uncertainties by strictly following the combined performance group guidelines, with an additional focus on the RNN shape systematics and the role played by the different smoothing options applied in the fit step.

The major sources of uncertainties are deeply connected with all of the physics objects we have to reconstruct to efficiently calculate and fit the  $\phi_{CP}^*$ ; they are listed in experimental uncertainties chapter 7.1. In addition to that the other big sources such as luminosity, Pile-UP reweighting and Fake Backgrounds have to be studied. All the uncertainties are considered in the region  $\pm 1\sigma$ .

The experimental uncertainties are shown in section 7.1 while the theoretical uncertainties and the analysis-specific uncertainties are respectively shown in section 7.2 and section 7.3.

Dedicated studies are then performed to measure the RNN shape uncertainties in section 7.5, but also to calculate the Beam Spot (BS) constraint of the primary vertex in section 7.6 as well as the impact of different smoothing options in the final fit results in section 7.7.

## 7.1 Experimental Uncertainties

An important source of the systematic uncertainties is represented by the experimental uncertainties. Within this source are considered uncertainties related to the limited accuracy of the measuring apparatus but also from the limitation and simplifications of the experimental procedure.

Overall, MET and JET uncertainties will be discussed in detail due to their higher impact in the final fit of this analysis. Jet uncertainties are discussed in section 7.1.1, MET uncertainties in section 7.1.2, lepton uncertainties in section 7.1.3 and the luminosity uncertainties in section 7.1.4.

### 7.1.1 Jet uncertainties

The jet energy scale (JES) and resolution (JER) are determined in a multi step process, as shown in figure 7.1, starting with Monte Carlo simulations and comparing them to several measurements in data at various energy ranges, resulting in about 80 correlated uncertainties that are cooked down to some 40 nuisance parameters, designed to maintain correlations important for analyses [9]. In almost every calibration stage the full four-momentum is corrected, thereby adjusting the jet  $p_T$ , energy and scale.

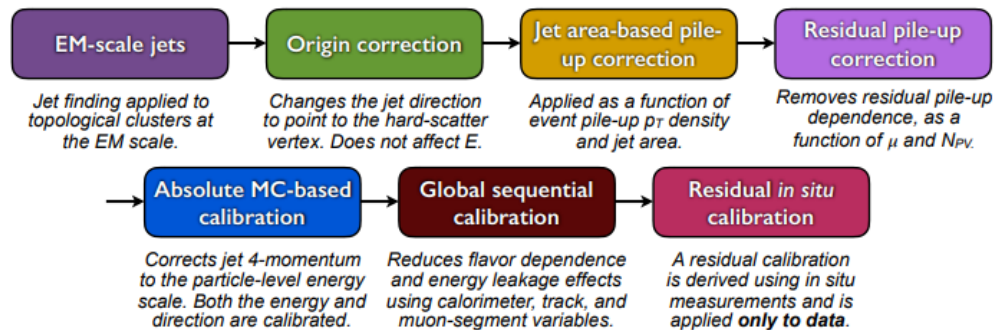


Figure 7.1: Calibration stages for EM-scale jets. Other than the origin correction, each stage of the calibration is applied to the four-momentum of the jet [9].

The origin correction recalculate the jet direction to point to the hard-scatter primary vertex in order to improve the  $\eta$  resolution of jets.

Reconstructed jets associated to a hard scatter vertex are corrected with a pile-up energy density derived per area in  $\eta$  and  $\phi$  and as a function of  $p_T$  and number of primary vertices  $N_{PV}$ . Four systematic uncertainties are related to the  $N_{PV}$  (in-time pile-up), the instantaneous luminosity  $\mu$  (out-of time pile-up), the per-event median energy density  $\rho$ , and the  $p_T$  dependence of the residual pile-up correction derived from comparing reconstructed jets and truth jets [9].

In the next step the reconstructed jet 4-momentum is corrected to particle level energy scale, providing a directional correction in  $\eta$  and setting an absolute energy scale. This correction mainly accounts for hardware imperfections and transitions in granularity.

While previous steps focused on dependencies in  $\eta$ , the following Global Sequential Calibration (GSC) corrects for differences in the longitudinal shower development. In this step differences between quark and gluon initiated jets, particle composition and energy distribution within a jet, are calibrated. Apart from variables measuring the longitudinal shower development, tracks are used as well in this step. The 2011 calibration procedure, described in Ref. [165], identifies five observables, listed in Ref. [9], that enhance the resolution of the JES.

As a final step in-situ calibrations using various physics processes are used to correct for remaining data and MC differences. Di-jet events,  $Z/\gamma$ +jet events and Multi-jet Balance (MJB) events cover a wide energy range from about 30 GeV to 1.2 TeV and the forward region up to  $|\eta| < 4.5$ .

67 variables modelling the systematic uncertainty are coming from the Z+jet and MJB in-situ calibration. Three uncertainties each are due to  $\eta$  inter calibration, jet flavour and composition, and GSC response, resulting in a total of 80 uncertainties. Typical uncertainties range from 4.5% at low  $p_T$  to 1% at 200 GeV. An eigen-decomposition results in a set of 19 nuisance parameters describing jet energy uncertainties, which is used in this analysis.

### 7.1.2 Missing transverse energy uncertainties

The techniques used to estimate the MET uncertainties are described in [10] and include the sources listed in table 7.1. Every reconstructed physics objects undergoes a specific calibration process, which converts the detector signals into a completely corrected four-momentum measurement. In each event, the hard component of the missing transverse energy Missing Transverse Energy ( $E_T^{miss}$ ) is calculated based on the selected muons, electrons, photons, hadronically decaying tau, and jets. On the other hand, the soft component poses a necessary but challenging aspect of  $E_T^{miss}$  reconstruction. It comprises all the detector signals that do not correspond to the previously defined reconstructed objects. This component can include contributions from the hard scattering process, as well as pile-up. This analysis uses the uncertainty on the soft term. To calculate the uncertainty three quantities illustrated in figure 7.2 are calculated:

**The parallel scale** ( $\Delta_L$ ) the mean value of the parallel projection of  $p_T^{soft}$  along  $p_T^{hard}$ .

**The parallel resolution** ( $\sigma_{\parallel}$ ) the root-mean-square of  $p_{\parallel}^{soft}$ .

**The transverse resolution ( $\sigma_{\perp}$ )** The root-mean-square of the perpendicular component of  $p_T^{soft}$  with respect to  $p_T^{hard}$  defined as  $p_{\perp}^{soft}$ .

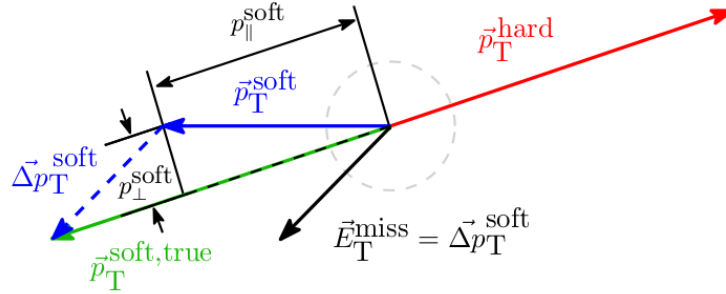


Figure 7.2: Illustration of the track-based soft term projections with regard to  $p_T^{hard}$  for the systematic uncertainty calculation, ref. [10].

The results, presented in Ref. [10], show a  $p_T$  dependent systematic uncertainty between 1.8 and 5 GeV for a  $0 < p_T < 100$ .

<b>MET Resolution</b>	MET-SoftTrk-ResoPara, MET-SoftTrk-ResoPerp
<b>MET Soft track energy scale</b>	MET-SoftTrk-Scale

Table 7.1: MET uncertainties list.

### 7.1.3 Lepton uncertainties

Systematic uncertainties affect the reconstruction of leptons, such as  $e$ ,  $\mu$ ,  $\tau$ . Systematic uncertainties belonging to  $\tau$  particles are well studied in detail. With the introduction of the full leptonic channel in the fit stage, in addition to the hadronic one, the reconstruction of muons and electrons needs to be more precise.

Those uncertainties take into account different source of errors, in particular energy resolution, energy scale, uncertainties in the reconstruction and identification of the object as well as trigger related uncertainties.

The  $\tau$  uncertainties used in this thesis are the 2019-summer set. They include similar sources to both electron and muon uncertainties, with additional sources on the eVeto on fake  $\tau_{had-vis}$ . Section 11.1 provides a list of the systematic uncertainties employed in this analysis for every lepton.

### 7.1.4 Luminosity uncertainties

Luminosity uncertainties have to be calculated for the HIGGS signal and the simulated background not constrained in by Control Region (CR) in data. The value measured by the LUCID-2 [166] detector for 2015-1028 is 1.7% [167].

## 7.2 Theoretical Uncertainties

Most of the theoretical uncertainties, belonging to the signal  $H \rightarrow \tau\tau$  and to the main background  $Z \rightarrow \tau\tau$ , are inherited from the full Run-2  $H \rightarrow \tau\tau$  analysis, as described in [92].

When considering the uncertainties described in table 7.2, which are estimated before pruning procedure, only the shape and migration components are taken into account as the final absolute normalization values are determined during the fit.

Regarding background processes, the choice to consider only the theoretical uncertainties of  $Z \rightarrow \tau\tau$ +jets and  $Z \rightarrow \text{ll}$ +jets samples is justified from the fact that theory uncertainties belonging to other background processes are expected to be irrelevant.

Therefore, the main source of theory uncertainties signal prediction are:

**QCD scale uncertainties** High perturbative orders missing in the calculation.

**QCD scale uncertainties** Non perturbative orders of the calculation.

**Experimental input parameters** An example is the PDFs as well as the  $\alpha_s$ .

While, the main sources of uncertainties taken into account for both  $Z \rightarrow \tau\tau$ +jets and  $Z \rightarrow \text{ll}$ +jets samples are:

**PDF central value** evaluated considering the standard deviation of the 100 NNPDF replicas event weights of NNPDF3.0nnlo PDF set used in Sherpa.

**renormalization and factorization scales** evaluated using Sherpa.

**ckkw** jet-to-parton matching uncertainty, evaluated using truth-level parameterisation as a function of jet multiplicity and  $p_T$  ( $Z$ ).

**qsf** resummation scale uncertainty, evaluated using truth-level parameterisation as a function of jet multiplicity and  $p_T$  ( $Z$ ).

$\alpha_s$  evaluated using event-weights provided by Sherpa.

**PDF alternative value** evaluated comparing predictions from NNPDF3.0nnlo PDF set (nominal) with MMHT2014nnlo68cl PDF set [168].

Moreover, the analysis makes use of different normalisation factors for each production cross section such as ggH, VBF and VH and therefore the respective theoretical uncertainties are not expected to impact the fit result.

A comparison between the nominal event generator and alternative event generators is performed to estimate the MC modelling uncertainties.

Nuisance Parameter (NP) name and description
<i>Higgs</i> $\rightarrow \tau\tau$
<i>theory_sig_qcd</i> _[0 – 8]_ggH LHCHSWG scheme for perturbative uncertainties on ggH
<i>theory_VBFH_mur_muf_envelope</i> Renormalisation and factorisation scales $\mu_R/\mu_F$ acceptance uncertainty for VBFH
<i>theory_VH_mur_muf_envelope</i> Renormalisation and factorisation scales $\mu_R/\mu_F$ acceptance uncertainty for VH
<i>theory_ttH_mur_muf_envelope</i> Renormalisation and factorisation scales $\mu_R/\mu_F$ acceptance uncertainty for ttH
<i>theory_sig_AlphaS</i> $\alpha_S$ acceptance uncertainty for the signal
<i>theory_sig_pdf</i> _[0 – 29] PDF acceptance uncertainties for the signal (PDF4LHC)
<i>Higgs</i> $\rightarrow \tau\tau$
<i>theory_z_AlphaS</i> $\alpha_S$ acceptance uncertainty, evaluated using event-weights provided by Sherpa
<i>theory_z_alt_pdf_envelope</i> Comparison of predictions from alternative PDF sets
<i>theory_z_pdf_envelope</i> Standard deviation of the 100 NNPDF replicas event weights of NNPDF3.0nnlo PDF set in Sherpa
<i>theory_z_mur_muf_envelope</i> Renormalisation and factorisation scales $\mu_R/\mu_F$ uncertainty, evaluated using event-weights provided by Sherpa
<i>theory_z_ckkw</i> Jet-to-parton matching uncertainty, evaluated using truth-level parameterisation as a function of jet multiplicity and $p_v(Z)$
<i>theory_z_qsfy</i> Resummation scale uncertainty, evaluated using truth-level parameterisation as a function of jet multiplicity and $p_v(Z)$
<i>theory_z_CT14_pdfset</i> Comparison of predictions from NNPDF3.0nnlo PDF set (nominal) with CT14nnlo PDF set
<i>theory_z_MMHT_pdfset</i> Comparison of predictions from NNPDF3.0nnlo PDF set (nominal) with MMHT2014nnlo68cl PDF set

Table 7.2: Theoretical uncertainties.

### 7.3 Analysis-specific uncertainties

The whole set of systematic uncertainties used in this analysis included the uncertainties described so far which are mostly in common with the Full Run-2  $H \rightarrow \tau\tau$  cross-section analysis plus a set of specific systematic uncertainties, specifically developed for this analysis, that are directly linked to the main observable  $\varphi_{CP}^*$ , such as the track uncertainties on the final state reconstructed objects, uncertainties on the angular resolution of the reconstructed  $\pi^0$  and theoretical uncertainty on the  $\varphi_{CP}^*$  shape needed to consider the effect of the TauSpinner reweighting, as well as the decay mode classification of the di- $\tau$  system. Most systematics are shape systematics, meaning that the  $\pm 1\sigma$  systematics found by the Combined Performance ATLAS groups result in a shape variation in  $\varphi_{CP}^*$  with a  $\pm 1\sigma$  envelope. This envelope is propagated into the final fit with a nuisance parameter as variable scale factor. The impact of the most important systematics is shown in figures 8.13, while in the following table 8.4 the names of the nuisance parameters and their relation to the systematic uncertainty are listed.

#### Tau decay mode classification uncertainties

The uncertainties developed to account for the different decay mode classifications are based on the Run-1 strategy [4] where the relative decay mode classification efficiencies are calculated using data to extract the scale factors as well as the uncertainties built to match the MC distribution with data. Scale factors are obtained, for each decay mode and for all the tau ID working point to take into account the whole set of scenarios, but also to account for the migration of nearby decay modes, see table 7.3.

Systematic uncertainties	Unc. (%)
<i>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_1P0N_RECO_1P0N_TOTAL</i>	5.8
<i>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_1P1N_RECO_1P0N_TOTAL</i>	19.5
<i>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_1P1N_RECO_1P1N_TOTAL</i>	11.2
<i>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_1P1N_RECO_1PXN_TOTAL</i>	8.1
<i>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_1PXN_RECO_1P1N_TOTAL</i>	1.9
<i>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_1PXN_RECO_1PXN_TOTAL</i>	8.3
<i>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_3P0N_RECO_3P0N_TOTAL</i>	8.4
<i>TAUS_TRUEHADTAU_EFF_JETID_DECAY_MODE_TRUE_3PXN_RECO_3P0N_TOTAL</i>	1.9

Table 7.3: Tau decay mode classification uncertainties.



---

 **$\varphi_{CP}^*$  systematic uncertainties**

This analysis makes use of the track uncertainties, see table 7.4, generated during the smearing process of all reconstructed objects in both semi-leptonic and full hadronic channels, and they are estimated following the Tracking Combined Performance (CP) ATLAS group recommendation [169]. Those systematics are applied to consider the difference between data and MC.

Moreover, a set of uncertainties in table 7.5 accounting for the reconstructed  $\pi^0$  angular resolutions and energy scale are considered in the fit.

Eventually, the TauSpinner reweighting process produces another source of uncertainty (PHISTAR-THEORY-SHAPE).

---



---

```

TRK_BIASD0_WM
TRK_BIAS_Z0_WM
TRK_BIAS_QOVERP_SAGITTA_WM
TRK_RES_D0_DEAD
TRK_RES_D0_MEAS
TRK_RES_Z0_DEAD
TRK_RES_Z0_MEAS

```

---

Table 7.4:  $\varphi^*$  Track uncertainties.

---



---

```

TAUS_PI0_RECALC_PHI_STAR_ETA_SMEAR
TAUS_PI0_RECALC_PHI_STAR_PHI_SMEAR
TAUS_PI0_RECALC_PHI_STAR_ENERGY_SCALE

```

---

Table 7.5:  $\pi$  reconstruction uncertainties.

 **$\tau_{lep}\tau_{had}$  - Fake Systematic Uncertainties.**

The sources of systematic uncertainties considered in this analysis, arising from the background estimate using the Fake-Factor method, are:

1. The uncertainty on  $F_{QCD}$  and on  $F_W$  due to low statistics.
2. The uncertainty on the  $R_{QCD}$  estimation with contribution from different sources.
3. MC subtraction term uncertainty.

In closure test the same fake estimation method is applied in same sign region, and take the discrepancy between Data and prediction as closure uncertainty.

### $\tau_{\text{had}}\tau_{\text{had}}$ - Fake Systematic Uncertainties.

Four different sources of uncertainties are considered for the fake  $\tau_{\text{had}}\tau_{\text{had}}$  data-driven method.

1. Statistical uncertainty on the W+jets fake-factor.
2. MC subtraction uncertainty in the fake-factor calculation.
3. Uncertainty on the limitation of the fake-factor parametrization.
4. Uncertainty on the closure of the fake estimation performed with the fake-factor derived in the  $\tau_{\text{had}}\tau_{\text{had}}$  high- $\Delta\eta$  region.

## 7.4 Data-driven fake estimation uncertainties

Another source of uncertainties is introduced in the fit to treat the fake  $\tau$  estimations and its contribution can be split in the semi-leptonic and fully hadronic contribution.

The uncertainty sources considered for the  $\tau_{\text{lep}}\tau_{\text{had}}$  include the statistical uncertainties on the fake-factor measurement, statistical uncertainty on the extrapolation of the ratio of QCD contributions in different signal and control regions (RQCD) and the relative Isolation Factor used for the same derivations, uncertainties on MC-subtraction for fake-factor measurements as well as Isolation Factors and RQCD derivations, and eventually closure uncertainties extrapolated from the same-sign region.

Table 7.6 7.7 summarize the uncertainties on fake estimation for both  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  channels, where the uncertainties are shown in percent difference compared to the nominal fake estimation.

	Boost				VBF			
	$\ell$ -1p0n	$\ell$ -1p1n	$\ell$ -1pXn	$\ell$ -3p0n	$\ell$ -1p0n	$\ell$ -1p1n	$\ell$ -1pXn	$\ell$ -3p0n
lh fake closure 1p0n	10.8				100.2			
lh fake closure 1p1n		-4.4				-27.0		
lh fake closure 1pXn			16.6				1.9	
lh fake closure 3p0n				13.3				-15.8
lh fake FF QCD 1p0n stat	1.5				2.3			
lh fake FF QCD 1p1n stat		1.1				2.9		
lh fake FF QCD 1pXn stat			1.8				3.1	
lh fake FF QCD 3p0n stat				2.4				4.1
lh fake FF W 1p0n stat	10.7				21.6			
lh fake FF W 1p1n stat		7.5				19.4		
lh fake FF W 1pXn stat			12.4				26.0	
lh fake FF W 3p0n stat				13.9				26.5
lh fake msubstr 1p0n	-11.0				-9.3			
lh fake msubstr 1p1n		-11.9				-16.2		
lh fake msubstr 1pXn			-9.8				-7.3	
lh fake msubstr 3p0n				-10.3				-8.6
lh fake RQCD 1p0n stat	0.3				-0.2			
lh fake RQCD 1p1n stat		0.1				0.8		
lh fake RQCD 1pXn stat			-0.1				-0.5	
lh fake RQCD 3p0n stat				0.0				-0.1
Total (systematic)	19	15	23	22	103	37	27	32
Statistical	1.1	0.6	0.8	0.5	3.2	1.5	2.3	1.6

Table 7.6: Systematic uncertainties in the fake-factor method used for the  $\tau_{\text{ep}}\tau_{\text{had}}$  channel by decay mode regions in the Boost and VBF SRs (where the fake estimation is derived in each decay mode region and separately in Boost or VBF regions). The uncertainties are presented in percent difference of the varied and nominal values of the fake estimation. The total systematic uncertainty is calculated assuming all components are uncorrelated. The statistical uncertainty of the fake yield is also shown in the last row for comparison. (The size of the Isolation Factor systematics in the  $\tau_{\text{ep}}\tau_{\text{had}}$  channel are not present in the table but their sizes are expected to be very small [6].)

	1p0n-1p0n	1p0n-1p1n/1p1n-1p0n	1p1n-1p1n	1p0n-1pXn/1pXn-1p0n	1p1n-1pXn/1pXn-1p1n	1p1n-3p0n/3p0n-1p1n
hh fake ff composition	-18.2	-17.4	-13.6	-9.7	-9.3	-10.6
hh fake ff param	-0.9	-0.9	-0.9	-0.9	-0.9	-0.9
hh fake ff stat 1p0n lnm	-11.2	-5.5		-4.7		
hh fake ff stat 1p0n nm	12.0	6.2		5.5		
hh fake ff stat 1p1n lnm		-4.4	-7.5		-4.0	-4.0
hh fake ff stat 1p1n nm		3.8	8.2		4.0	4.0
hh fake ff stat 1pXn lnm				-6.1	-5.5	
hh fake ff stat 1pXn nm				5.7	5.9	
hh fake ff stat 3p0n lnm						-9.7
hh fake ff stat 3p0n nm						9.2
hh fake ff mcsubstr 1p0n nm	-8.8	-4.6		-4.1		
hh fake ff mcsubstr 1p0n lnm	6.3	3.1		2.7		
hh fake ff mcsubstr 1p1n nm		-5.1	-11.0		-5.3	-5.4
hh fake ff mcsubstr 1p1n lnm		4.2	7.1		3.7	3.8
hh fake ff mcsubstr 1pXn nm				-3.9	-4.0	
hh fake ff mcsubstr 1pXn lnm				3.1	2.8	
hh fake ff mcsubstr 3p0n nm						-6.1
hh fake ff mcsubstr 3p0n lnm						5.0
Total (systematic)	27	22	22	16	16	21
Statistical	32	13	1.3	2.5	1.3	1.7

Table 7.7: Systematic uncertainties in the fake-factor method used for the  $\tau_{\text{had}}\tau_{\text{had}}$  channel by regions of decay mode combination. The fake factors used in the  $\tau_{\text{had}}\tau_{\text{had}}$  channel are derived from the ones from the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel. The uncertainties are presented in percent difference of the varied and nominal values of the fake estimation. The total systematic uncertainty is calculated assuming all components uncorrelated. The statistical uncertainty of the fake yield is also shown in the last row for comparison [6].

## 7.5 RNN shape uncertainties

The effects of tau identification on the  $\phi^*$  shape were studied in detail for this analysis. The tau identification procedure is based on an RNN classifier [89], employing information from the reconstructed charged-particle tracks and the clusters of energy in the calorimeter, associated to the  $\tau_{had}$ -vis candidates.

The dependence of the  $\phi^*$  shape on the RNN score can be estimated using  $Z \rightarrow \tau\tau$  events. Systematic uncertainties based on the differences between data and MC can be extracted comparing data-driven and  $Z_{MC} \rightarrow \tau\tau$   $\phi^*$  shape in ZCR region. To check that the MC predictions in the SR and the ZCR are compatible, the ratio of  $\phi^*$  distributions in  $Z_{MC} \rightarrow \tau\tau$  events in the ZCR region and  $Z_{MC} \rightarrow \tau\tau$  events in the SR is also constructed.

The Z control region is divided into Boost tight, Boost loose and VBF regions, where each of them is sub-divided into High and Low. All channels from the extended scenario of signal regions were added together for the purposes of this analysis. Based on the RNN scores of the two taus, each region is split into 4 different RNN bins using the criteria shown in Table 7.8.

	$\tau_0$	$\tau_1$
RNN bin 11	$nTrks^\pm = 1 \ \& \ 0.25 < RNNs < 0.65$ $nTrks^\pm = 3 \ \& \ 0.4 < RNNs < 0.7$	$nTrks^\pm = 1 \ \& \ 0.25 < RNNs < 0.65$ $nTrks^\pm = 3 \ \& \ 0.4 < RNNs < 0.7$
RNN bin 12	$nTrks^\pm = 1 \ \& \ 0.25 < RNNs < 0.65$ $nTrks^\pm = 3 \ \& \ 0.4 < RNNs < 0.7$	$nTrks^\pm = 1 \ \& \ RNNs > 0.65$ $nTrks^\pm = 3 \ \& \ RNNs > 0.7$
RNN bin 21	$nTrks^\pm = 1 \ \& \ RNNs > 0.65$ $nTrks^\pm = 3 \ \& \ RNNs > 0.7$	$nTrks^\pm = 1 \ \& \ 0.25 < RNNs < 0.65$ $nTrks^\pm = 3 \ \& \ 0.4 < RNNs < 0.7$
RNN bin 22	$nTrks^\pm = 1 \ \& \ RNNs > 0.65$ $nTrks^\pm = 3 \ \& \ RNNs > 0.7$	$nTrks^\pm = 1 \ \& \ RNNs > 0.65$ $nTrks^\pm = 3 \ \& \ RNNs > 0.7$

Table 7.8: RNN Bins criteria.

For each RNN score bin and for each region the following distributions are produced:

- $\phi^*$  shape for  $Z_{MC} \rightarrow \tau\tau$  in the ZCR
- $\phi^*$  shape for  $Z_{MC} \rightarrow \tau\tau$  in the signal region (SR)
- Data-driven  $\phi^*$  shape for  $Z \rightarrow \tau\tau$

The  $\phi^*$  distributions are sampled in 3 bins in the cases of Boost loose and Boost tight regions and in 2 bins, due to statistical limitations, in the VBF region. The  $\phi^*$  distributions in the  $Z_{MC} \rightarrow \tau\tau$  SR and ZCR and the data-driven ZCR are shown in figure 7.3 for Boost loose, while Boost tight and VBF channels are shown in section 11.3.

The  $Z_{MC} \rightarrow \tau\tau$  ZCR/SR (Z peak/SR) does not display significant discrepancies in the Boost loose and Boost tight cases, while the discrepancies in the VBF region are mostly covered by the statistical uncertainties.

The ratio of data-driven  $Z \rightarrow \tau\tau$  and  $Z_{MC} \rightarrow \tau\tau$  in the ZCR region is used to estimate a systematic uncertainty on the  $\phi^*$  shape. The size of the systematic uncertainty on the  $\phi^*$  distribution in the signal region is shown in Figure 7.4.

Those uncertainties are not considered in the final fit for mainly two reasons. The first one is that at preselection level they appeared to be small and therefore they are not impacting the final fit results, while the second reason is due to the double counting of many other uncertainties such as the mis-modelling, tracks uncertainties,  $\pi^0$  uncertainties, cluster and all the other uncertainties being used in the production of the RNN score.

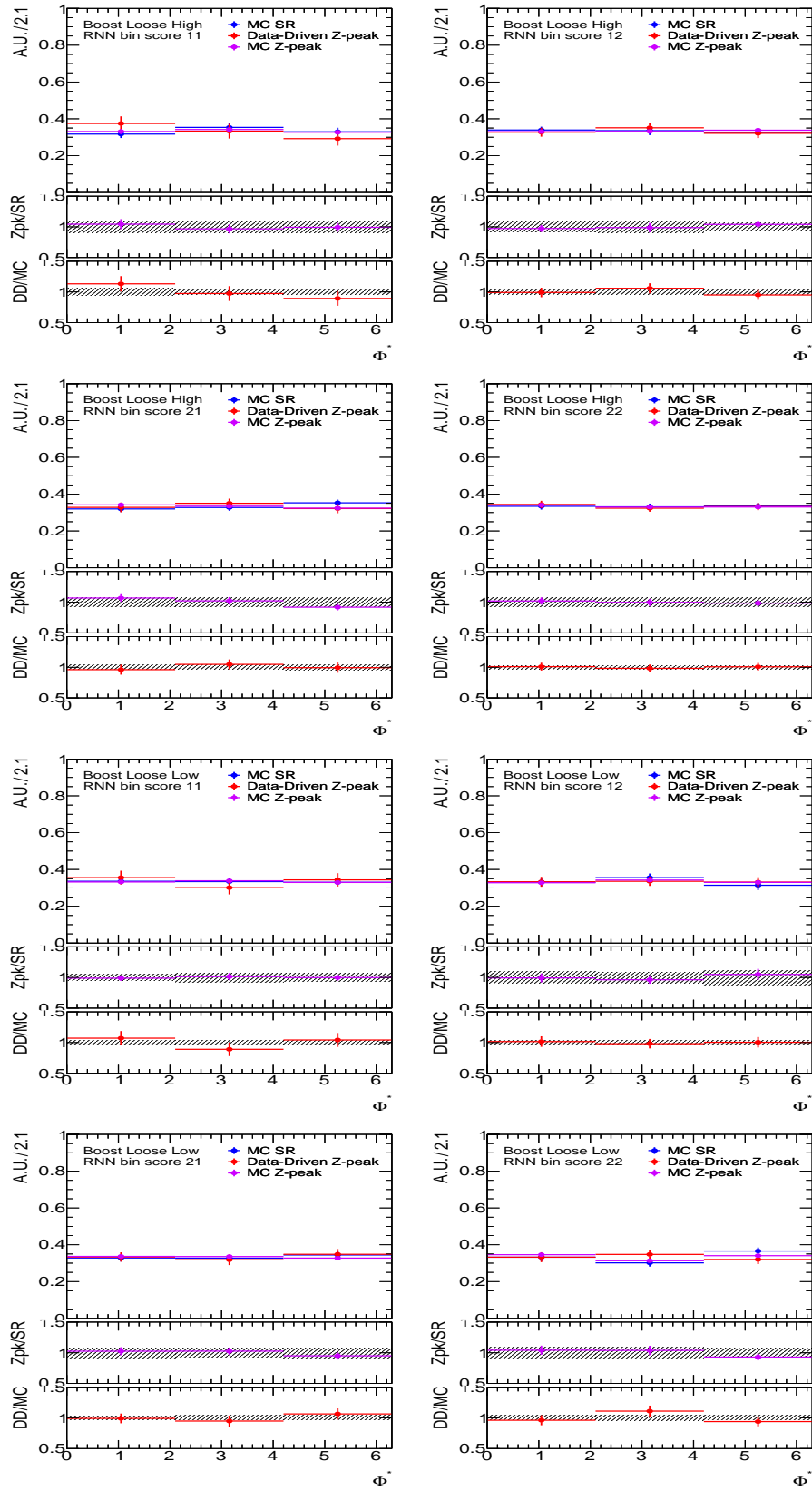


Figure 7.3: Boost Loose  $\phi^*$  shape distributions shown for the different RNN bins.

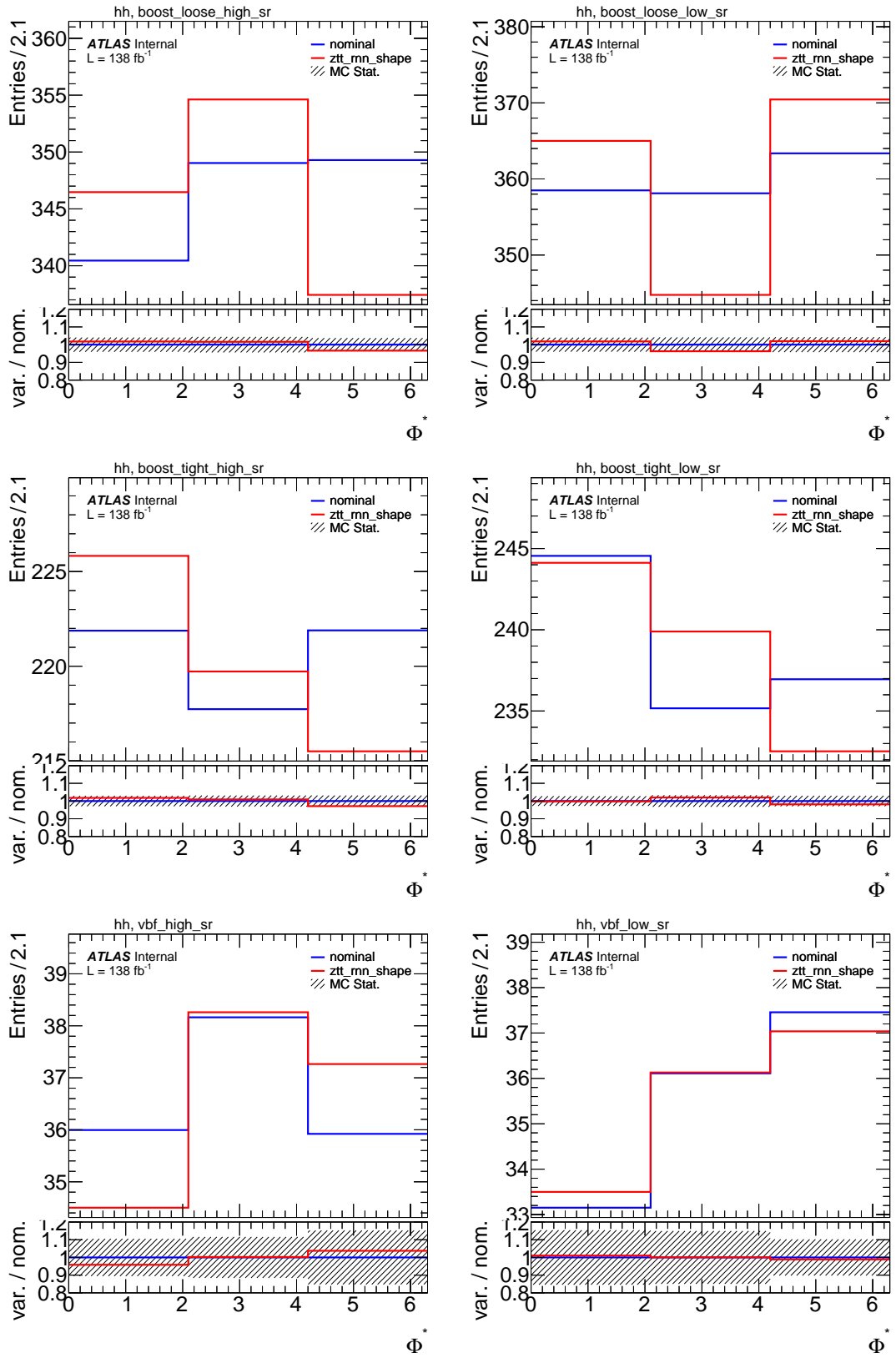


Figure 7.4: The 3 bins  $\phi^*$  envelopes and the " $ztt\_rnn\_shape$ " systematic uncertainty are shown for the different signal regions.



## 7.6 Primary vertex uncertainties

Variations of the beam spot (BS) constraint applied in the calculation of the primary vertex can lead to another source of systematic uncertainty that potentially could affect the  $\varphi^*$  distribution and therefore dedicated studies were done to check if the primary vertex (PV) used to calculate the impact parameter vector, as well as the  $\varphi^*$  calculation, needed to be revised.

The PV reconstruction is performed using tools called vertex finders. The reconstruction of the vertices [170] is performed in two steps. The first one, called vertex finding, associates tracks to a particular vertex, while the second, known as vertex fitting, performs the reconstruction of the vertex position with its covariance matrix as well as an estimation of the fit quality. A beamspot constraint is applied by default when performing the vertex fit in ATLAS.

The aim of the study is to estimate the effect of changing the BS constraint on the primary vertex position and therefore on the main fit observable  $\varphi^*$ . Three different setups were tested on MC events, the first one concerning the default BS, the second one concerning the scenario where the BS is removed and the last one concerning a softened BS, reduced by 55% of the default one. A comparison with the MC distributions is also performed.

This study makes use of four additional MC samples, VBF CP-even and CP-odd and similarly ggH CP-even and CP-odd, generated using Pythia8 with angular decay correlations in the decay of the Higgs boson provided by Pythia8. The samples were used for cross-check purposes and have much smaller statistics than the signal MC samples described in section 5.

The beamspot is a representation of the 3D profile of the luminous region, where the LHC beams collide in the ATLAS detector. Its reconstruction is based on an unbinned maximum-likelihood fit to the spatial distribution of primary vertices collected from many events [170] and is updated several times during a run. The uncertainties in the beamspot parameters are small and are incorporated in the primary vertex fitting algorithms to provide an additional constraint on the PV position.

All MC samples are produced with a beamspot at  $b = (x_b, y_b, z_b) = (-0.5, -0.5, 0)$  and a covariance matrix:

$$c_b = \begin{bmatrix} 0.01^2 & 0 & 0 \\ 0 & 0.01^2 & 0 \\ 0 & 0 & 42^2 \end{bmatrix} = \begin{bmatrix} 0.0001 & 0 & 0 \\ 0 & 0.0001 & 0 \\ 0 & 0 & 1764 \end{bmatrix} \quad (7.1)$$

A covariance weighted estimate of the fitted vertex without the BS constraint and

a subsequent application of a weaker BS constraint is used for these studies. If  $v = (x_v, y_v, z_v)$  and  $c_v$  are the position and covariance of a given primary vertex, obtained from the default PV reconstruction, the corresponding vertex position,  $d = (x_d, y_d, z_d)$  without a BS constraint and its covariance matrix  $c_d$  can be obtained using:

$$d = c_d(v c_v^{-1} - b c_b^{-1}), \quad (7.2)$$

where

$$c_d^{-1} = c_v^{-1} - c_b^{-1} \quad (7.3)$$

Similarly, to obtain a primary vertex position with a softened beamspot constraint  $v'$  and its covariance  $c_{v'}$ , the following transformations are applied:

$$v' = c_{v'}(d c_d^{-1} + b c_b^{-1}), \quad (7.4)$$

where

$$c_{v'}^{-1} = c_d^{-1} + c_b^{-1} \quad (7.5)$$

and

$$c_{b'} = \begin{bmatrix} 0.01^2 \times f^2 & 0 & 0 \\ 0 & 0.01^2 \times f^2 & 0 \\ 0 & 0 & 42^2 \times f^2 \end{bmatrix} \quad (7.6)$$

The factor for softening the BS constraint used for these studies is chosen to be  $f = 1.55$ . This results in closer shapes of the distributions of the vertex positions to the MC truth distributions, while the default PV distributions show a bias (squeezing in the transverse plane) and much broader distributions in the case of no-BS constraint as shown, for the VBF even sample, in figure 7.5 and in section 11.4 for the VBF odd sample as well as ggH even and odd samples.

The residual distributions on the other hand, shown in figure 7.6 for the VBF even sample and in section 11.5 for the other samples, show marginally better resolution for the default PV, compared to the one with a softer BS constraint. The fit is to a single Gauss function in the range  $(-0.02, +0.02)$  in the x- and y-coordinates and  $(-0.06, +0.06)$  in the z-coordinate to include the core of the distribution only. The worst resolution is still in the case of no BS constraint. The z-coordinates are not affected, as expected. 2D distributions were also considered: Figure 7.7 for the VBF even sample shows the distribution of the reconstructed vertex coordinates and the corresponding residuals vs the MC truth coordinates. All the other sample results are shown in section 11.6. The z-coordinate shows no biases in all 3 cases, while the x-coordinate (similarly for y, not shown) suffers from some biases - larger distances to the z-axis result in a larger squeeze of

---

the vertex position in the transverse plane. The smallest bias in the x-residual distribution is observed in the case of no-BS constraint, which gives an indication that the improved vertex resolution in x and y is at the cost of a bias in the reconstructed vertex position. This makes it difficult to predict how the choice of BS constraint would affect the reconstruction of the CP-sensitive angle  $\varphi^*$ . Hence the three BS setups were also used in the  $\varphi^*$  calculation and compared to the MC truth value.

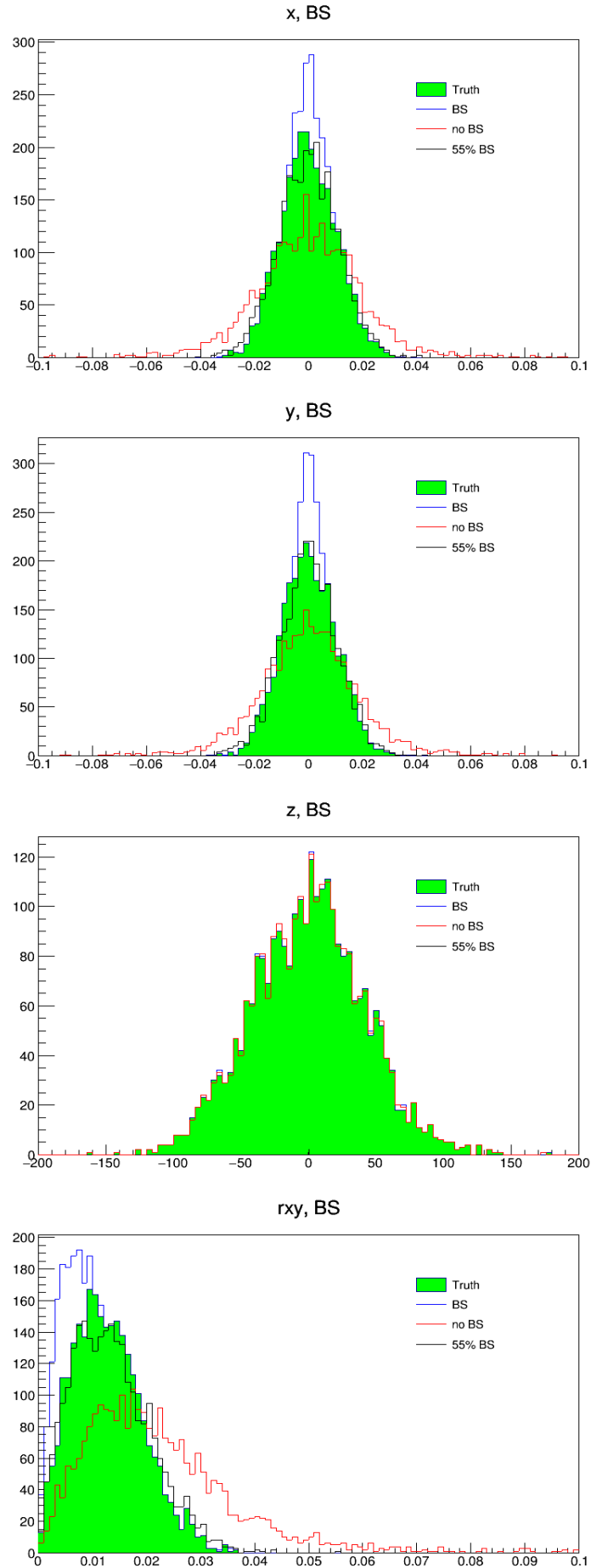


Figure 7.5: PV coordinates ( $x$ ,  $y$ ,  $z$ ) and the distance in the detector transverse plane ( $R_{xy}$ ) for the VBF CP-even sample. Comparison between the default BS constraint, no BS constraint and 55% of the BS constraint is performed.

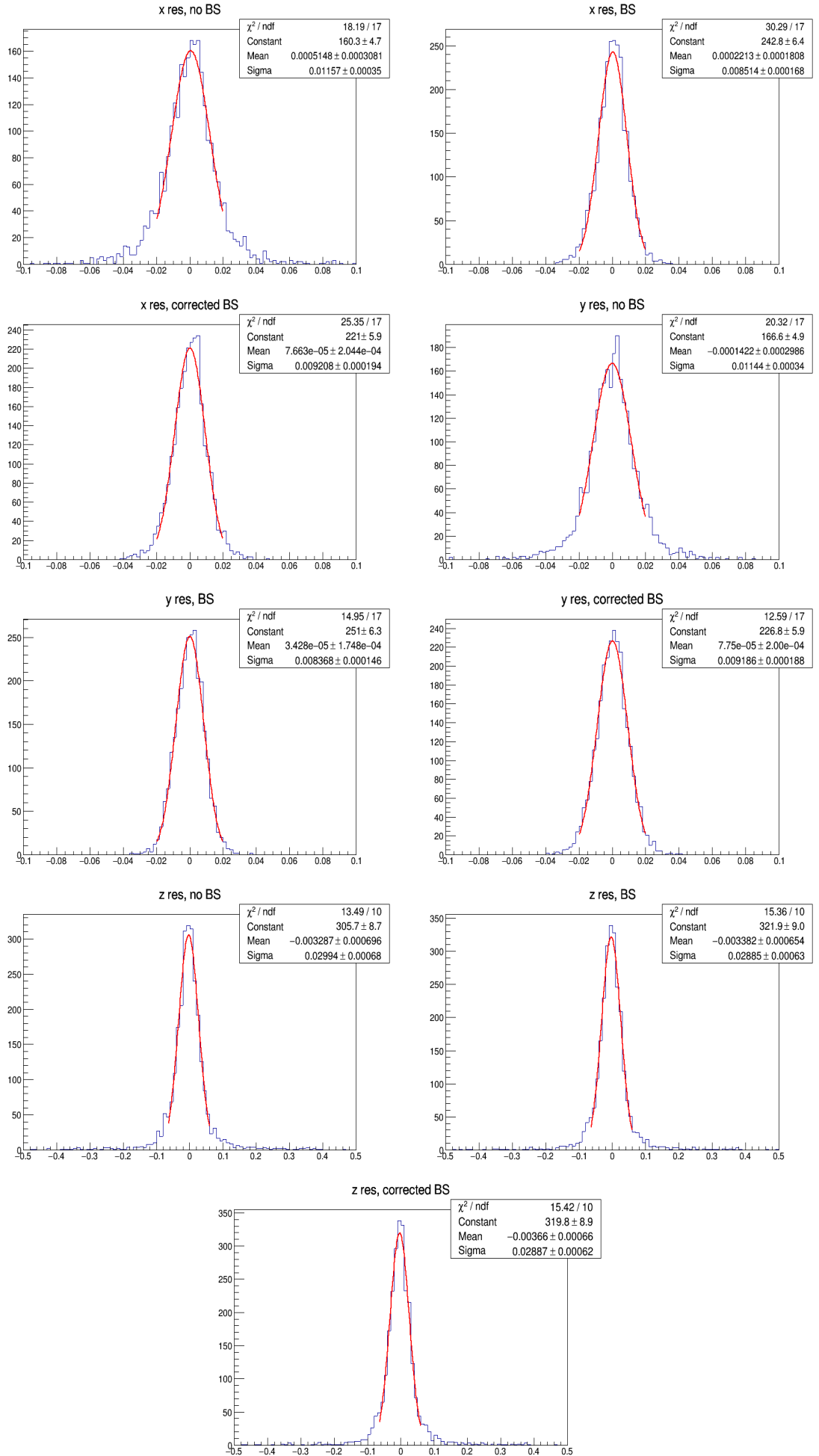


Figure 7.6: Coordinate residual plots of the VBF CP-even sample.

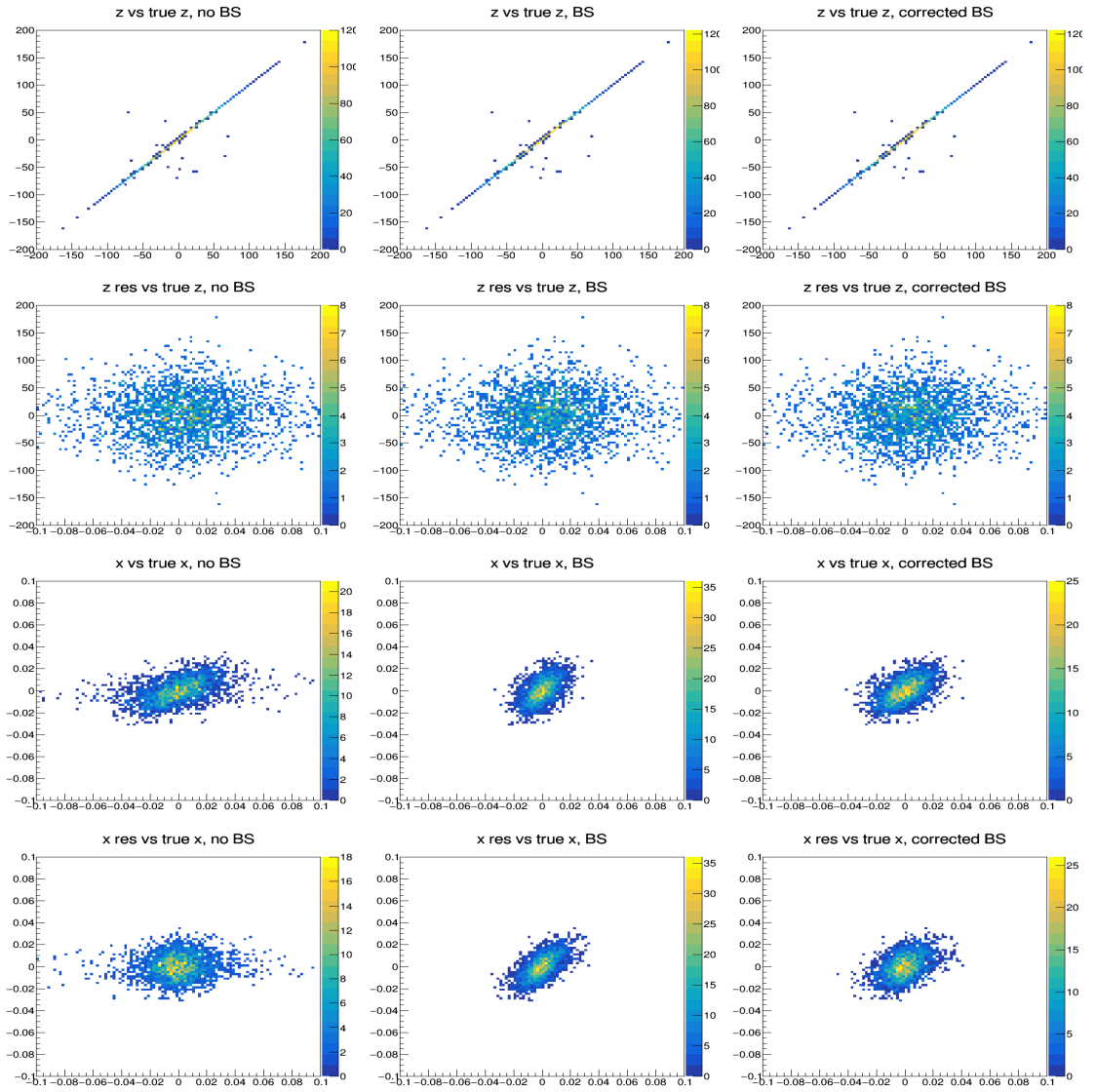


Figure 7.7: First row shows true  $z$  position (y-axis) vs  $z$  position (x-axis) plots, second row shows true  $z$  position (y-axis) vs  $z$  residual (x-axis) plots, third row shows true  $x$  position (y-axis) vs  $x$  position (x-axis) plots and fourth row shows true  $x$  position (y-axis) vs  $x$  residual (x-axis) plots. From left to right the plots correspond to the noBS case, default case and soft BS case. All of them are calculated using the VBF even sample.

The first step in the  $\varphi^*$  calculation, in the case of the IP-IP method, is to construct the IP vector when considering a reconstructed PV. It is obtained by the following formula

$$\mathbf{IP} = ip + k \cdot \pi_{4p} - PV, \quad (7.7)$$

where the parameters used in the formula are:

$ip = (x_{MC}^\pi, y_{MC}^\pi, z_{MC}^\pi)$  if using MC truth, where  $x_{MC}^\pi, y_{MC}^\pi, z_{MC}^\pi$  are the coordinates of the  $\pi$  production ( $\tau$  decay) vertex;

$\pi_{4p}$  = four-momentum of the pion;

$k = (PV - ip) \cdot \hat{\pi}_{4p}$ , where  $\hat{\pi}_{4p}$  is the unit vector.

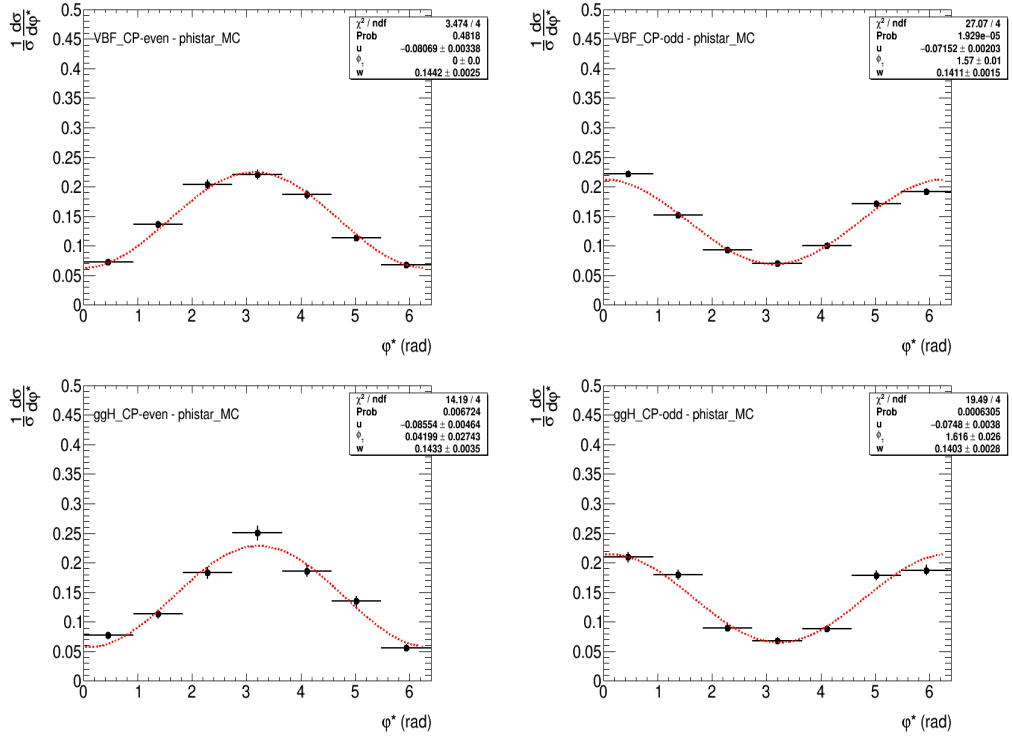
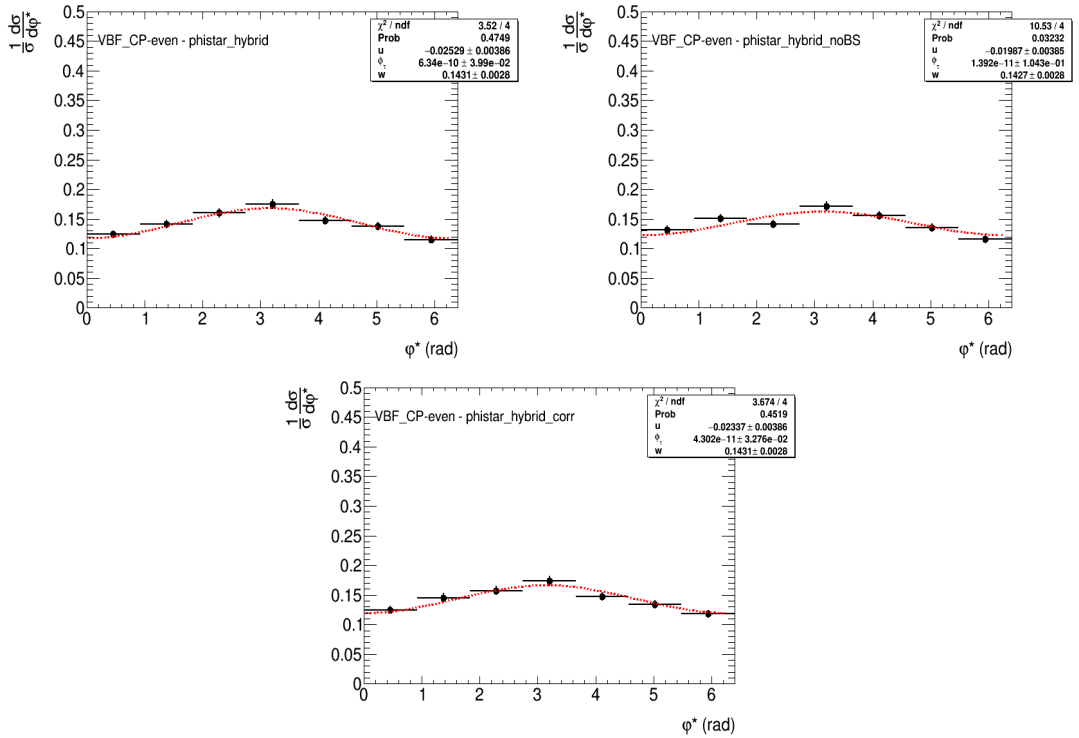
The pion four-momentum is obtained from the MC record, while the PV is the reconstructed primary vertex position (Hybrid method). This allows one to make an estimate of the effect due to the PV only, without the additional effects of momentum smearing. The IP vector is then calculated for both  $\pi^+$  and  $\pi^-$ , as well as its parallel and perpendicular projections onto the relative  $\pi$  direction and lastly the two IP vectors,  $IP(\pi^+)$  and  $IP(\pi^-)$  and both four-momenta,  $\pi_{4p}^+$  and  $\pi_{4p}^-$ , are boosted into the reference frame, defined as the zero momentum reference frame of the two pions.

The  $\varphi^*$  distribution obtained, plotted using 7 bins between  $[0, 2\pi]$ , is shown for the MC truth in figure 7.8 and for the hybrid method, using the VBF even sample, in figure 7.9. The  $\varphi^*$  distributions related to the other samples are shown in section 11.7. The fit function used is  $u \cos(\varphi^* - 2\phi_\tau) + w$ .

To validate the choice for using the PV with the default BS constraint,  $\varphi^*$  residual comparison plots are produced between the three different scenarios and the MC distribution, in figure 7.10, as well as between themselves in figure 7.11.

The resolutions are then obtained from a double Gaussian fit, defined by two Gaussian function, a narrow and wide, to the  $\varphi^*$  residual plots. The results are shown for the MC method either in figure 7.12 concerning the VBF even sample or in section 11.8 for the other samples. Similarly results are shown for the Hybrid method either in figure 7.13 for the VBF even sample or in section 11.8 for the other samples. The standard deviation tables for the hybrid method are shown in tables 7.9 7.10.

The obtained resolutions are equivalent within the statistical uncertainties, with the largest difference between the default BS constraint and the no-BS constraint cases. Using a BS constraint different to the default one is not expected to lead to a sizeable change in

Figure 7.8:  $\varphi^*$  distributions for MC truth for the four signal samples.Figure 7.9: From left to right, in the upper row, the VBF CP-even  $\varphi^*$  hybrid method distributions for default and no BS while in the lower row soft BS.

the final results and no additional systematic has been assigned due to the uncertainty of the PV reconstruction.



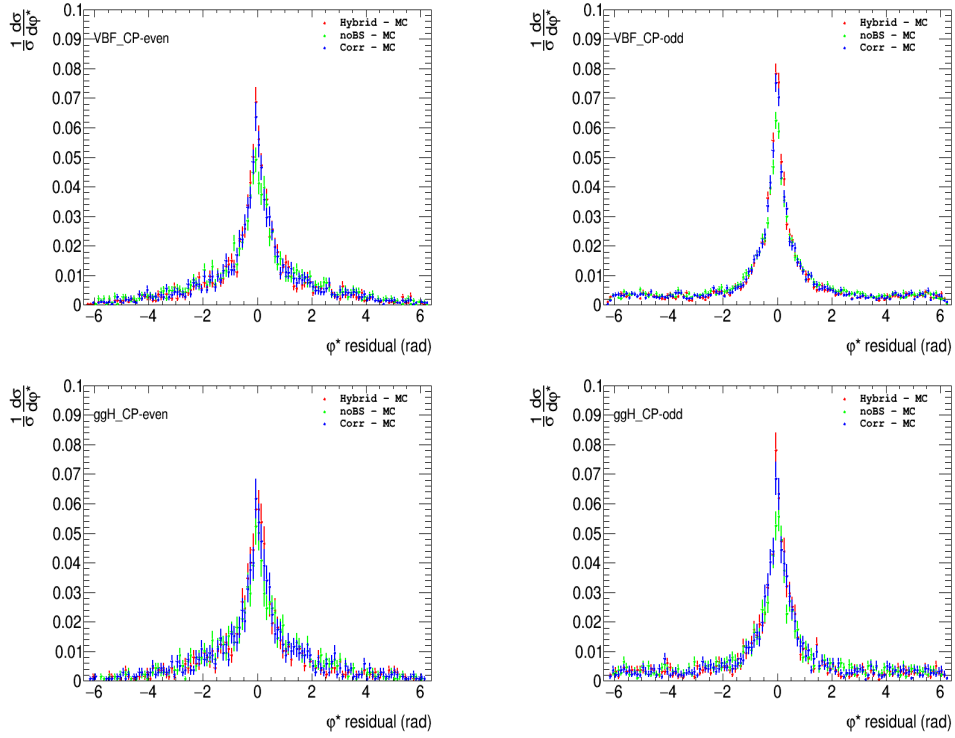


Figure 7.10:  $\varphi^*$  residual plots of the three different hybrid scenarios against the MC for the ggH sample (top(left CP-even, right CP-odd)) and the VBF sample (bottom(left CP-even, right CP-odd)).

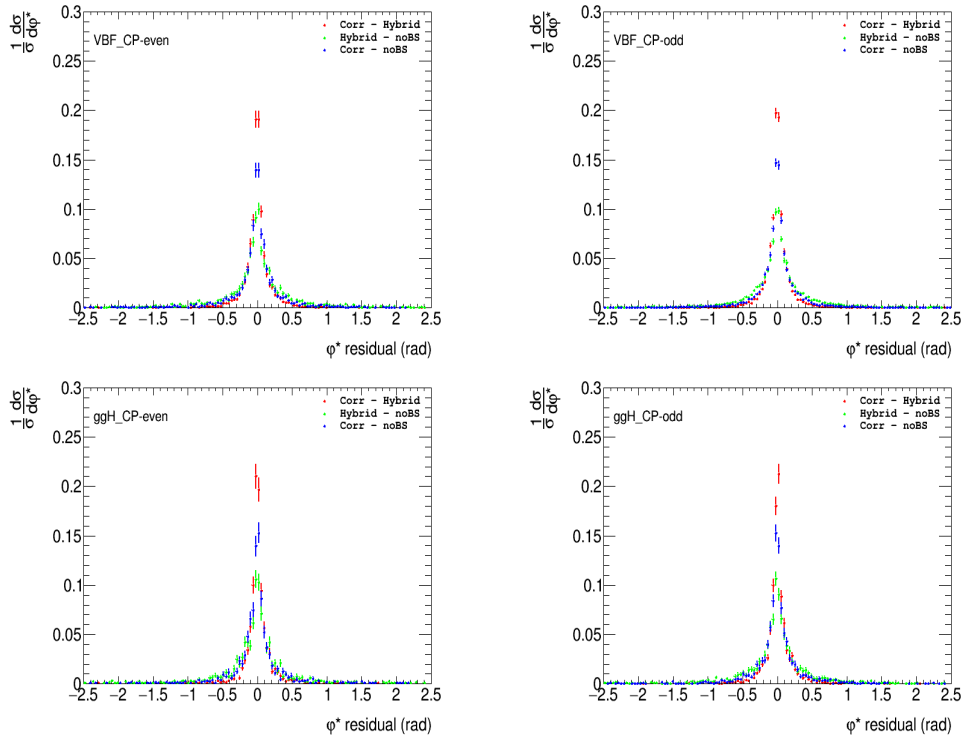


Figure 7.11:  $\varphi^*$  residual plots of the three different scenarios against themselves for the ggH sample (top(left CP-even, right CP-odd)) and the VBF sample (bottom(left CP-even, right CP-odd)).

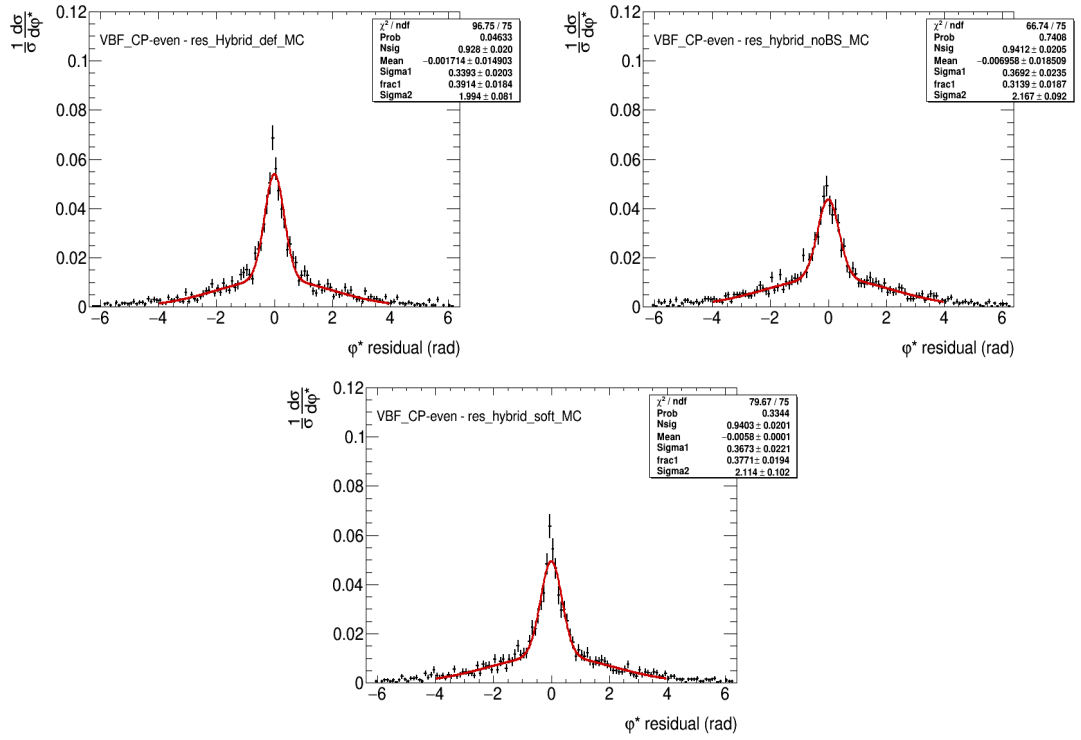


Figure 7.12: From left to right, in the upper row, the VBF CP-even  $\varphi^*$  residual plot when considering (hybrid default - MC) and (hybrid noBS - MC) while in the lower row the (hybrid soft - MC) case.

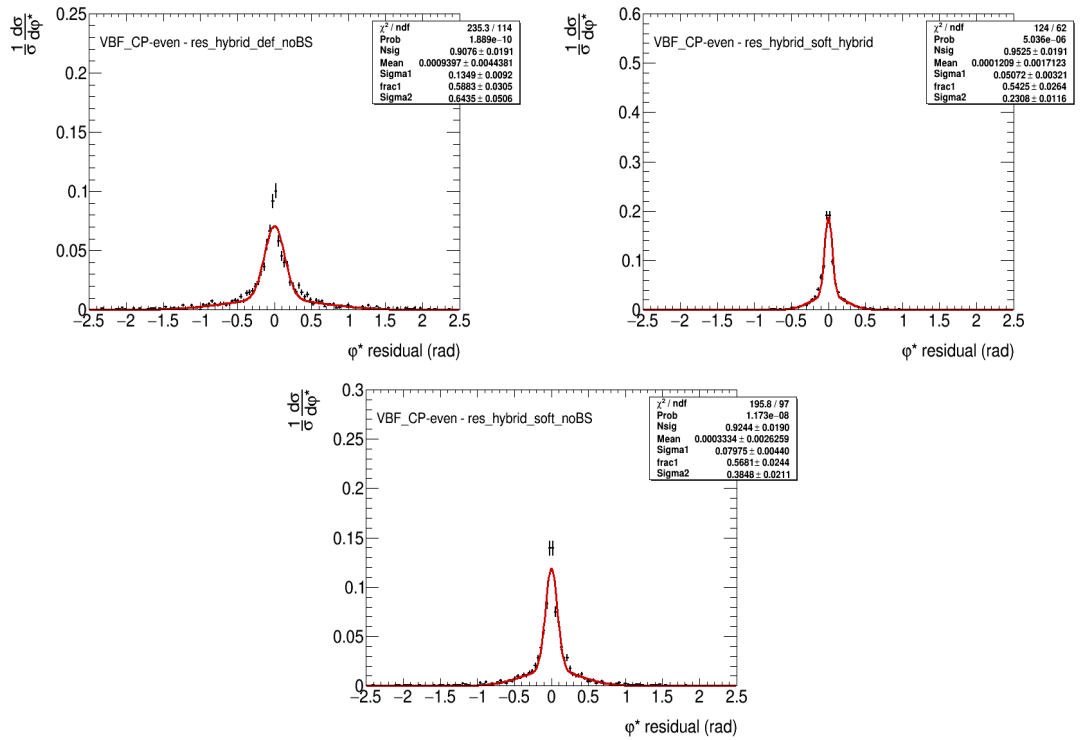


Figure 7.13: From left to right, in the upper row, the VBF CP-even  $\varphi^*$  residual plot when considering (hybrid default - hybrid noBS) and (hybrid soft - hybrid default) while in the lower row the (hybrid soft - hybrid noBS) case.

residual	VBF CP-even		VBF CP-odd	
	$\sigma_{\text{narrow}}$	$\sigma_{\text{wide}}$	$\sigma_{\text{narrow}}$	$\sigma_{\text{wide}}$
default - MC	$0.339 \pm 0.020$	$1.994 \pm 0.081$	$0.334 \pm 0.013$	$2.053 \pm 0.079$
noBS - MC	$0.369 \pm 0.023$	$2.167 \pm 0.092$	$0.387 \pm 0.016$	$2.389 \pm 0.100$
soft - MC	$0.367 \pm 0.033$	$2.114 \pm 0.102$	$0.361 \pm 0.014$	$2.193 \pm 0.091$
default - noBS	$0.134 \pm 0.009$	$0.643 \pm 0.050$	$0.121 \pm 0.006$	$0.56 \pm 0.02$
soft - default	$0.050 \pm 0.003$	$0.230 \pm 0.011$	$0.052 \pm 0.002$	$0.225 \pm 0.010$
soft - noBS	$0.079 \pm 0.004$	$0.384 \pm 0.021$	$0.069 \pm 0.003$	$0.334 \pm 0.016$

Table 7.9: VBF standard deviation table of the Gaussian functions, narrow and wide.

residual	ggH CP-even		ggH CP-odd	
	$\sigma_{\text{narrow}}$	$\sigma_{\text{wide}}$	$\sigma_{\text{narrow}}$	$\sigma_{\text{wide}}$
default - MC	$0.339 \pm 0.027$	$1.868 \pm 0.101$	$0.406 \pm 0.025$	$2.078 \pm 0.152$
noBS - MC	$0.334 \pm 0.050$	$1.746 \pm 0.092$	$0.404 \pm 0.028$	$2.531 \pm 0.207$
soft - MC	$0.331 \pm 0.029$	$1.914 \pm 0.109$	$0.384 \pm 0.022$	$2.038 \pm 0.129$
default - noBS	$0.135 \pm 0.013$	$0.589 \pm 0.088$	$0.102 \pm 0.007$	$0.486 \pm 0.025$
soft - default	$0.1 \pm 0.0$	$0.531 \pm 0.116$	$0.1 \pm 0.0$	$0.357 \pm 0.023$
soft - noBS	$0.081 \pm 0.006$	$0.339 \pm 0.030$	$0.069 \pm 0.004$	$0.346 \pm 0.018$

Table 7.10: ggH standard deviation table of the Gaussian functions, narrow and wide.

## 7.7 Smoothing Options

In order to speed-up and improve the stability of the fit, smoothing and pruning procedures are applied using TRexFitter. Several algorithms are available for smoothing, each has its advantages and disadvantages.

The algorithm rebins the input systematic variation histograms until the number of slope variations is less than a certain threshold. The smoothing procedure is used to get rid of larger-than-expected systematic variations due to statistical fluctuations, making sure that genuine shape effects from systematic uncertainties are kept, while statistical fluctuations in the systematic variations are smoothed out.

The smoothing options available in TRexFitter [171] are:

1. CommonToolSmoothMonotonic
2. CommonToolSmoothParabolic
3. MaxVariation
4. TTBARResonance

The two simplest algorithms, CommonToolSmoothMonotonic and CommonToolSmoothParabolic, rely on rebinning the input distributions of the systematic variations to perform the smoothing. If the statistical uncertainty of the nominal distribution is larger than 5% of the total yield, the algorithm rebins it to 1 final bin. The next step of the algorithm is to merge bins that are most compatible until two extrema, a bin different than the neighbouring one, are found in the case of the CommonToolSmoothMonotonic smoothing function, while until three bins are found in the case of the CommonToolSmoothParabolic smoothing function. The estimation of compatible bins is performed by considering the  $\chi^2$  value between the considered bin and the average of all neighbouring bins. This procedure is repeated until the statistical uncertainty is smaller than 5% in all bins.

The TTBARResonance algorithm has two variations, one related to the correlated source of the events and the second one for the uncorrelated source of events. Two different thresholds are then defined according to the source of the events,  $\delta M = \max(\delta S, \delta N)$  for the correlated case and  $\sqrt{\delta S^2 + \delta N^2}$ , where  $\delta S$  is the statistical uncertainty of the systematic variation S and  $\delta N$  is the statistical uncertainty of the nominal histogram N. For each i-th bin the quantity  $x_{i-1,i}$  is calculated using both statistical uncertainties of the i-th bin,  $S_i$  and  $N_i$ , and the bin before the i-th bin  $i - 1$ ,  $S_{i-1}$  and  $N_{i-1}$ :

$$x_{i-1,i} = \left| \frac{S_i - N_i}{N_i} - \frac{S_{i-1} - N_{i-1}}{N_{i-1}} \right| \quad (7.8)$$

For each  $i$ -th bin is also calculated its relative statistical uncertainty  $\delta x_{i-1,i}$ . If the condition  $x_{i-1,i} < \delta x_{i-1,i}$  is satisfied for at least one bin, the smoothing algorithm looks for neighbouring bins "b-1" and "b" presenting the highest ratio  $\frac{\delta x_{b-1,i}}{x_{b-1,b}}$  and the bins are merged together. The process finishes when there are no bins satisfying the condition  $x_{i-1,i} < \delta x_{i-1,i}$ .

Finally, The MaxVariation algorithm limits the number of variations in the ratio between the Systematic (syst) and nominal by following two steps. In the first one, the distribution is firstly rebinned until the relative statistical uncertainty is below a tolerance value, if not, the process is repeated by halving the threshold. The second step takes place when the uncertainty is below the selected threshold, in which case the ROOT [172] smoothing algorithm is used to create smooth transition between merged bins.

The effects of the different smoothing options on selected systematics are illustrated in figures 7.14 7.15 for JET\_JER\_EffectiveNP\_4 and in section 11.9 for TAUS\_TRUEHADTAU\_EFF\_RNNID\_HIGHPT and TAUS\_TRUEHADTAU\_EFF\_JETID\_DECAY\_MODE\_TRUE\_1P0N\_RECO\_1P0N\_TOTAL. For each systematic, the boost tight high and the boost loose high regions are shown.

Having in mind the reduction of the migration of events between signal regions (SRs) and control region (ZCR), as well as reducing large yield variation, it was decided to study the whole set of smoothing options applied on MC samples. When applying smoothing, the systematics are treated as correlated with the nominal distribution.

To explore the impact on the fit results, the  $\Delta\text{NLL}$  distributions are compared in figure 7.16 for the different smoothing scenarios. The post-fit nuisance parameters ranking are shown in tables 7.11 7.12 7.13 7.14 for the different options. It shows that the biggest number of systematics with a large impact on the fit, and therefore the worst fit-environment, occurs when CommonToolSmoothMonotonic is used. The CommonToolSmoothParabolic instead shows a huge number of systematics pruned away. The TTBAR and the MaxVariation options, in terms of number of systematics with a considerable impact on the fit, are comparable. The MaxVariation option, as illustrated in figure 7.16, shows a higher sensitivity compared to the TTBARResonance algorithm, that reflects the lower impact on the fit of the whole set of nuisance parameters with a particular reduced impact, more than the 5%, by the Jet MET group. The Impact on the final fit of the grouped sources of systematic uncertainties for each smoothing option is shown in tables 7.15 7.16. All of these improvements allowed us to choose the MaxVariation algorithm as the best performing smoothing algorithm.

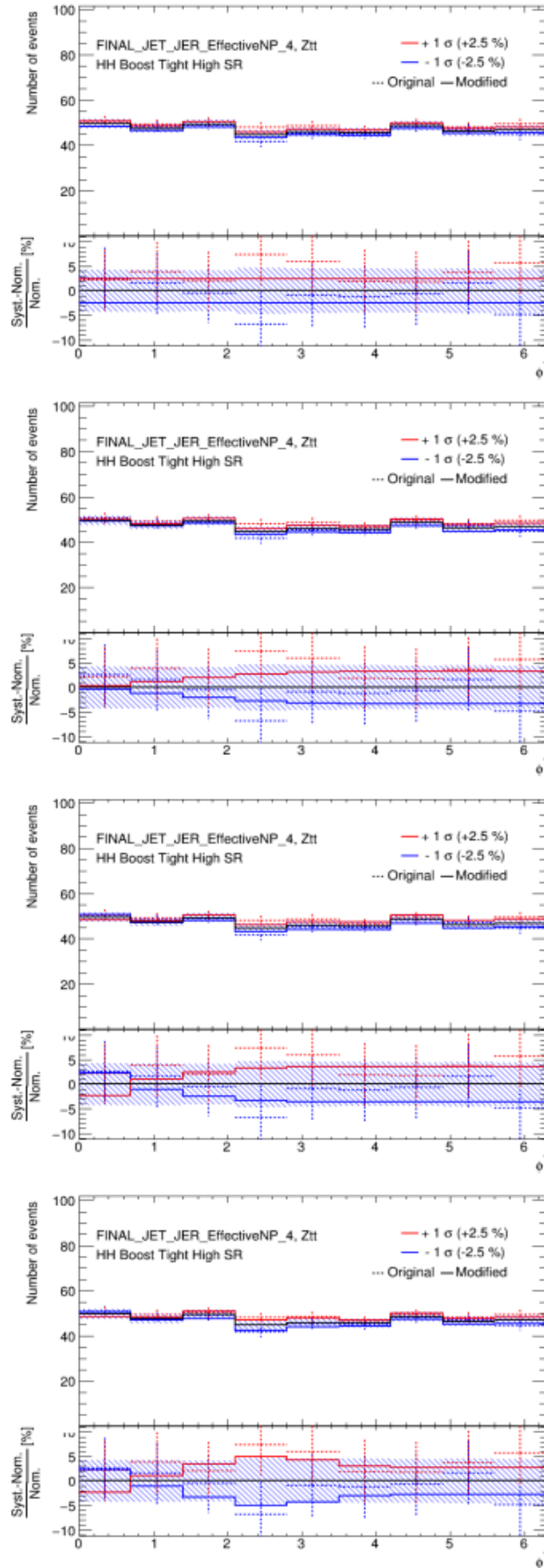


Figure 7.14: The  $JET\_JER\_EffectiveNP\_4$  is shown, for the  $\tau_{had}\tau_{had}$  channel, in the boost tight high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic.

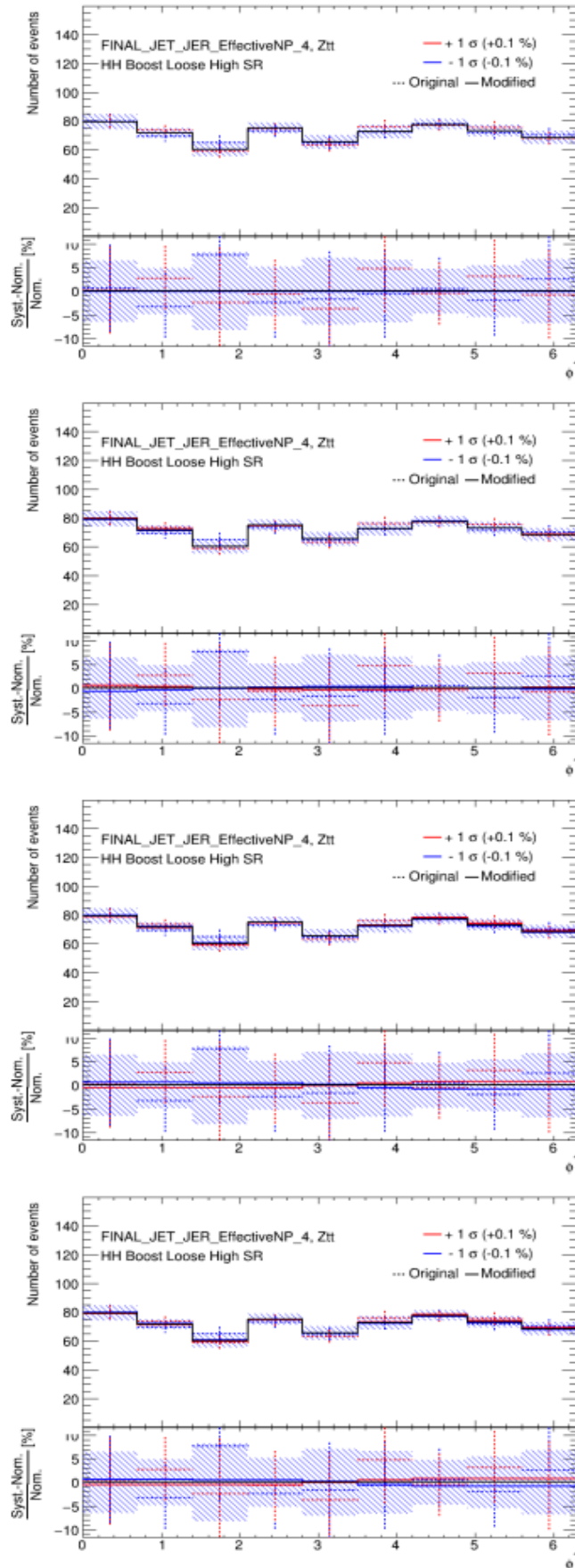


Figure 7.15: The  $JET\_JER\_EffectiveNP\_4$  is shown, for the  $\tau_{had}\tau_{had}$  channel, in the boost loose high signal region for the  $Ztt$  QCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic.

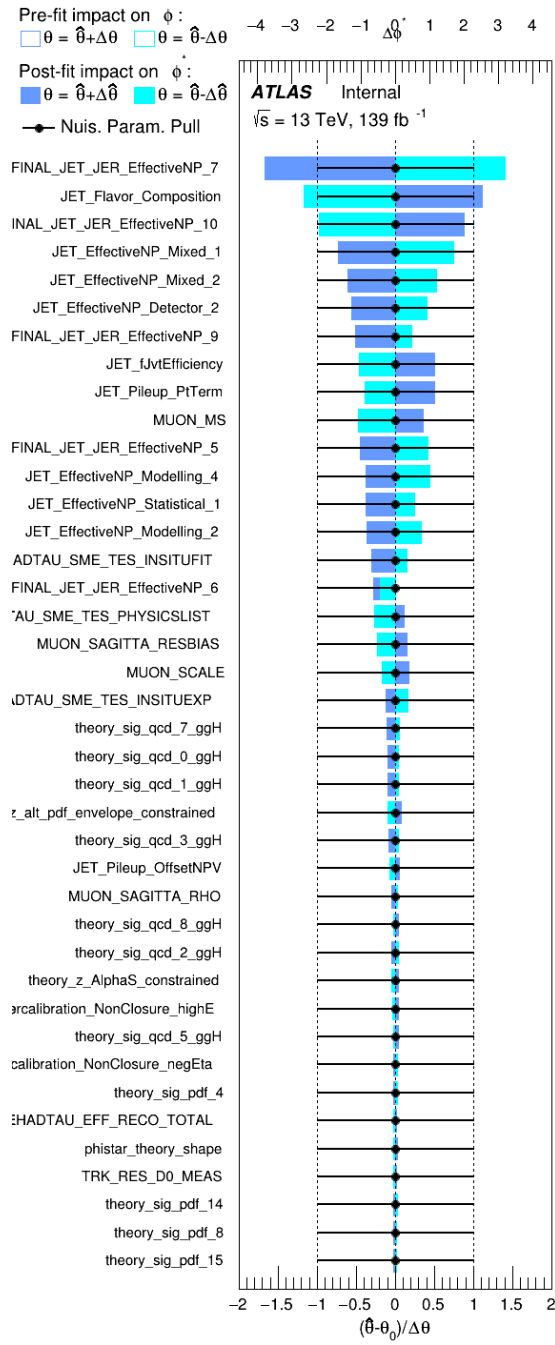


Table 7.11: The nuisance parameters ranking is shown for the MxVariation smoothing option.



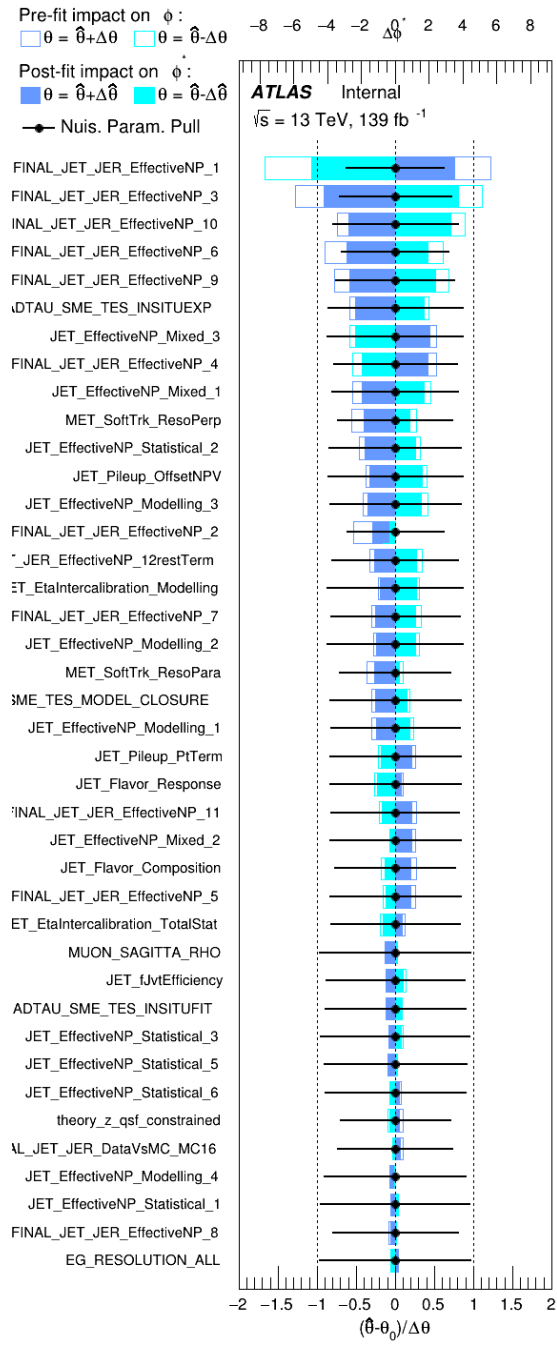


Table 7.12: The nuisance parameters ranking is shown for the TTBAR smoothing option.

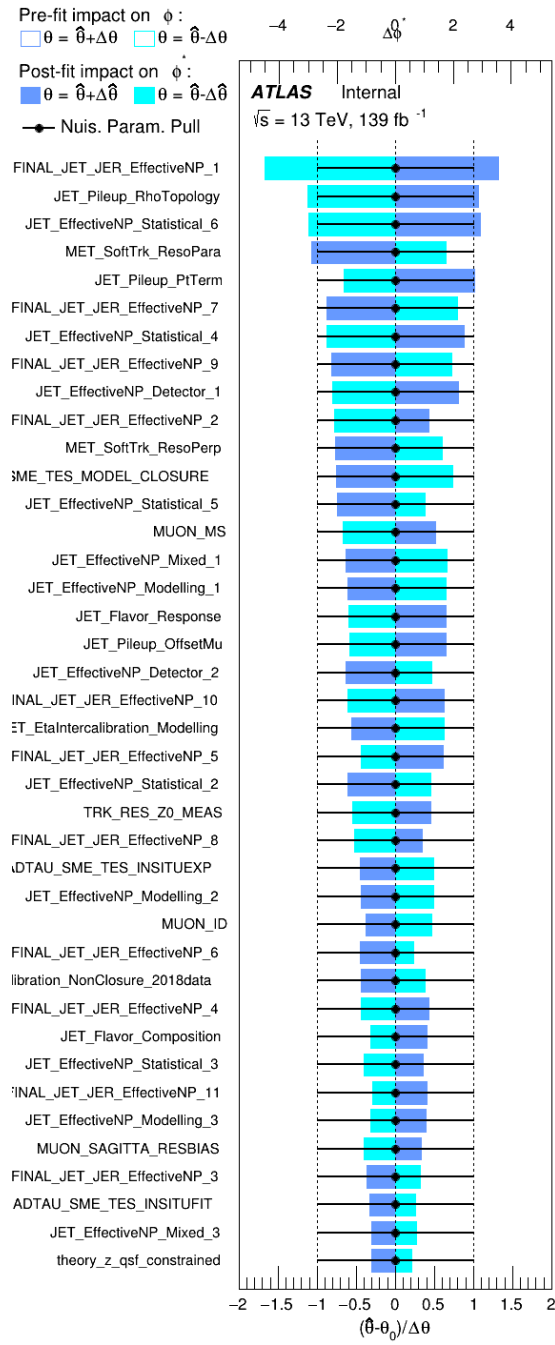


Table 7.13: The nuisance parameters ranking is shown for the Monotonic smoothing option.

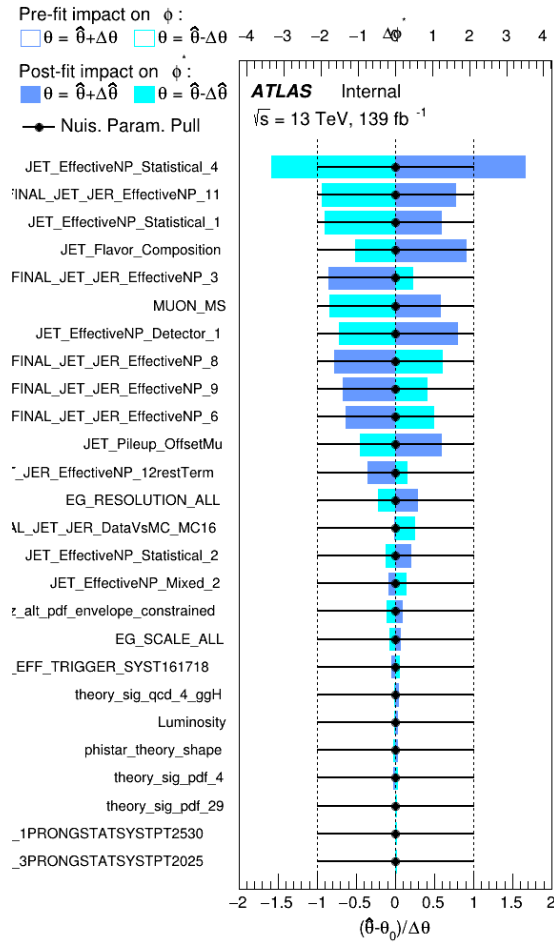


Table 7.14: The nuisance parameters ranking is shown for the Parabolic smoothing option.

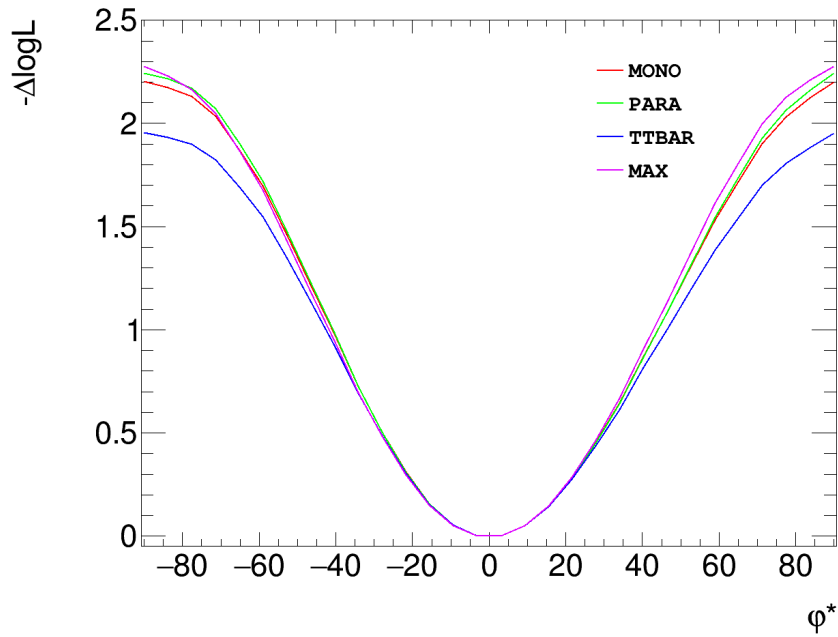


Figure 7.16:  $\Delta \text{NLL}$  distribution showing the comparison among the different smoothing options.

Parameter of interest (POI)	Central Value (0.0°)	Central Value (0.0°)
nuisance parameters	Impact on uncertainty MONO	Impact on uncertainty PARA
Flavour Tagging	+0.08307, -0.01696	+0.01866, -0.01274
DecayMode	+0.71437, -0.30370	+0.68408, -0.64884
Electron	+0.49081, -0.31822	+1.00731, -0.10374
Fake Estimation	+2.01361, -1.82658	+2.10815, -1.09309
Jet	+9.52456, -7.64456	+11.0531, -9.00824
Luminosity	+0.02291, -0.08057	+0.08585, -0.01262
MET	+2.76571, -1.84342	+2.94612, -1.88019
Muon	+2.17679, -2.15648	+1.93121, -2.99726
$\pi_0$	+0.22032, -0.35615	+0.13652, -0.45726
SignTheory	+1.75337, -0.82439	+1.93428, -1.23445
Tau	+2.71177, -1.85315	+2.93096, -1.89197
Track	+1.00715, -1.58984	+0.48840, -1.46757
ZttTheory	+3.87409, -2.74642	+3.58987, -1.94353

Table 7.15: Impact on the final fit of the grouped sources of systematic uncertainties. From left to right, CommonToolSmoothMonotonic and CommonToolSmoothParabolic.

POI	Central Value (0.0°)	Central Value (0.0°)
nuisance parameters	Impact on uncertainty MaxVar	Impact on uncertainty TTBAR
Flavour Tagging	+0.12504, -0.01503	+0.07221, -0.05147
DecayMode	+0.60973, -0.60185	+0.88167, -1.65005
Electron	+0.93888, -0.83942	+0.22930, -0.19635
Fake Estimation	+1.79151, -1.19243	+1.05469, -1.12376
Jet	+9.82024, -8.04397	+9.22727, -9.98675
Luminosity	+0.07417, -0.13837	+0.04472, -0.07421
MET	+1.46022, -0.98830	+1.79812, -1.82268
Muon	+1.74405, -1.32031	+0.21513, -0.45233
$\pi_0$	+0.31920, -0.71896	+0.10468, -0.30292
SignTheory	+1.83211, -0.77907	+1.60111, -1.25583
Tau	+1.79929, -1.84424	+2.13869, -2.67085
Track	+1.36206, -0.6865	+0.08612, -0.11379
ZttTheory	+3.32901, -2.20039	+2.56561, -3.13917

Table 7.16: Impact on the final fit of the grouped sources of systematic uncertainties. From left to right, MaxVariation and TTBAR.

# RESULTS

## 8

In this section an overview about the fit model is given in section 8.1, in particular the different CP templates are then discussed in subsection 8.1.1, while the treatment of the nuisance parameters is evaluated in subsection 8.1.2 as well as the sensitivity regions used in this analysis in subsection 8.1.3. The likelihood function used to verify the sensitivity of both blinded and unblinded fits is then discussed in section 8.2 and finally the results are shown in section 8.3, where the Asimov dataset results are presented in section 8.3.1 while the unblinded results in section 8.3.2.

### 8.1 Fit Model

The main observable of this analysis is  $\varphi_{CP}^*$ , which allows the extraction of the CP mixing angle  $\phi_\tau$ .  $\phi_\tau$  is estimated using a Maximum Likelihood (ML) fit performed using TRexFitter package [171] with additional tools such as Histfactory [173], RooFit [174] and RooStats [175]. The packages are also equipped with other components able to perform template morphing and either pruning or smoothing of the systematic variation histograms.

The  $\varphi_{CP}^*$  fit is performed using the 8  $\tau_{lep}\tau_{had}$  signal regions (SR) and the 8  $\tau_{had}\tau_{had}$  signal regions plus 2 control regions (CR), defined in section 6.3.2. To get a unified value of the CP mixing angle, the  $\phi_{CP}^*$  templates of all the decay mode combinations are merged together in each signal region or control region. The control regions are needed to constrain the background normalisation and the nuisance parameters and are useful to describe the systematic uncertainties, in particular a single-bin histogram is used to retrieve the  $Z \rightarrow \tau\tau$  normalisation. The signal normalisation is not computed in the ML fit in order to estimate  $\phi_\tau$  from only the  $\phi_{CP}^*$  shape distribution.

Merging the whole set of decay modes in each SR, brings with it the need to firstly fix the decay mode combination normalisation and their respective ratios and secondly

to include as fit nuisance parameters the normalisation uncertainties and the decay mode migration.

As defined previously, the analysis includes, for both the semileptonic channel  $\tau_{\text{lep}}\tau_{\text{had}}$  and the full hadronic channel  $\tau_{\text{had}}\tau_{\text{had}}$ , 8 different signal regions.

The first 4 belong to the vector boson fusion Higgs (VBF) and VH production modes:

- VBF 1 high
- VBF 1 low
- VBF 0 high
- VBF 0 low

while the other 4 belong to the gluon gluon fusion Higgs (ggH) and ttH production modes:

- Boosted tight high
- Boosted tight low
- Boosted loose high
- Boosted loose low

Finally, the two  $Z \rightarrow \tau\tau + jets$  control regions, for each channel:

- VBF 0
- Boosted 0

The schematic summary of the whole set of regions is shown in figure 8.1, and, on top of the already discussed regions, a set of normalisation factors (NF) is defined in order to have the  $H \rightarrow \tau\tau$  signal sample and  $Z \rightarrow \tau\tau$  background samples free to float in some of the regions:

- NF H
- NF VBF  $Z \rightarrow \tau\tau$
- NF Boosted  $Z \rightarrow \tau\tau$

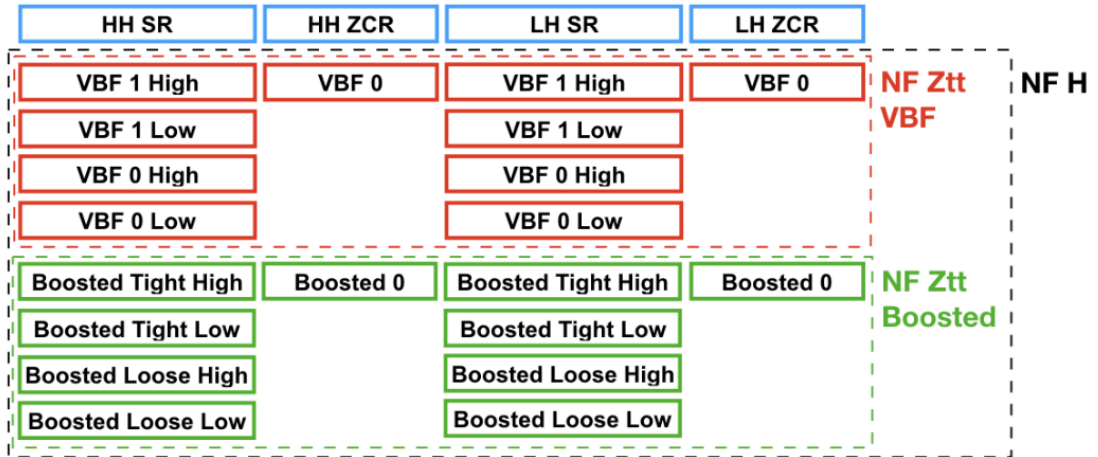


Figure 8.1: Schematic summary of the fit models used in the analysis. They are grouped by topologies (Boosted (green) and VBF (red)) and by decay channels (HH for  $\tau_{\text{had}}\tau_{\text{had}}$  channel and LH for  $\tau_{\text{lep}}\tau_{\text{had}}$  channel). The boxes with dashed lines as boundaries indicate the free floating normalisation factors which are acting on various regions.

### 8.1.1 CP templates

The templates chosen for the binned likelihood function included either CP-even (SM hypothesis) or CP-odd signals plus the expected background. The likelihood function adopted in the following is a Poisson probability density function where each term of the function corresponds to a different bin of the main analysis observable  $\varphi_{CP}^*$  distributions and the POI is estimated maximising the likelihood. Section 8.2 will provide a more comprehensive explanation of the methodology employed in conducting the test statistic. Several hypotheses are being considered, particularly CP-even hypothesis is satisfied for the value of  $\phi_\tau = 0^\circ$ , CP-odd hypothesis for the value  $\phi_\tau = \pm 90^\circ$  and every value between the CP-even and CP-odd degree is therefore considered as a hint of a CP-mixed coupling. The 19 signal templates, created by reweighting the CP-flat VBFH, ggH, VH, and  $ttH \rightarrow \tau\tau$  signal samples obtained via ATLAS simulation with TAUSPINNER [100], correspond to  $\phi_\tau$  values from  $-90^\circ$  to  $+90^\circ$  with  $10^\circ$  intervals.

The fit uses, for the CP hypothesis templates, a one dimensional morphing with a linear interpolation [176] on the parameter of interest  $\phi_\tau$ .

### 8.1.2 Treatment of nuisance parameters (NP)

As described in section 7, this analysis includes several heterogeneous nuisance parameters. Some of them, which take into account the normalisation of the CP-even and -odd templates as well as the  $Z \rightarrow \tau\tau$  background sample, are treated by the fit model as unconstrained nuisance parameters, while most of the other systematic uncertainties, such as experimental and theoretical systematic uncertainties, make use of a Gaussian or Log

constraint. For each nuisance parameter the correlated bin-by-bin  $+1\sigma$  ( $-1\sigma$ ) variations of the input histogram represent the upward (downward) variation.

Another source of nuisance parameters belong to the independent bin-by-bin fluctuation that can appear in the MC templates, thus they are also included in the fit as a Poisson constraint term where the statistical uncertainty of each bin is correlated to the fitted error of the Poisson term itself.

For each systematic uncertainty a symmetrization option is applied to any bins showing both variations  $+1\sigma$  and  $-1\sigma$  lying in the same direction compared to the nominal distribution. The option acts by keeping the biggest variation among them and varying by the same amount but in the opposite direction to the smaller one. In the case of negative bins, or bins with zero content, the yield is set to a very tiny value ( $1e^{-16}$ ). The  $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$  analysis combined fit considers one-sided and "absolute" symmetrisation options.

An additional improvement of the fit, in terms of speed and stability, is done by applying either smoothing or pruning. In order to get rid of the larger than expected systematic variations, the analysis team has chosen the MAXVARIATION smooth function which is used to modify the input systematic variation histograms until the number of slope variations is less than the pruning threshold that is set at 1% and, as a result, the fit ignores the NP if bin-by-bin variations are less than 1%.

The last step of having a more stable and fast fit is performed through a pruning system where, if the total yield of the shape variation of any of the systematics is under 1%, the systematic is pruned away from the fit. The total yield is also affected by the binning of the main fit observable  $\varphi_{CP}^*$ , and to take into account this effect, the pruning threshold is modified just after the rebinning in order to prove that the fit results are independent by the pruning threshold chosen. Thus, it was decided to set the pruning threshold to 0.01, so that regions where the statistical noise is dominant, which would lead to an over-estimation of the shape component of the systematic variation in the fit, are kept out of the likelihood calculation.

### 8.1.3 Sensitivity regions

Before applying the whole set of systematic uncertainties, a statistic-only fit is performed to explore the sensitivity for the combined case and per decay channel, as shown in figure 8.2. The results in table 8.1, show that the contribution of some signal regions is lower than 0.5% of the overall sensitivity, as assessed by  $\Delta\text{NLL}(\pm 90^\circ)$ . In order to improve the stability of the combined fit while taking care about the overall sensitivity, a single bin histogram is used for each of those regions. Signal regions are also used to constrain the normalisation



factors of the Higgs (NF H) and the background (NF  $Z \rightarrow \tau\tau$ ).

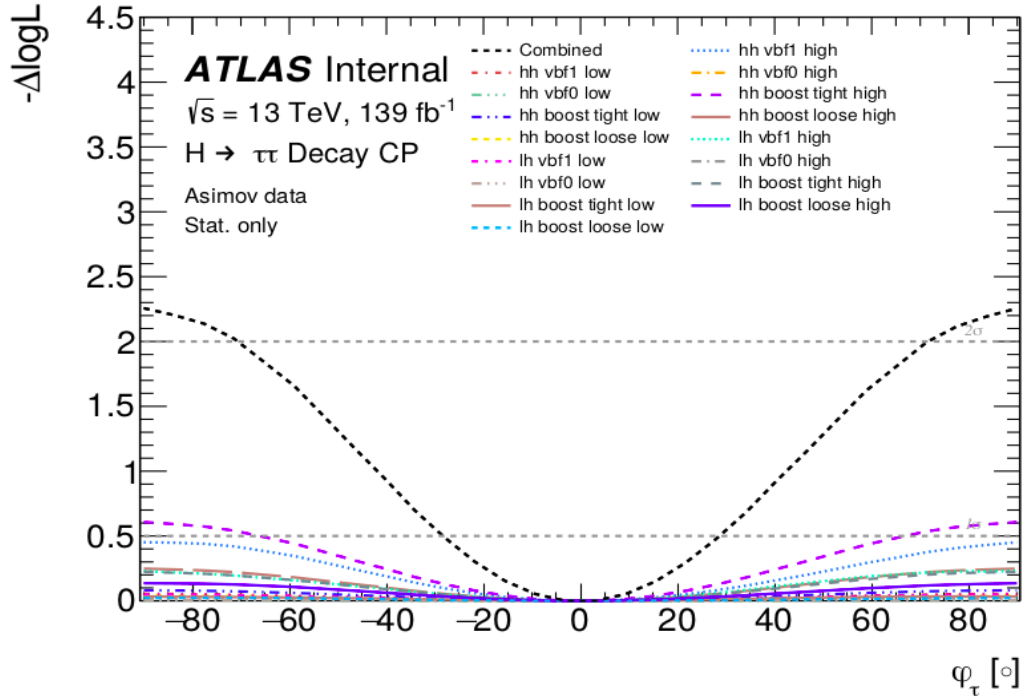


Figure 8.2: Expected  $\Delta\text{NLL}$  curves for the  $\tau_{\text{had}}\tau_{\text{had}}$  plus  $\tau_{\text{lep}}\tau_{\text{had}}$  combined fit with only the statistical uncertainty of the Asimov dataset. The sensitivity of each fit can be estimated by  $\Delta\text{NLL}$  at  $\phi_\tau = \pm 90^\circ$ . The intercepts of the two  $\Delta\text{NLL}$  curves with  $\Delta\text{NLL}=0.5$  (grey horizontal dashed line) correspond to the 68% confidence levels on  $\phi_\tau$  from the combined fit.

Signal Region	Expected $\Delta\text{NLL}$ (+89.9)	Expected $\Delta\text{NLL}$ (-90)
lh vbf 1 signal high	-0.184827	0.1849
hh vbf 1 signal high	0.303696	0.3041
hh vbf 1 signal medium	0.0471471	0.0471574
hh vbf 1 signal low	0.0517643	0.051772
lh vbf 1 signal medium	0.0150913	0.0151166
lh vbf 1 signal low	0.0232238	0.0231985
hh vbf 0 signal high	0.0432063	0.0432313
hh vbf 0 signal medium	0.00828052	0.0082875
hh vbf 0 signal low	0.0079066	0.00790576
lh vbf 0 signal high	0.0369403	0.0369201
lh vbf 0 signal medium	0.00737764	0.00739102
lh vbf 0 signal low	0.00248572	0.00248579
hh boost tight signal high	0.17646	0.17646
hh boost tight signal medium	0.0503772	0.0503771
hh boost tight signal low	0.014189	0.0141891
lh boost tight signal high	0.0949515	0.0949515
lh boost tight signal medium	0.0155676	0.0155654
lh boost tight signal low	0.00571889	0.00571909
hh boost loose signal high	0.0795746	0.0795747
hh boost loose signal medium	0.027808	0.0278081
hh boost loose signal low	0.00592716	0.00592738
lh boost loose signal high	0.0449817	0.0449818
lh boost loose signal medium	0.0146505	0.0146473
lh boost loose signal low	0.00237323	0.0023733

Table 8.1: Expected  $\Delta\text{NLL}(\pm 90^\circ)$  for  $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$  combined fit with only the statistical uncertainties of Asimov data. A binning of 7 bins are used in the observable histogram for all the regions in this fit. Regions contributing less than 0.5% of the overall sensitivity are identified and fewer bins are used for these regions in the final combined fit (to reduce overall fit time without affecting sensitivity).

### 8.1.4 Binning Optimisation

The optimization of observable binning in this analysis, which uses the entire Run-2 dataset as opposed to the prior round of analysis, which only used data acquired in 2015-2016, represents a significant advancement. A three-bin arrangement of  $[0, \frac{\pi}{2}, \frac{3}{2}\pi, 2]$  was developed to address the constraints imposed by both data and MC statistical uncertainty. The three-bin arrangement of  $[0, \frac{\pi}{2}, \frac{3}{2}\pi, 2]$  demonstrates the most effective capacity to discriminate between CP-even and CP-odd states, as shown in Ref. [6]. Typically, when comparing shapes through a likelihood fit, the sensitivity improves as the number of bins increases. However, when the bins increases, the effect of statistical uncertainties of data and MC samples implemented in the likelihood fit will dampen the expected sensitivity causing the sensitivity to level off. According to figure 8.3, the expected sensitivity of the  $\phi_{CP}^*$  levels out at about seven to nine bins.

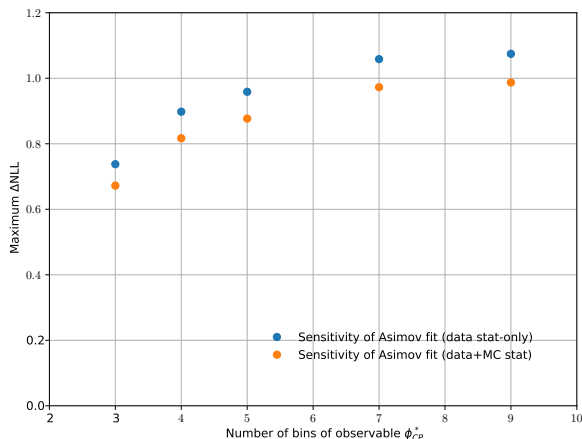


Figure 8.3: Maximum  $\Delta\text{NLL}$  as a function of the number of bins of observable for the Asimov fit using 3, 4, 5, 7, and 9 bins in the observable  $\phi_{CP}^*$  [6].

### 8.1.5 Prefit distribution

After treating the whole set of systematic uncertainties with symmetrisation, smoothing and pruning, as described in sec 8.1.2, this section shows the distributions that are fed to the likelihood model.

Nine bins are used for the signal regions, except for the VBF 1 High and Medium that, for sample size reason, use only 7 bins. All Low regions are represented by 4 bins due to their very low sensitivity. Prefit plots are shown in figure 8.4 for the hh channel and in figure 8.5 for the lh channel.

A single-bin distribution is used for the VBF and boost production modes, leading to 8

---

different control regions (CR), as shown in figure 8.6, 4 related to the  $\tau_{\text{had}}\tau_{\text{had}}$  channel and 4 for the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel, which are used to transfer only yield, but no shape information from the control regions to the signal regions.

Two additional control regions, shown in figure 8.7, are employed to make use of the shape information of the  $\rho$  mass to constrain  $\pi_0$  related uncertainties (smearing and energy rescaling).

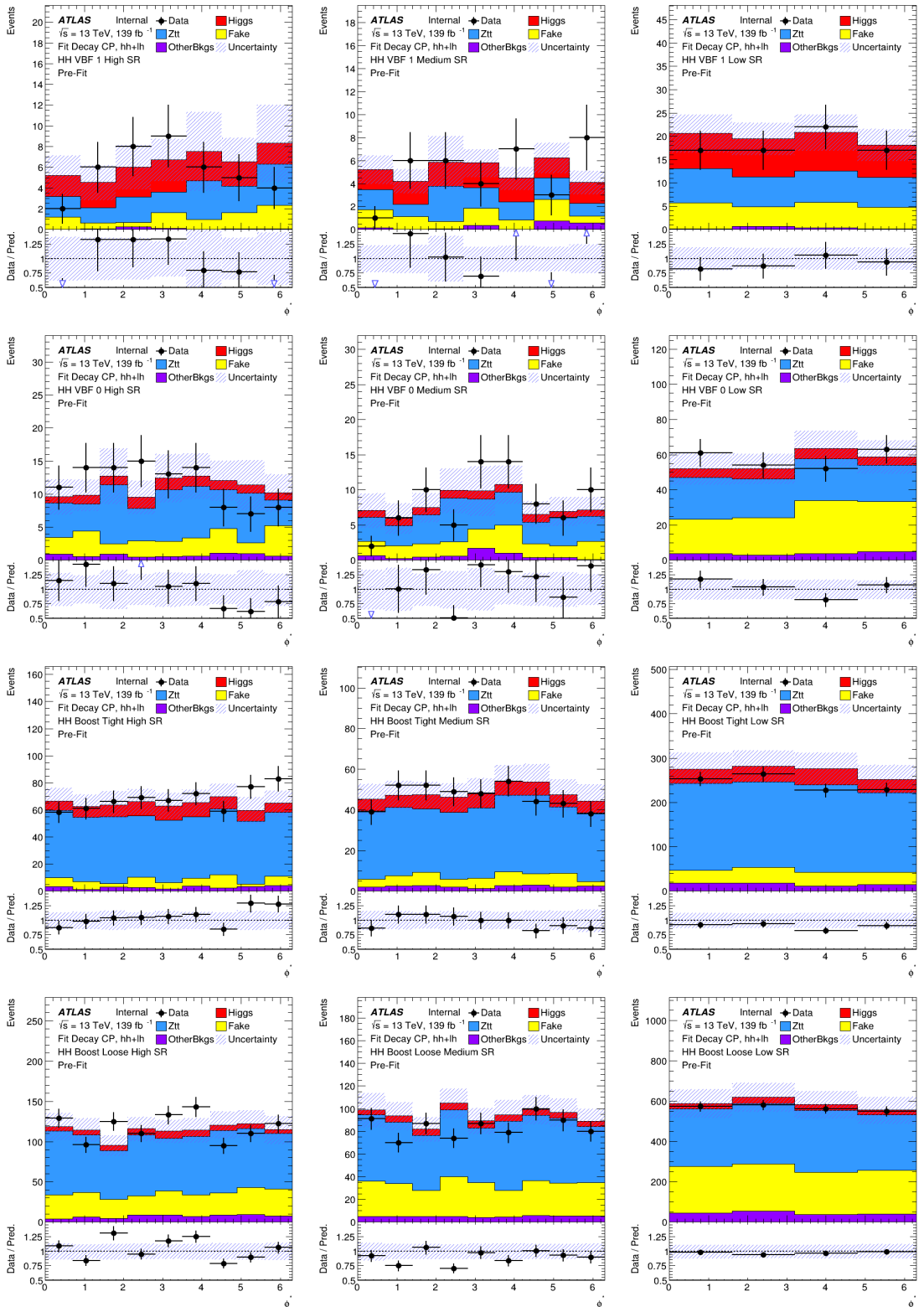


Figure 8.4: Pre-fit distributions in the  $H \rightarrow \tau\tau$  signal regions in the  $\tau_{\text{had}}\tau_{\text{had}}$  channel, which will be included in the likelihood model. ‘OtherBkgs’ includes  $W$ , diboson, top,  $Z \rightarrow \text{ll}$ ,  $H \rightarrow WW$  samples.

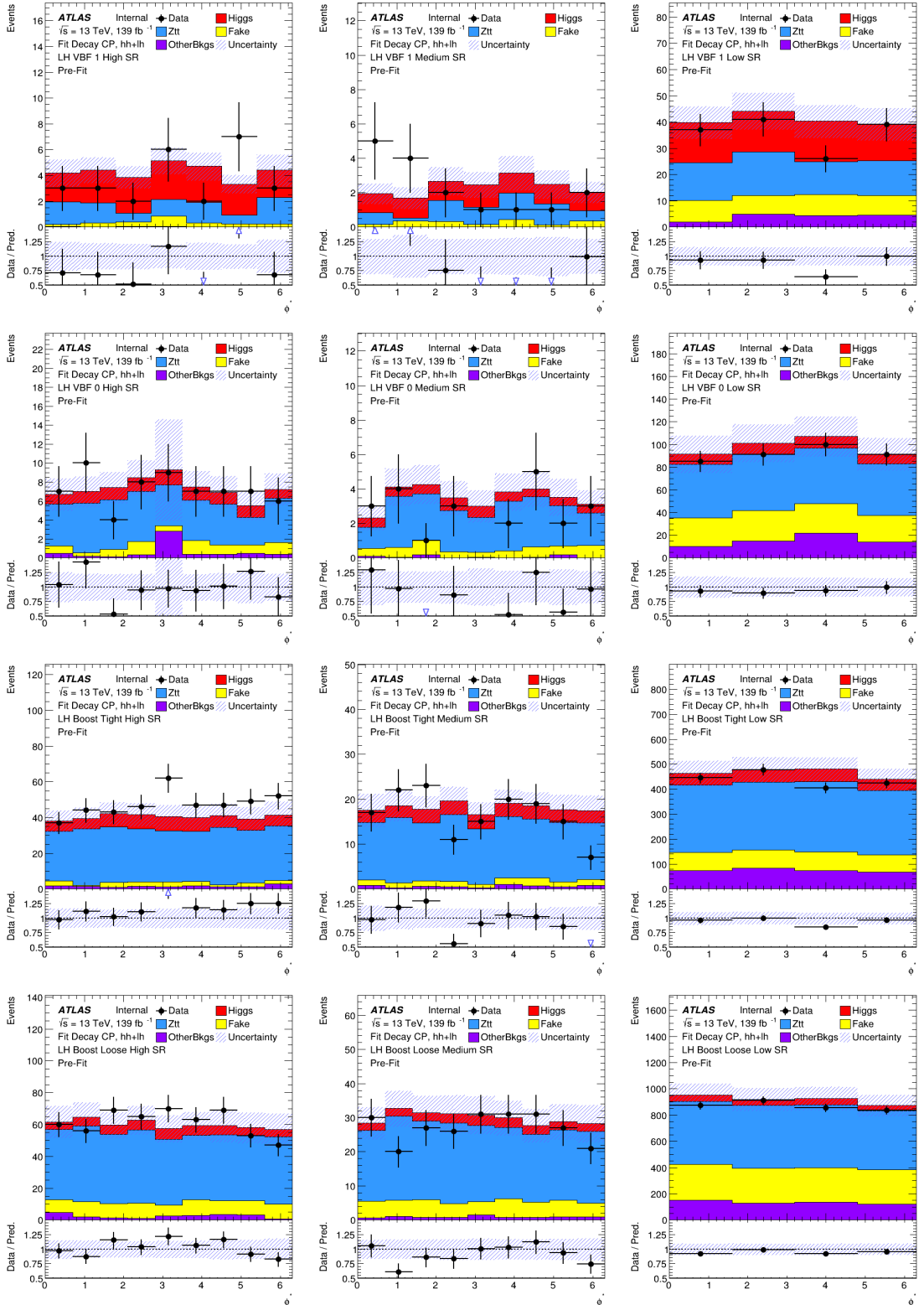


Figure 8.5: Pre-fit distributions in the  $H \rightarrow \tau\tau$  signal regions in the  $\pi_{\text{lep}}\tau_{\text{had}}$  channel, which will be included in the likelihood model. ‘OtherBkgs’ includes  $W$ , diboson, top,  $Z \rightarrow \text{ll}$ ,  $H \rightarrow WW$  samples.

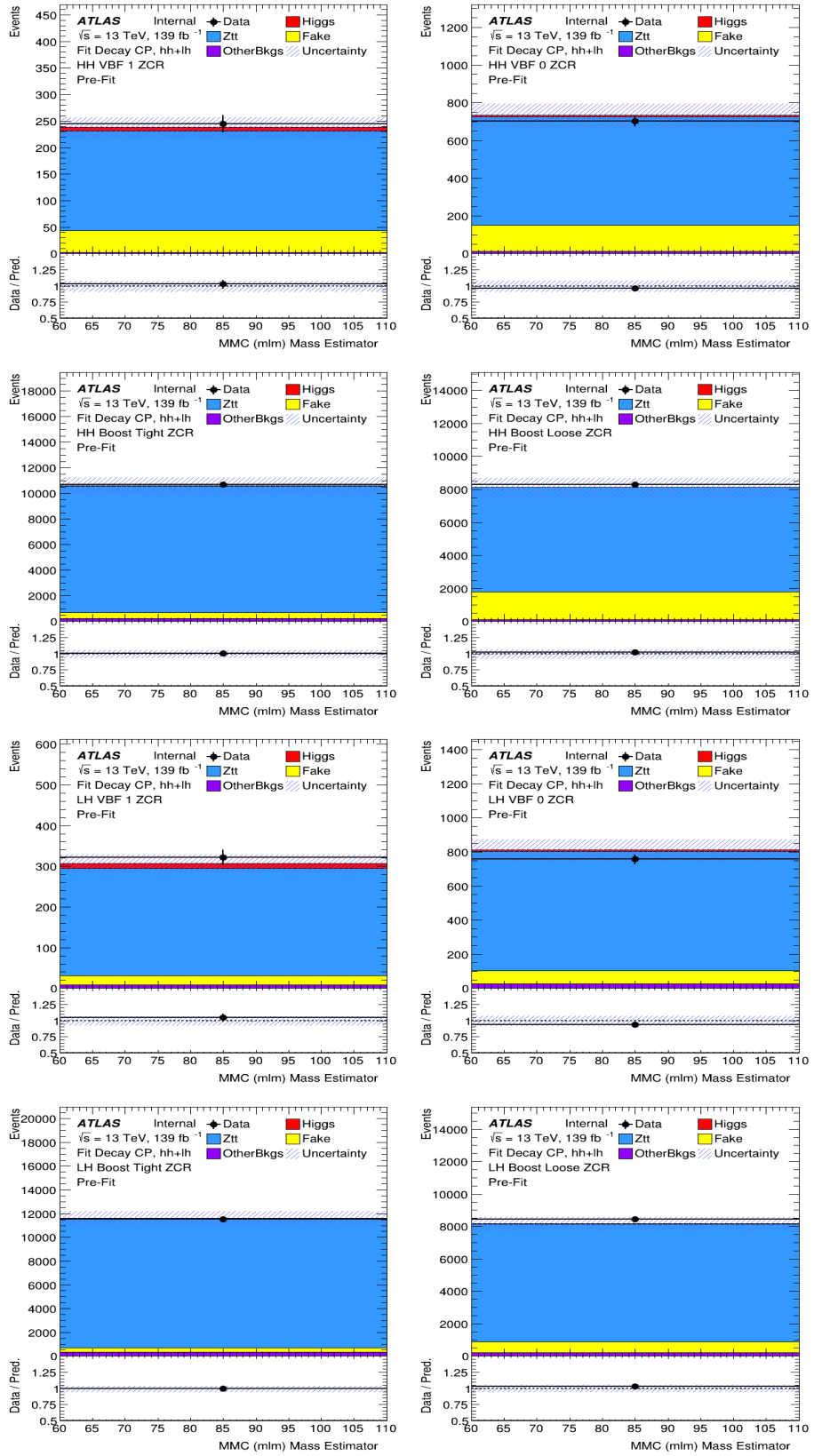


Figure 8.6: Pre-fit distributions in the  $Z \rightarrow \tau\tau$  control regions in both semi-leptonic and full-hadronic channels, which will be included in the likelihood model. ‘OtherBkgs’ includes W, diboson, top,  $Z \rightarrow ll, H \rightarrow WW$  samples.

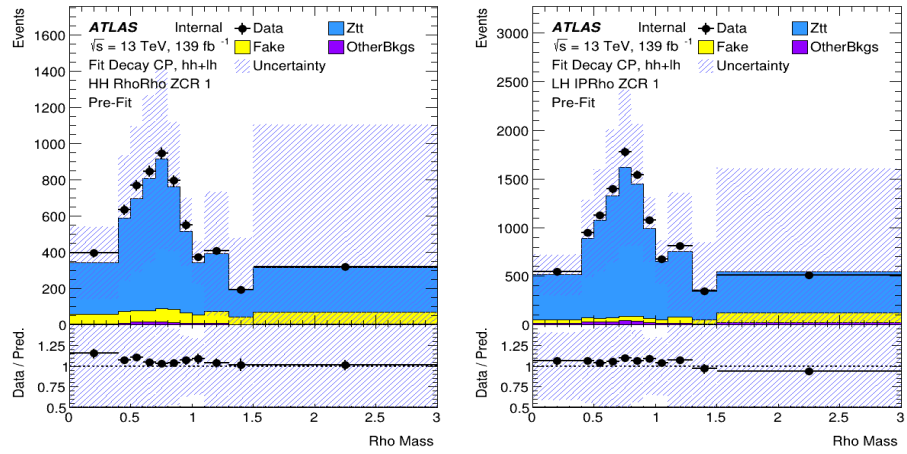


Figure 8.7: Pre-fit distributions in the  $Z \rightarrow \tau\tau$  control regions for the  $\rho$  mass distribution.



## 8.2 Sensitivity

The binned likelihood function  $\mathcal{L}(x, \phi_\tau, \theta)$  is a function to the data  $x$ , the free floating mixing angle  $\phi_\tau$ , and nuisance parameters  $\theta$  corresponding to the systematic uncertainties described in section 7. The weighted signal templates defined in section 8.1.1 for each  $\phi_\tau$  hypothesis are used to evaluate the likelihood function while keeping the background processes, as shown in section 5.2, unchanged.

From the likelihood function, a Negative Log Likelihood (NLL) is obtained by taking the negative logarithm of the likelihood function. The negative log-likelihood is a measure of the discrepancy between the observed data and the expected data predicted by the statistical model. Comparing the negative log-likelihood under different hypotheses helps to estimate model parameters or comparing the quality of the fit between different models. The NLL is the test statistic used in this analysis. The test statistic is validated performing the fit over the Asimov dataset, discussed afterwards in section 8.3.1.

The test statistic can be constructed as a function of  $\phi_\tau$ , and it is expressed by the formula 8.1, Ref. [177].

$$q = -2\ln \frac{\mathcal{L}(\mathbf{x}; \phi_\tau, \theta)}{\mathcal{L}(\mathbf{x}; \phi_\tau, \theta)} = -2\ln \lambda(k) \quad (8.1)$$

where the numerator corresponds to the conditional maximum-likelihood estimator while the denominator represents the unconditional maximum-likelihood estimator. In this analysis, we adopted as a NLL the  $q/2$  function, and thus the confidence range  $[\phi_\tau - \sigma_{\phi_\tau}, \phi_\tau + \sigma_{\phi_\tau}]$  is extracted from the NLL using the Neyman construction [178]:

$$-\log \mathcal{L}(\phi_\tau \pm N\sigma_{\phi_\tau}) = -\log \mathcal{L}_{max} + \frac{N^2}{2} \quad (8.2)$$

The NLL curve is produced by calculating the NLL value for each  $\phi_\tau$  hypothesis and for each specific dataset  $x$ . The best estimator  $\hat{\phi}_\tau$  at which the NLL curve is minimal, obtained by reading off the  $\Delta NLL = NLL - NLL_{min}$  at 0.5, determines the 68% central confidence interval ( $N = 1$  standard double-sided Gaussian quantiles “ $1\sigma$ ”). This interval contains the true value of  $\phi_\tau$  in 68% of all cases.

The highest contribution to the sensitivity mainly comes from the high categories, either VBF, boost loose or boost tight.

## 8.3 Fit results

Two different sets of results are produced in this analysis. An Asimov dataset is used to show simulated results, as shown in section 8.3.1, while the data results are shown in section 8.3.2.

### 8.3.1 Asimov fit results

An Asimov dataset [177] is typically constructed by assuming a specific set of parameter values and then generating simulated data based on those values. When it is used in a statistical test, it will return the result expected from that model configuration. An Asimov dataset is built as binned dataset, in which the event count in each bin is set to the expected event yield for the chosen model parameters.

An Asimov dataset is created in statistical hypothesis testing by assuming a certain set of model parameters that represent a potential hypothesis or scenario. Theoretical predictions for the observable quantities in a certain experiment are defined by these model parameters. To generate an Asimov dataset, a large number of events based on the assumed model parameters has to be simulated. These simulated events reflect the expected behavior of particles or physical processes as predicted by the chosen hypothesis. The resulting dataset contains simulated measurements of various observables, such as particle energies, momenta, or decay products, that would be observed in an idealized experiment. The Asimov dataset is used in a statistical study to determine the model's compatibility by measuring how well the chosen model's parameters match the observed data. By measuring the differences between the simulated dataset and the real experimental data, it enables users to evaluate the importance of their hypothesis or compare various hypotheses.

The Asimov fit results, shown in this section, belong to the  $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$  combined fit. Results are presented for both 1-dimensional and 2-dimensional fits.

The expected sensitivity on  $\varphi_\tau$  is shown as a function of  $-\Delta\ln(L)$ , in figure 8.10, as well as a function of the signal strength  $\mu^{\tau\tau}$  in figure 8.11.

The 1D result, as shown in figure 8.10, is calculated using 30 data points, from  $-90^\circ$  to  $+89.9^\circ$ . The expected sensitivity value for  $\varphi_\tau$  is  $0^\circ \pm 29^\circ$  at one standard deviation with a CP-odd exclusion limit of  $2.24\sigma$  for the stat-only case, as shown in figure 8.2 and  $2.1\sigma$  for the statistical plus systematic uncertainties fit as shown in table 8.2.

The expected post-fit values of the signal strength and background normalisations are shown in figure 8.8. Figure 8.9 shows the systematic uncertainties with the biggest impact on the final fit results.

The combined,  $\tau_{\text{had}}\tau_{\text{had}}$  and  $\tau_{\text{lep}}\tau_{\text{had}}$ , fit performed using the Asimov dataset is mostly affected by statistical uncertainties. The other two biggest sources of uncertainties belong to the MC statistical uncertainties, which include sample size uncertainties, and normalisation factors uncertainties as shown in table 8.3.

Visible effects in the jet uncertainties are mainly due to their dependence on the Missing Mass Calculator (MMC) between the control region and the signal regions. An additional

effect arises from the jet selection considered in the VBF and boost signal regions categorisation. As a result of those effects, most of the jet related uncertainties have a big impact on the fit and therefore they appear in the top 10 ranking positions as shown in figure 8.9, where the reason why some parameters are pulled in the fit to the maximum values is that they are left free floating and therefore unconstrained by the fit.

Tau related uncertainties do not have a big impact and are mainly dominated by systematics considering the tau energy scale. The  $\tau$  lepton decay mode classification uncertainties have a small impact on  $\varphi_\tau$ , as well as other sources such as theory prediction, fake estimation, track and MET. The luminosity uncertainty and the flavour tagging uncertainty have a negligible impact.

The 2D likelihood scan, as shown in figure 8.11, is performed between the signal strength  $\mu_{\tau\tau}$  and the CP mixing angle  $\phi_\tau$ . It is calculated using 29 data points in  $\phi_\tau$  and 29 data points in  $\mu_{\tau\tau}$  that correspond to 841 data points that are fit in parallel for timing reasons. The Standard Model expectation is within the  $1\sigma$  contour of the observed limits.

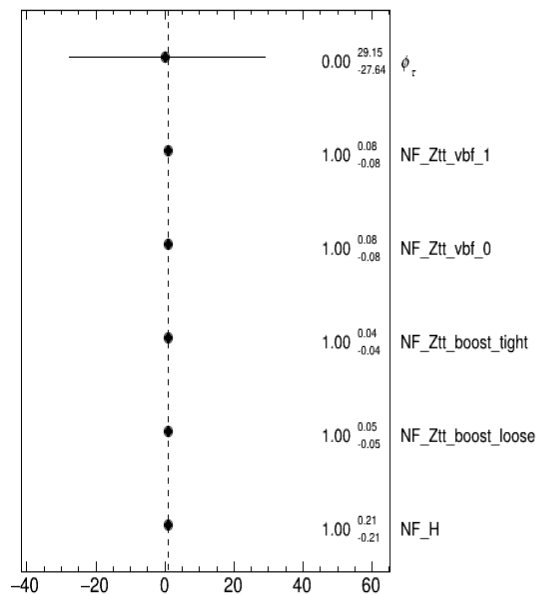


Figure 8.8: Expected post-fit values and uncertainties from the  $\tau_{\text{had}}\tau_{\text{had}}$  and  $\tau_{\text{lep}}\tau_{\text{had}}$  combined fit on the Asimov dataset for the signal strength and background normalisations (NF) for  $Z \rightarrow \tau\tau$  samples in VBF and Boost regions.

Fit setting	CP-odd exclusion limits
stat-only	$2.24\sigma$
stat-syst	$2.10\sigma$

Table 8.2: Asimov CP-odd exclusion limits.

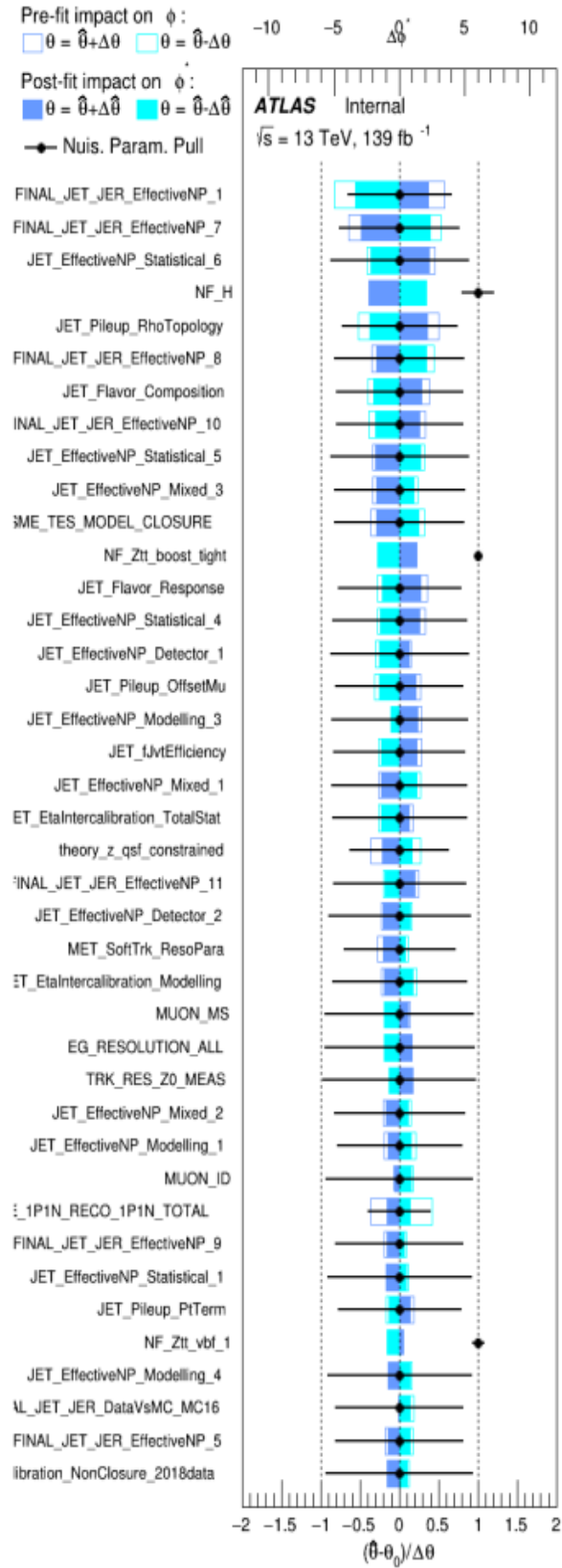


Figure 8.9: Expected ranking of 40 systematic uncertainties with the highest post-fit impact on  $\varphi_\tau$  in the  $\tau_{\text{had}}\tau_{\text{had}}$  and  $\tau_{\text{lep}}\tau_{\text{had}}$  combined fit on the Asimov dataset.

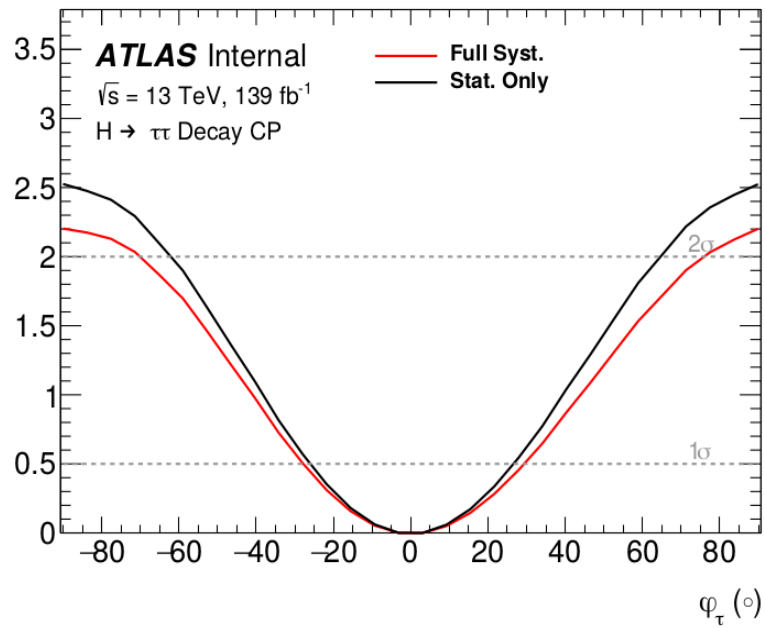


Figure 8.10: Expected  $\Delta\text{LNN}$  curves for the  $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$  combined Asimov fit with all statistical and systematic uncertainties. The expected value for  $\varphi_\tau$  is  $0 \pm 29$  at one standard deviation.

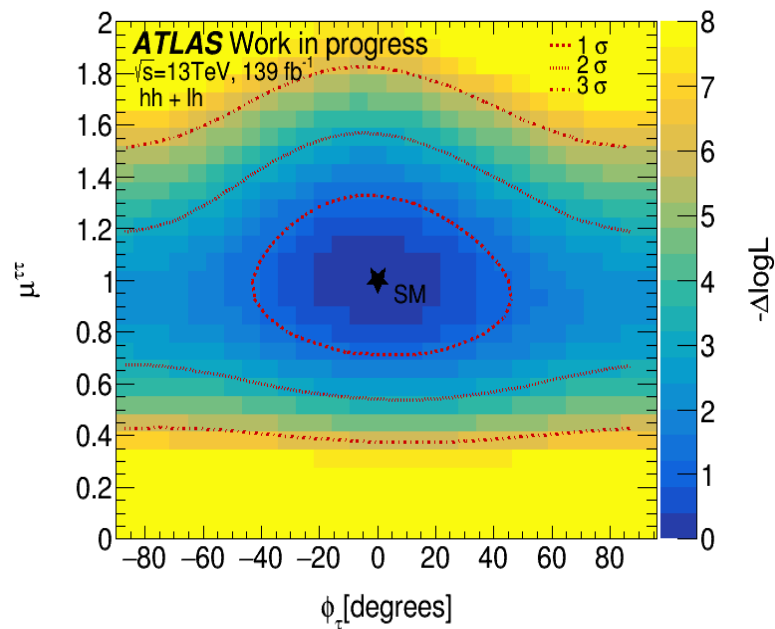


Figure 8.11: Expected 2D  $\Delta\text{NLL}$  contours as functions of  $H \rightarrow \tau\tau$  signal strength  $\mu^{\tau\tau}$  versus CP mixing angle  $\varphi_\tau$  for the  $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$  had combined Asimov fit with all statistical and systematic uncertainties (stat+syst).

Parameter of interest	Central Value (0.00) <sup>o</sup>
Set of nuisance parameters	Impact on uncertainty
Total	+29.12 / -27.65
DataStat	+26.46 / -25.94
FullSyst	+12.18 / -9.59
Jet	+9.82 / -8.04
Met	+1.46 / -0.99
Electron	+0.94 / -0.84
Muon	+1.74 / -1.32
Tau	+1.80 / -1.84
Fake Estimation	+1.79 / -1.19
Tau Decay mode Class.	+0.61 / -0.60
Neutral Pion	+0.32 / -0.72
Luminosity	+0.07 / -0.14
Track	+1.36 / -0.69
Flavour Tagging	+0.13 / -0.02
Theory unc.(H $\rightarrow\tau\tau$ )	+1.83 / -0.78
Theory unc.(Z $\rightarrow\tau\tau$ )	+3.33 / -2.20
MC stat.	+3.67 / -3.58

Table 8.3: Expected grouped systematic sources with the highest post-fit impact on  $\varphi_\tau$  in the  $\tau_{\text{had}}\tau_{\text{had}}$  and  $\tau_{\text{lep}}\tau_{\text{had}}$  combined fit on the Asimov dataset.

### 8.3.2 Unblinded fit results

Similarly to the Asimov fit, the fit to the data is mainly affected by statistical uncertainties as well as the jet related uncertainties which dominate the top positions of the ranking plot shown in figure 8.13 and therefore have the biggest impact in the data fit. Other big sources are the tau uncertainties, the signal theory uncertainty and the limited simulated sample size corresponding to the MC statistics. Similarly to figure 8.9, the reason why figure 8.13 shows most NF pulled to their maximum value is that they are unconstrained by the fit. All sources of uncertainties are shown in the grouped impact table 8.4.

The normalisation factors, as well as the expected post-fit values and uncertainties, are shown in figure 8.12.

The 1D and 2D fit are also produced for the data fit. The minimum of the  $\Delta\text{NLL}$  curve 8.14 is located at  $8.68^\circ$ , which is the measured value for  $\varphi_\tau$ , with the  $1\sigma$  limits corresponding to  $+17.77^\circ$  and  $-16.16^\circ$ . A 2-dimensional scan, see figure 8.15, is also performed between the signal strength and the CP mixing angle using the same number of data points as in the Asimov parallel fit (841 data points). Also for the data fit, the Standard Model expectation is within the  $1\sigma$  contour of the observed limits.

The pure CP-odd hypothesis is rejected at  $3.4\sigma$  (standard deviation). There is a  $1.3\sigma$  uncertainty reduction from the expected (Asimov) to the observed exclusion in data.

The measurement is consistent with the Standard Model expectations ( $\phi_\tau=0$ ). The tension between the expected and observed values is probably due to statistical fluctuations in data, mainly contributed by some High categories.

Figures 8.16 shows one such example, HH VBF High, where the background-subtracted  $\phi^*$  distribution exhibits a larger amplitude than the expected ones for the pure CP-even or CP-odd cases. Further work is ongoing to check if this is the right explanation for the better than expected sensitivity.

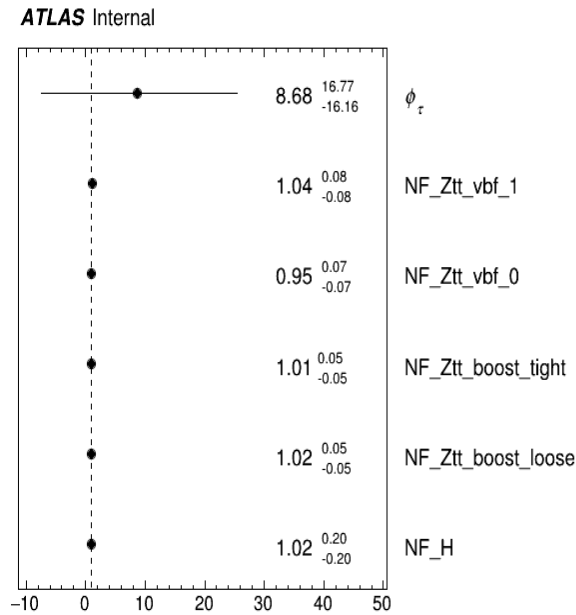


Figure 8.12: Expected post-fit values and uncertainties from the  $\tau_{\text{had}}\tau_{\text{had}}$  and  $\tau_{\text{lep}}\tau_{\text{had}}$  combined fit on the unblinded data for the signal strength and background normalisations (NF) for  $Z \rightarrow \tau\tau$  samples in VBF and Boost regions.

Parameter of interest	Central Value (8.68) <sup>o</sup>
Set of nuisance parameters	Impact on uncertainty
Total	+16.77 / -16.16
DataStat	+15.90 / -15.32
FullSyst	+5.32 / -5.14
Jet	+4.22 / -4.46
Met	+0.87 / -
Electron	+0.32 / -0.25
Muon	+0.87 / -0.91
Tau	+1.07 / -1.02
Fake Estimation	+0.89 / -0.33
Tau Decay mode Class.	+0.38 / -0.31
Neutral Pion	+0.30 / -
Luminosity	+0.24 / -
Track	+0.99 / -0.41
Flavour Tagging	+0.29 / -
Theory unc.(H $\rightarrow\tau\tau$ )	+2.08 / -0.89
Theory unc.(Z $\rightarrow\tau\tau$ )	+1.39 / -0.85
MC stat.	+1.44 / -1.34

Table 8.4: Grouped impact of different systematic sources for the data fit result of the  $\tau_{\text{had}}\tau_{\text{had}}$  and  $\tau_{\text{lep}}\tau_{\text{had}}$  combined fit with  $\varphi_\tau$  as parameter of interest.



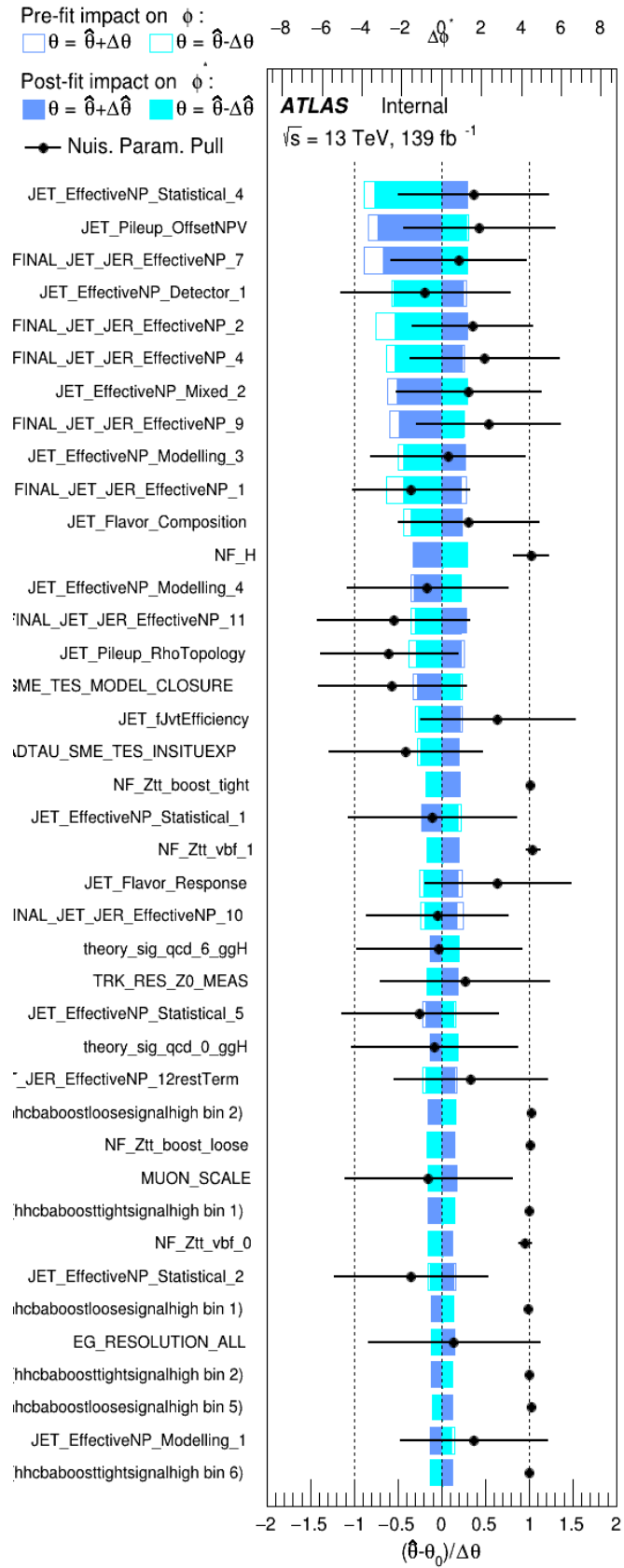


Figure 8.13: Expected ranking of 40 systematic uncertainties with the highest post-fit impact on  $\varphi_\tau$  in the  $\tau_{\text{had}}\tau_{\text{had}}$  and  $\tau_{\text{lep}}\tau_{\text{had}}$  combined fit on the unblinded data.

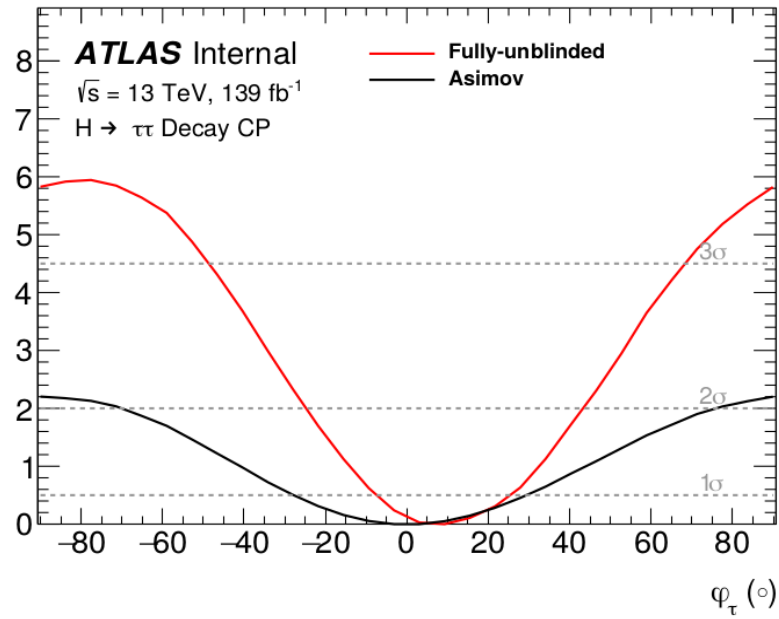


Figure 8.14: Expected  $\Delta LNN$  curves for the  $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$  combined data fit with all statistical and systematic uncertainties. The expected value for  $\varphi_\tau$  is  $0^\circ \pm 29^\circ$  at one standard deviation.

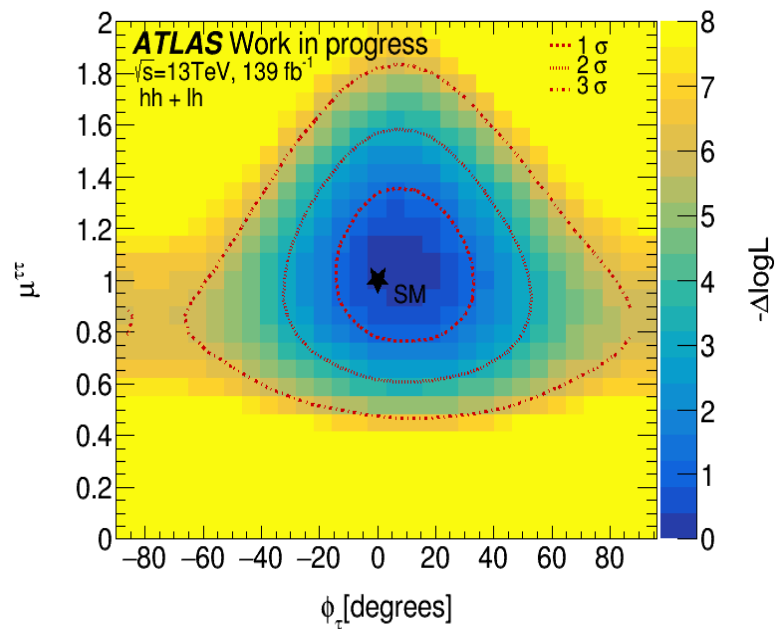


Figure 8.15: Expected 2D  $\Delta NLL$  contours as functions of  $H \rightarrow \tau\tau$  signal strength  $\mu^{\tau\tau}$  versus CP mixing angle  $\varphi_\tau$  for the  $\tau_{\text{had}}\tau_{\text{had}} + \tau_{\text{lep}}\tau_{\text{had}}$  had combined data fit with all statistical and systematic uncertainties (stat+syst).

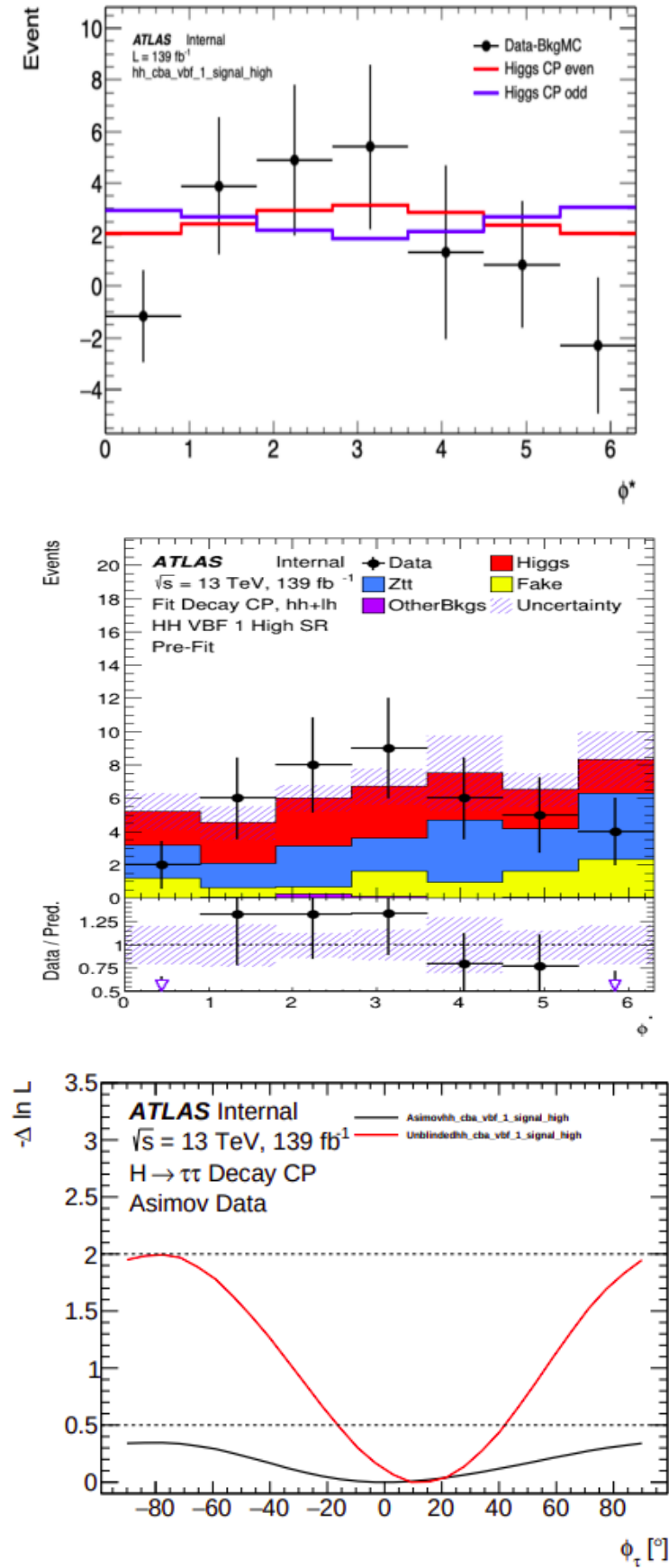


Figure 8.16: Data - background prefit distribution for the HH VBF High region, data and background distribution and the relative  $\Delta LNN$  distribution comparison between expected and observed.

# CONCLUSION

## 9

The thesis presents studies on the CP-properties of the Standard Model Higgs decaying into two  $\tau$  leptons. The analysis includes both semi-leptonic  $\tau_{\text{lep}}\tau_{\text{had}}$  and full hadronic  $\tau_{\text{had}}\tau_{\text{had}}$  channels, where the CP-violating interaction between the Higgs and the final states is studied by a mixing parameter  $\phi_\tau$  appearing in the Yukawa interaction.

The search is performed by using  $139 \text{ fb}^{-1}$  of the proton-proton collision data collected at LHC at a center-of-mass energy of  $\sqrt{s} = 13 \text{ TeV}$  between 2015 and 2018. For the Asimov results all signal and background are simulated using Monte Carlo methods with the exception of jets which are wrongly identified as  $\tau$  leptons as well as QCD multijet and top production that are estimated using a data driven method called Fake-Factor. The data-driven method is applied to both semi-leptonic and full hadronic channels.

To measure the CP properties, a maximum likelihood fit is performed to the analysis observable  $\varphi_{CP}^*$  which is calculated from the visible decay products of the Higgs decaying into a pair of  $\tau_{\text{had}}\tau_{\text{had}}$  or  $\tau_{\text{lep}}\tau_{\text{had}}$ . The  $\varphi_{CP}^*$  construction depends on how many charged and neutral particles are present in the final states. The analysis makes use of different kinematic selections for the different decay modes which improves the sensitivity to the observable even though the precision is constrained by the statistical uncertainties.

The observed (expected) limit on  $\phi_\tau$  is  $8.68^\circ \pm 16^\circ$  ( $0^\circ \pm 28^\circ$ ) at the 68% CL, and  $\pm 34^\circ$  ( $-70^\circ / +75^\circ$ ) at the 95.5% CL, as shown in table 9.1. The pure CP-odd hypothesis is rejected at  $3.4\sigma$  (standard deviation). The results achieved confirm that the Standard Model expectation for  $\phi_\tau$  equal to zero is consistent with the observation.

A similar and recent analysis of the CP structure of the Yukawa coupling involving the Standard Model Higgs and  $\tau$  leptons was also performed with the other general purpose LHC detector CMS [179]. The analysis is also performed using the  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  channels. The CMS data results exclude the pure CP-odd scenario at  $3.0\sigma$ , with an expected exclusion assuming the SM Higgs of  $2.6\sigma$ . The CP-mixing angle expected value

is found to be  $0^\circ \pm 21^\circ$  at the 63% confidence level and  $0^\circ \pm 49^\circ$  at 95.5%, as shown in table 9.1, while the observed value is  $-1^\circ \pm 19^\circ$  at 68% CL and  $\pm 41^\circ$  at 95% CL. The values found confirm that the results achieved are compatible with the SM predictions within the experimental uncertainties.

ATLAS			
Fit setting	CP-odd exclusion	Best Estimation $\phi_\tau(^{\circ})$	$1\sigma$ limit ( $^{\circ}$ )
Asimov	$2.10\sigma$	0.00	+29.13 / -27.65
unblinded	$3.41\sigma$	8.68	+16.77 / -16.16
CMS			
Fit setting	CP-odd exclusion	Best Estimation $\phi_\tau(^{\circ})$	$1\sigma$ limit ( $^{\circ}$ )
Asimov	$2.6\sigma$	0.00	$\pm 21$
unblinded	$3.0\sigma$	-1	$\pm 19$

Table 9.1: CP-odd exclusion limits comparison between ATLAS and CMS for both Monte Carlo samples and data samples.

The studies presented in this thesis are accepted by the The European Physical Journal C (EPJC) and available in [43]. Looking to the future, the use of the full Run-3 data is desirable to improve the sensitivity of the measurement and to improve the CP-odd exclusion to higher values of standard deviation.

# APPENDIX A

## 10

### 10.1 $\varphi_{CP}^*$ distributions at preselection level

Figures 10.1 10.2 10.3 10.4 show the  $\varphi_{CP}^*$  distributions, in high and low optimisation regions related to the  $0^\circ$  and  $90^\circ$  templates, while figures 10.5 10.6 10.7 10.8 show similar  $\varphi_{CP}^*$  distributions in high and low optimisation regions related to  $40^\circ$  and  $140^\circ$  templates.

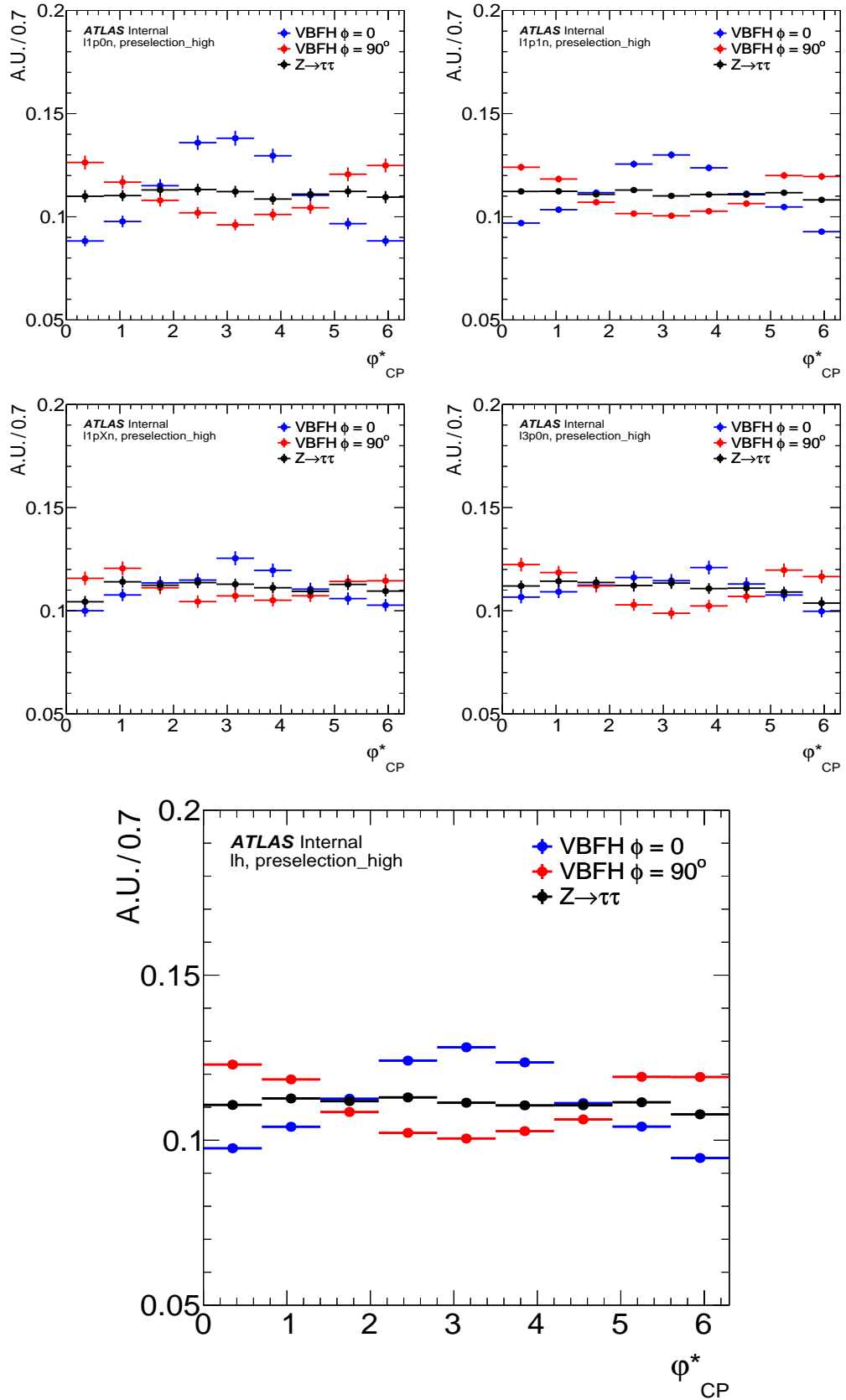


Figure 10.1: Semi-leptonic  $\varphi_{CP}^*$  distribution, for  $0^\circ$  and  $90^\circ$  degree templates, of different CP-hypotheses in the di- $\tau$  mode combination for the high optimisation regions.

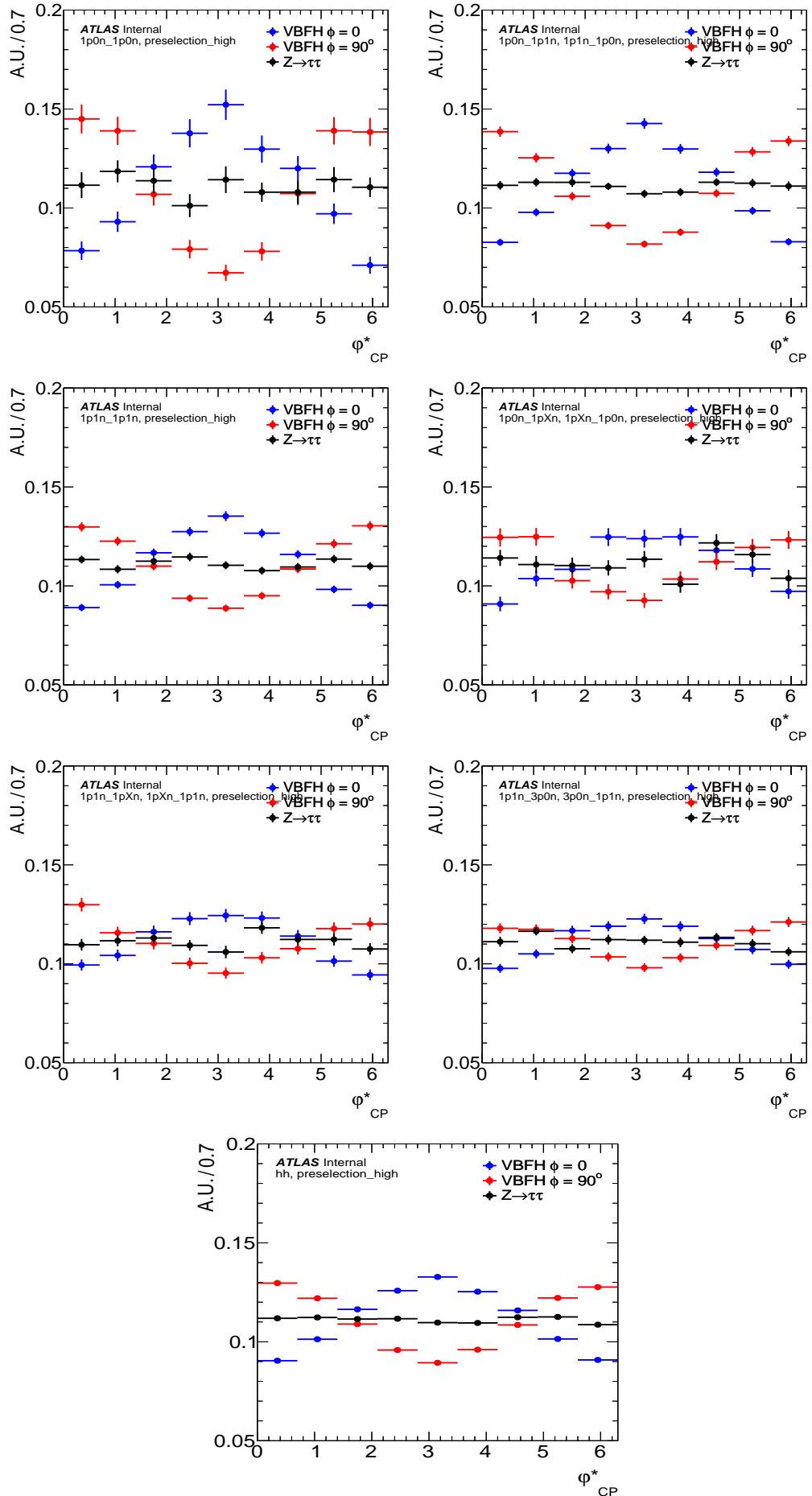


Figure 10.2: Fully hadronic  $\varphi_{CP}^*$  distribution, for  $0^\circ$  and  $90^\circ$  degree templates, of different CP-hypotheses in the di- $\tau$  mode combination for the high optimisation regions.



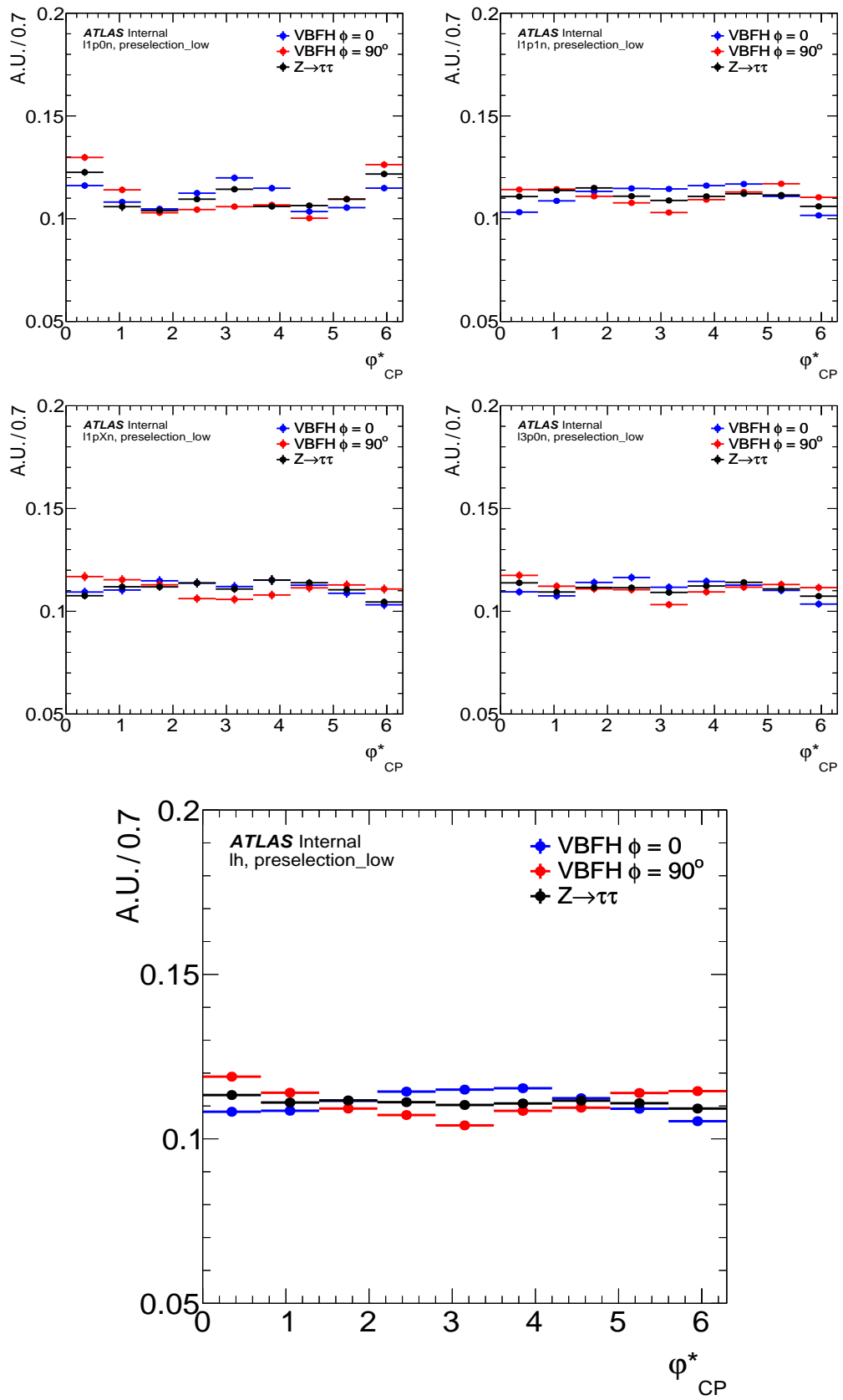


Figure 10.3: Semi-leptonic  $\varphi_{CP}^*$  distribution, for  $0^\circ$  and  $90^\circ$  degree templates, of different CP-hypotheses in the di- $\tau$  mode combination for the low optimisation regions.

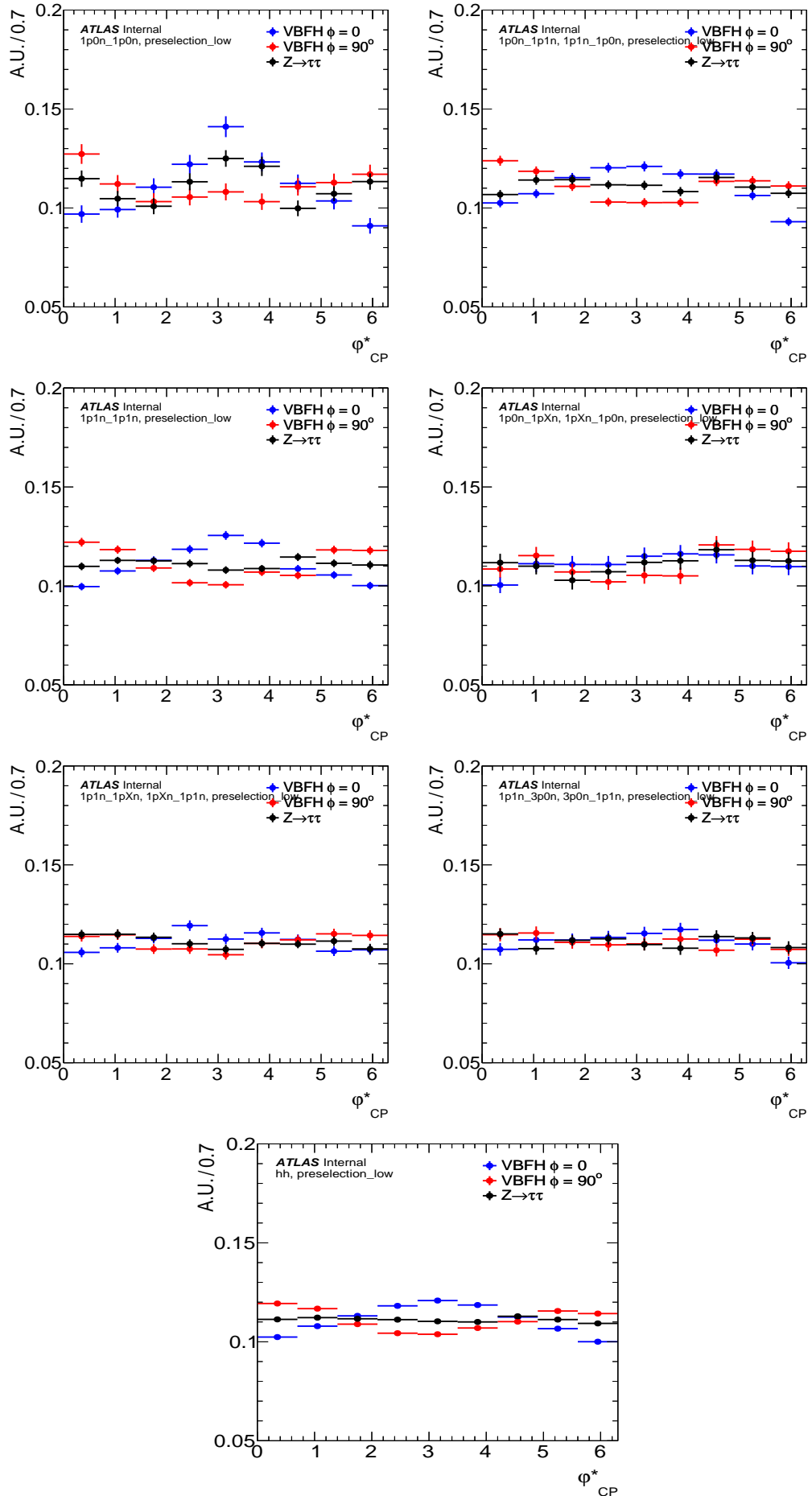


Figure 10.4: Fully hadronic  $\varphi_{CP}^*$  distribution, for  $0^\circ$  and  $90^\circ$  degree templates, of different CP-hypotheses in the di- $\tau$  mode combination for the low optimisation regions.

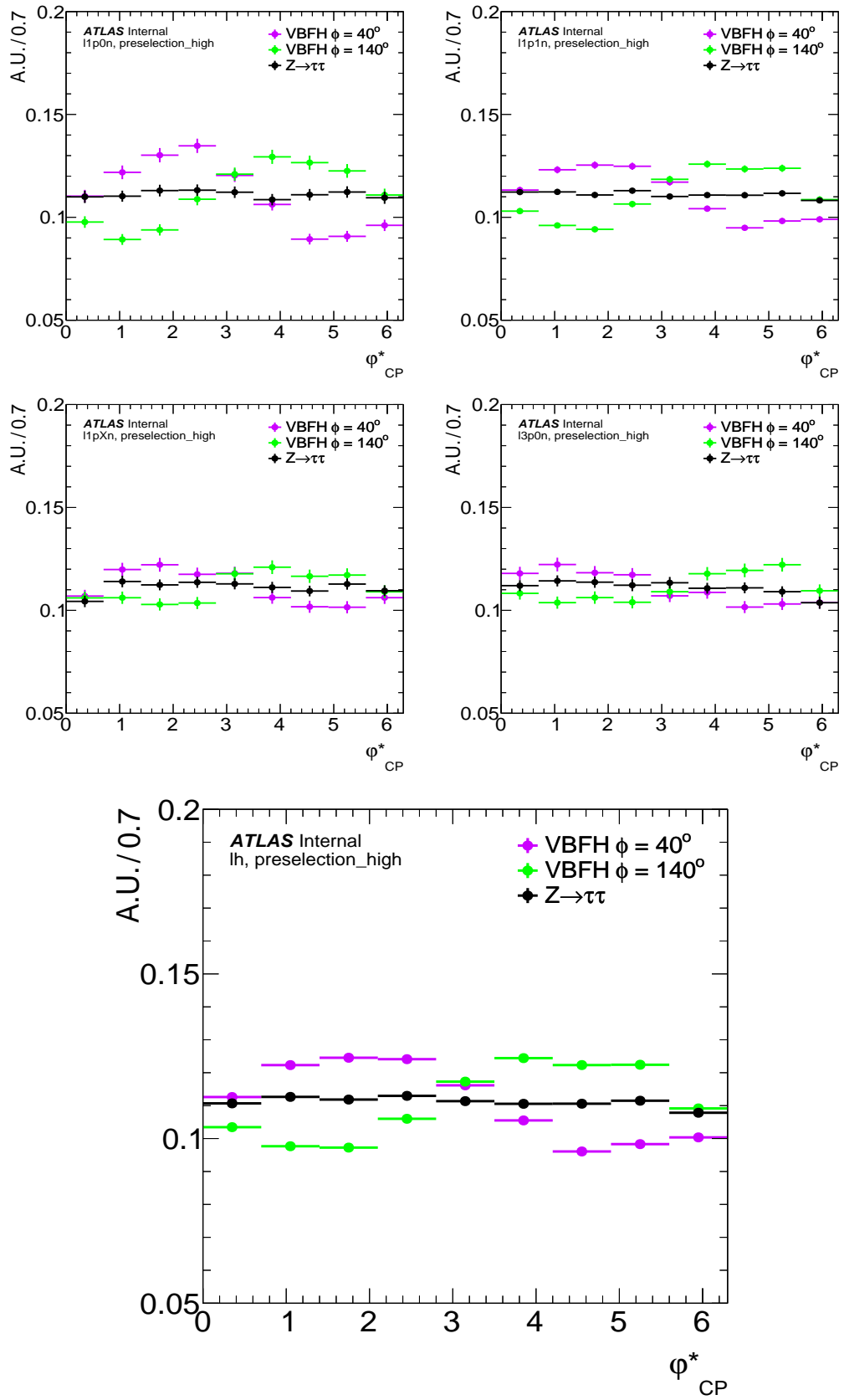


Figure 10.5: Semi-leptonic  $\varphi_{CP}^*$  distribution, for  $40^\circ$  and  $140^\circ$  degree templates, of different CP-hypotheses in the di- $\tau$  mode combination for the high optimisation regions.

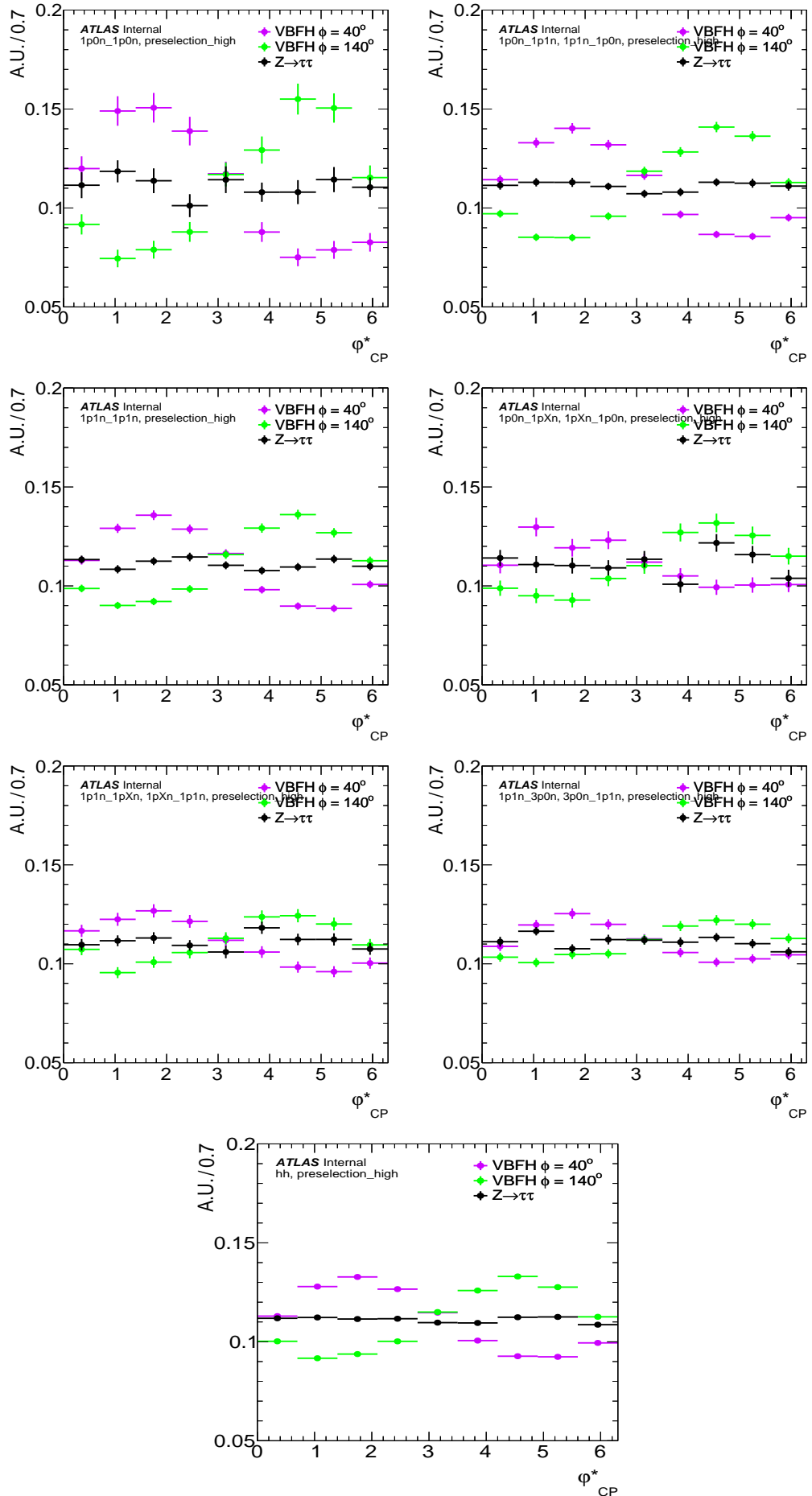


Figure 10.6: Fully hadronic  $\varphi_{CP}^*$  distribution, for  $40^\circ$  and  $140^\circ$  degree templates, of different CP-hypotheses in the di- $\tau$  mode combination for the high optimisation regions.

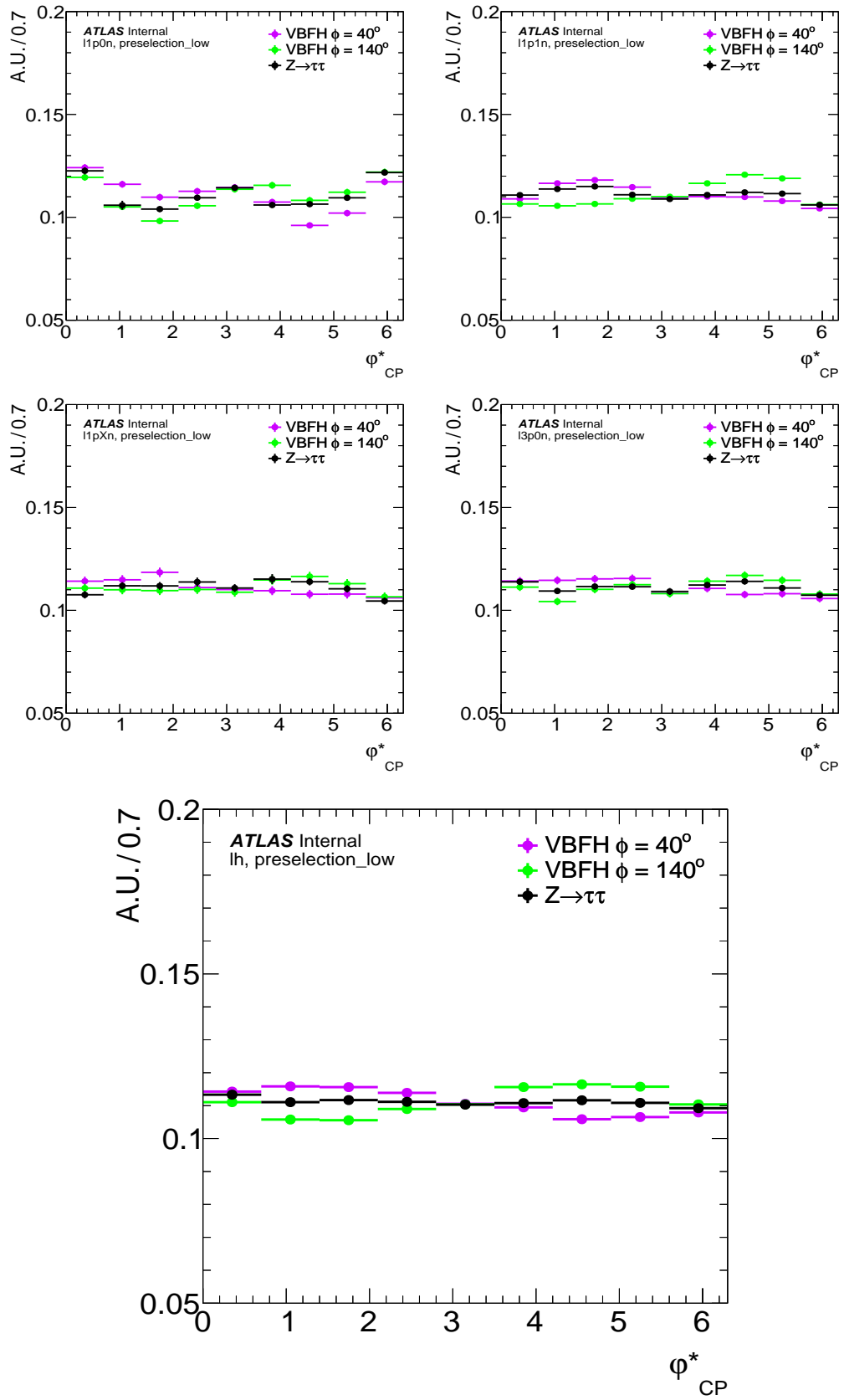


Figure 10.7: Semi-leptonic  $\varphi_{CP}^*$  distribution, for  $40^\circ$  and  $140^\circ$  degree templates, of different CP-hypotheses in the di- $\tau$  mode combination for the low optimisation regions.

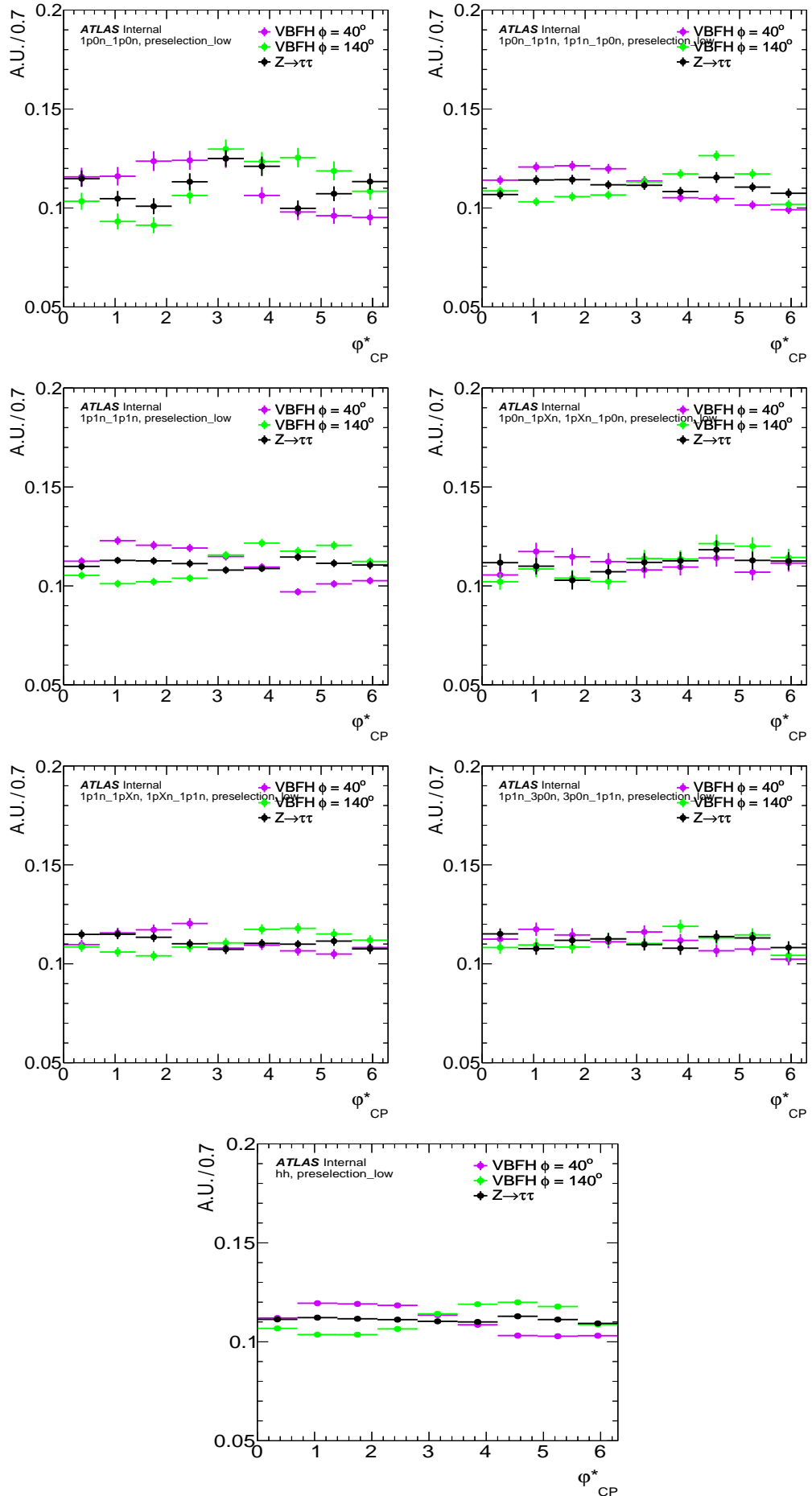
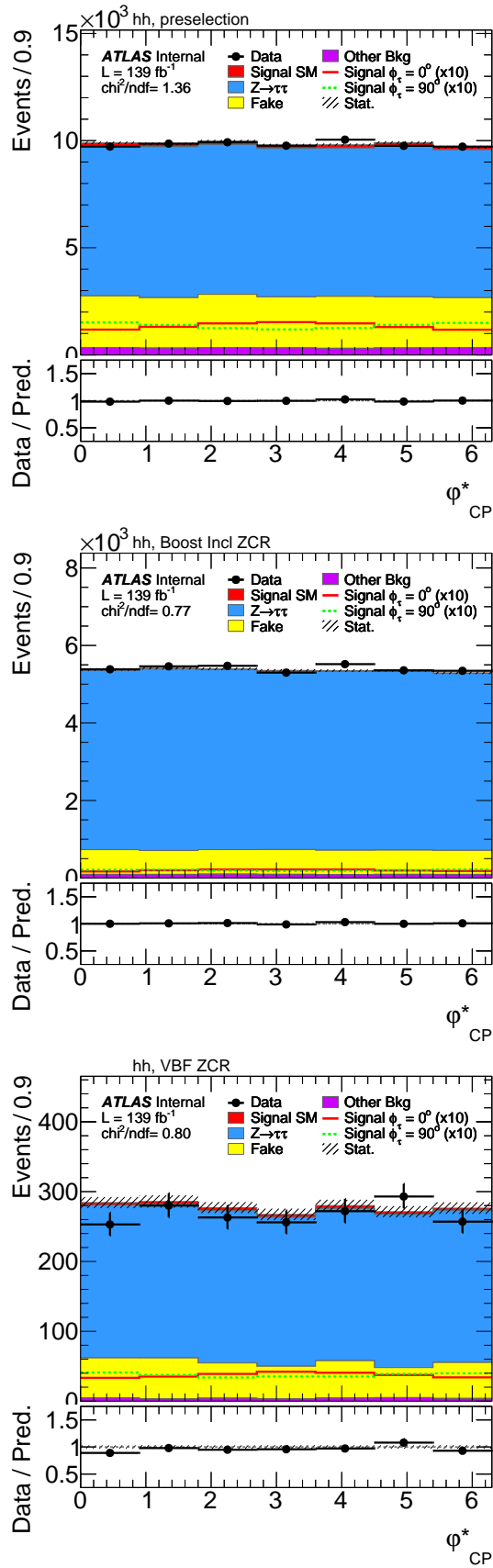


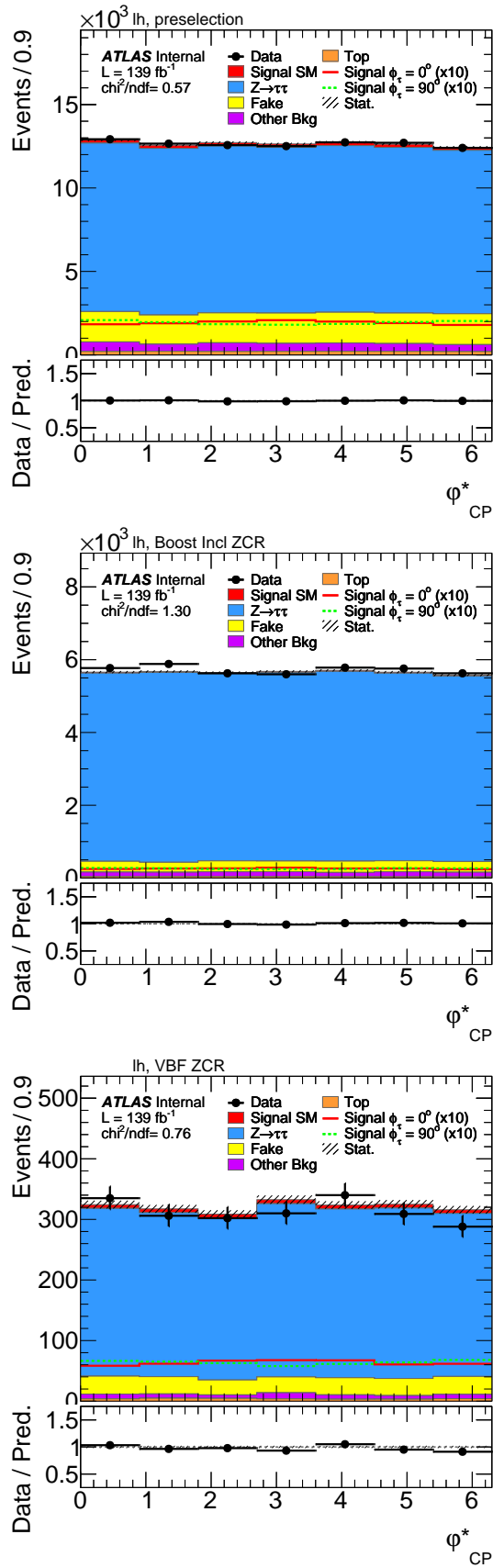
Figure 10.8: Fully hadronic  $\varphi_{CP}^*$  distribution, for  $40^\circ$  and  $140^\circ$  degree templates, of different CP-hypotheses in the di- $\tau$  mode combination for the low optimisation regions.

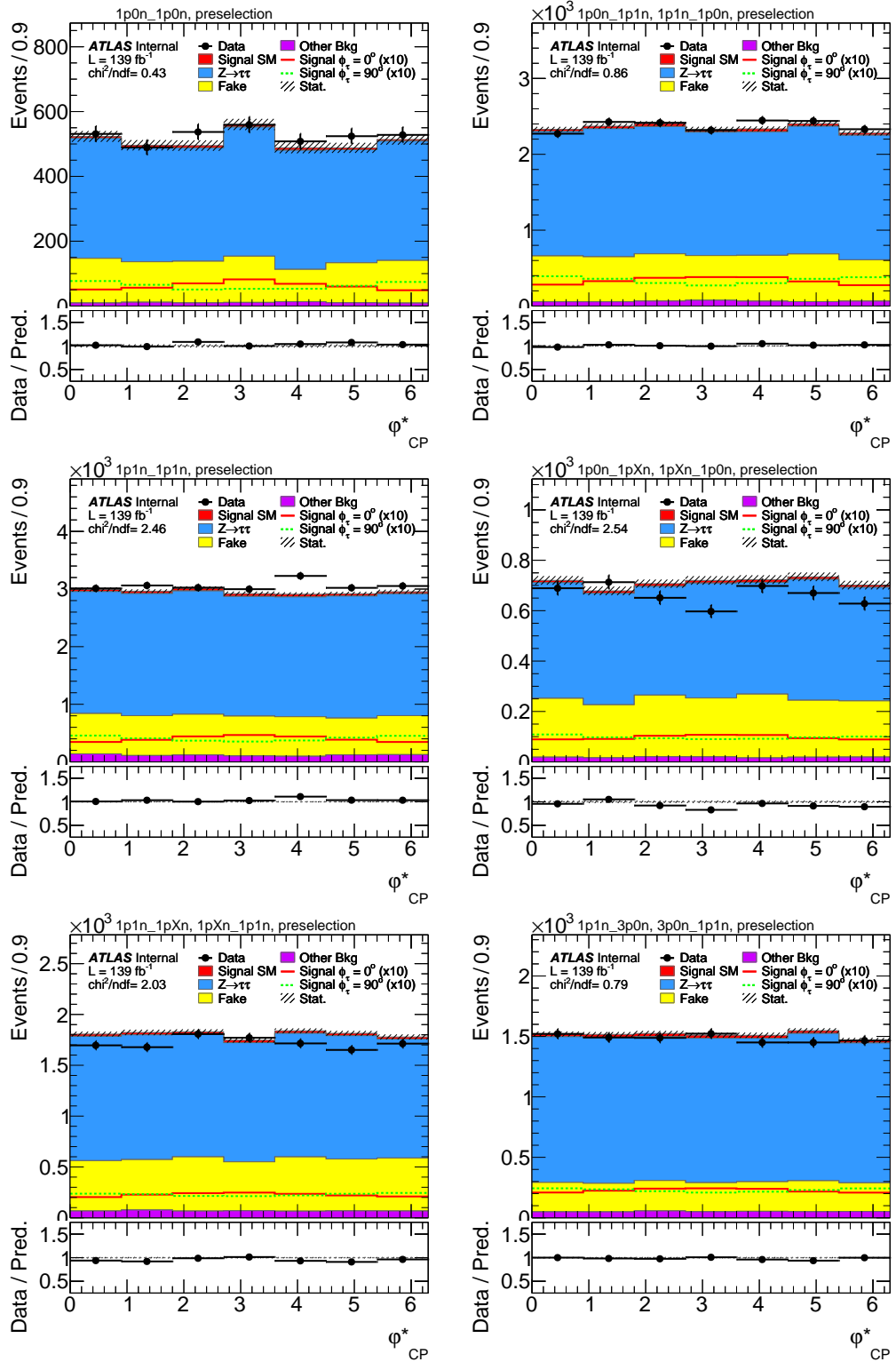
## 10.2 ZCR $\varphi_{CP}^*$ distributions at preselection level

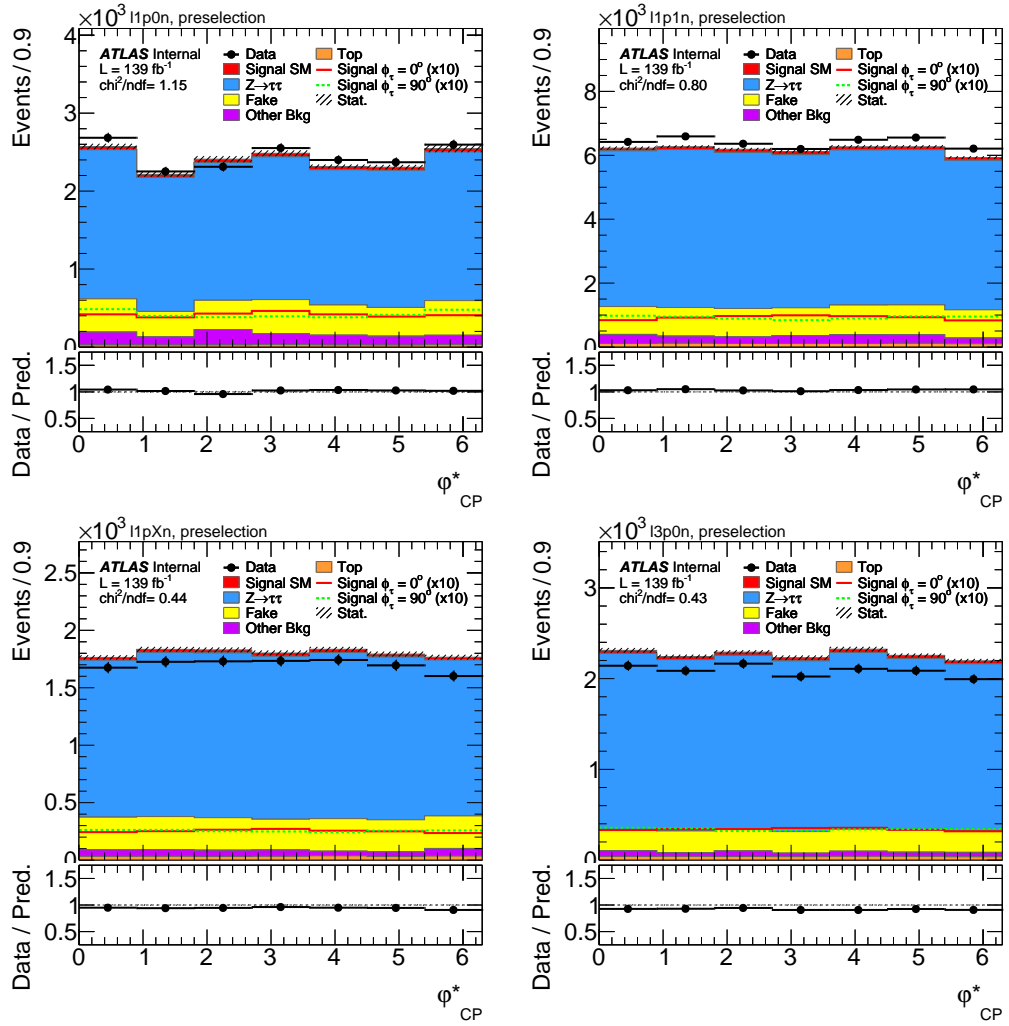
The modelling of the reconstructed  $\varphi_{CP}^*$  observables [6], in the preselection and Z control regions, is shown for the  $\tau_{had}\tau_{had}$  and the  $\tau_{lep}\tau_{had}$  channels in figure 10.9 and 10.10, respectively. The  $\phi^*$  shape is shown in preselection events for the hadronic decay channel in figure 10.11 and for the leptonic decay channel in figure 10.12. Figures 10.13 10.14 and figures 10.15 10.16 show, respectively, the  $\varphi_{CP}^*$  shape by decay channels in preselection events in the boost Z control region and in the VBF Z control region.

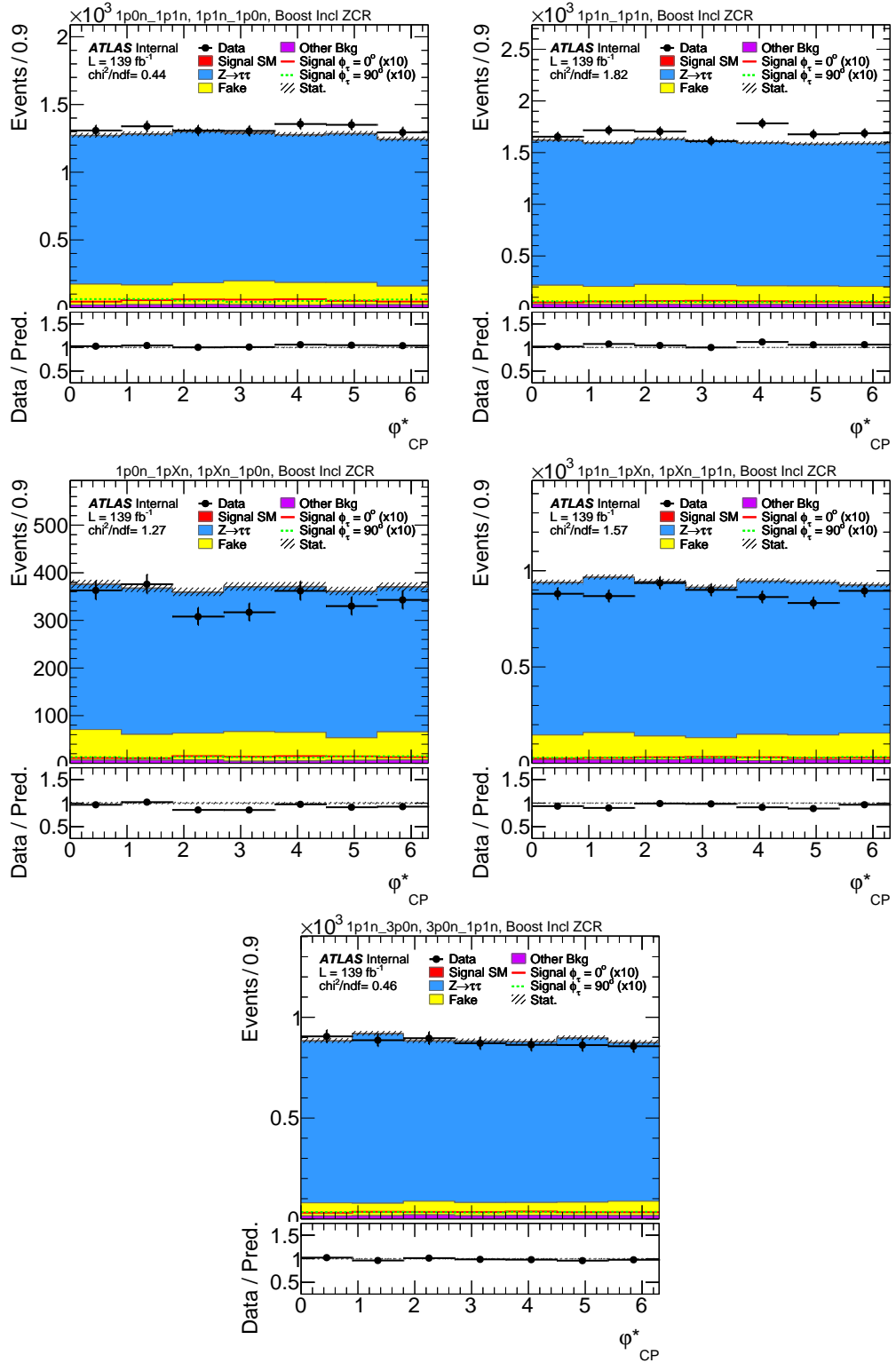
Figure 10.9:  $\phi^*$  in preselection and Z control regions in the  $\tau_{had}\tau_{had}$  channel.

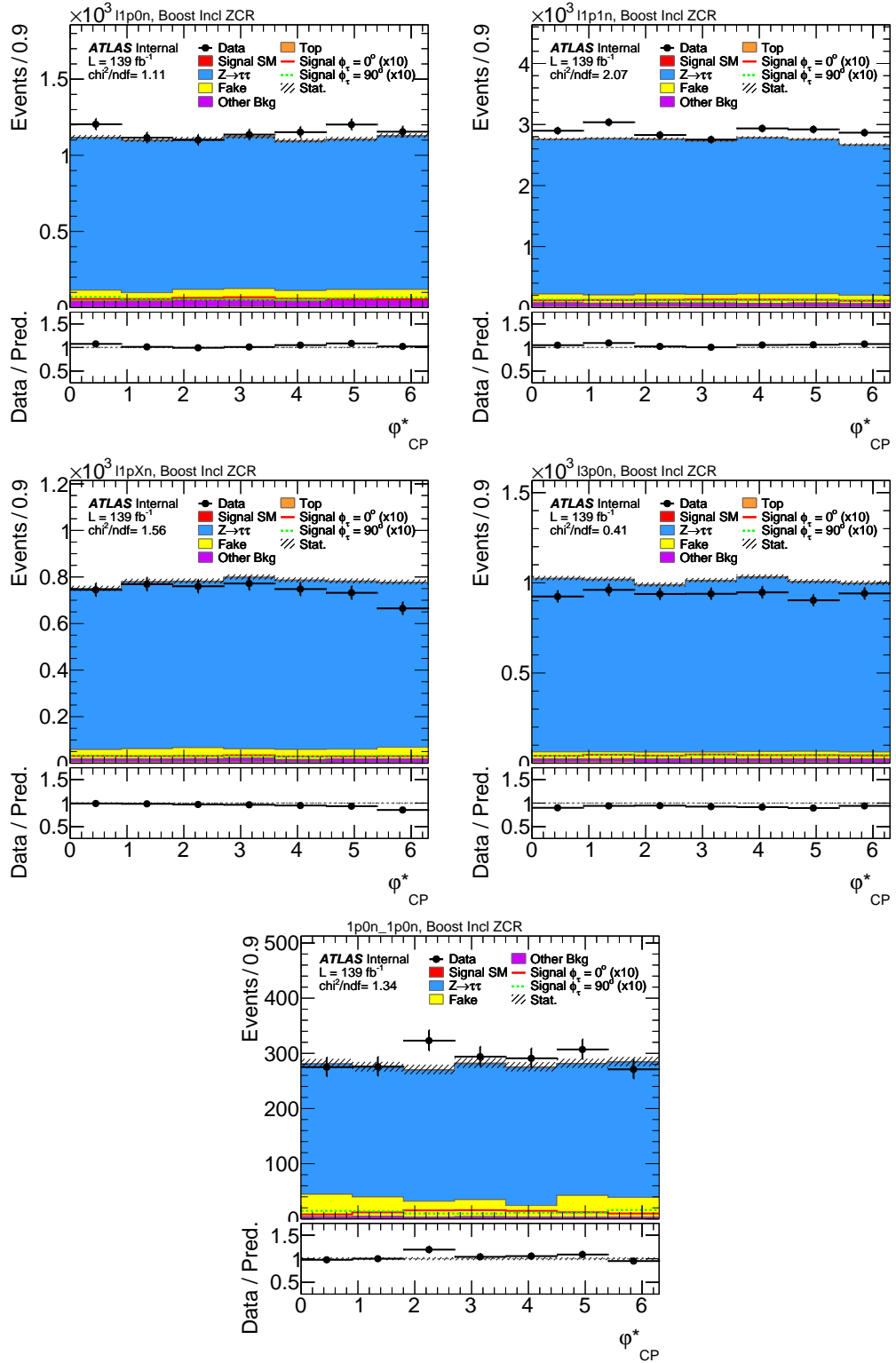


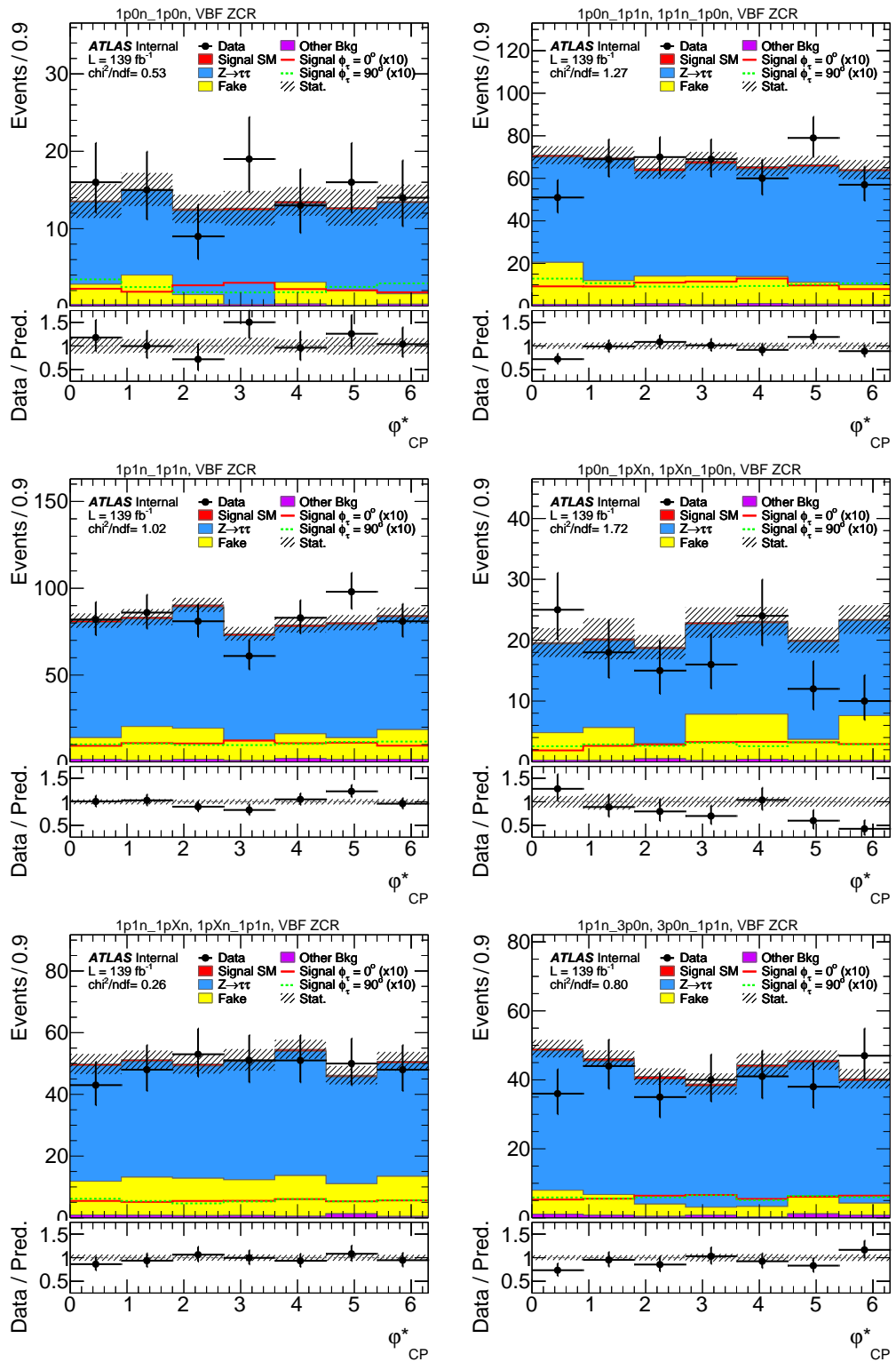
Figure 10.10:  $\phi^*$  in preselection and Z control regions in the  $\tau_{lep}\tau_{had}$  channel.

Figure 10.11:  $\phi_{CP}^*$  in different decay channels in preselection events in the  $\tau_{had}\tau_{had}$  channel.

Figure 10.12:  $\phi_{CP}^*$  in different decay channels in preselection events in the  $\tau_{lep}\tau_{had}$  channel.

Figure 10.13:  $\phi^*$  by decay channels in the boost Z control region in the  $\tau_{had}\tau_{had}$  channel.

Figure 10.14:  $\phi^*$  by decay channels in the boost Z control region in the  $\tau_{lep}\tau_{had}$  channel.

Figure 10.15:  $\phi^*$  by decay channels in the VBF Z control region in the  $\tau_{had}Thad$  channel.

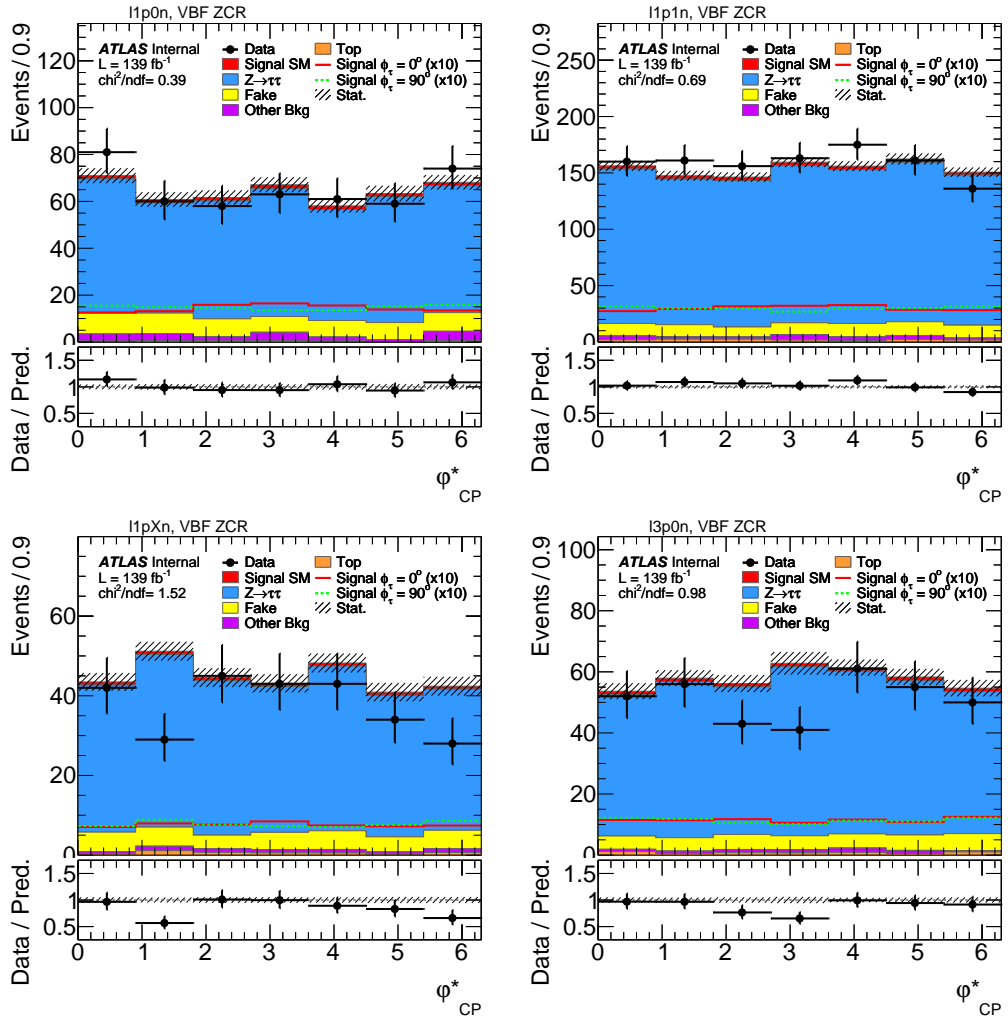


Figure 10.16:  $\phi^*$  by decay channels in the VBF Z control region in the  $\tau_{lep}\tau_{had}$  channel.

## 10.3 Observed post-fit distributions

$\phi_{CP}^*$  observed post-fit distributions are shown in figure 10.17 and figure 10.18, respectively, in all VBF and Boost signal regions, and in figure 10.19 for the  $Z \rightarrow \tau\tau$  control regions [6].

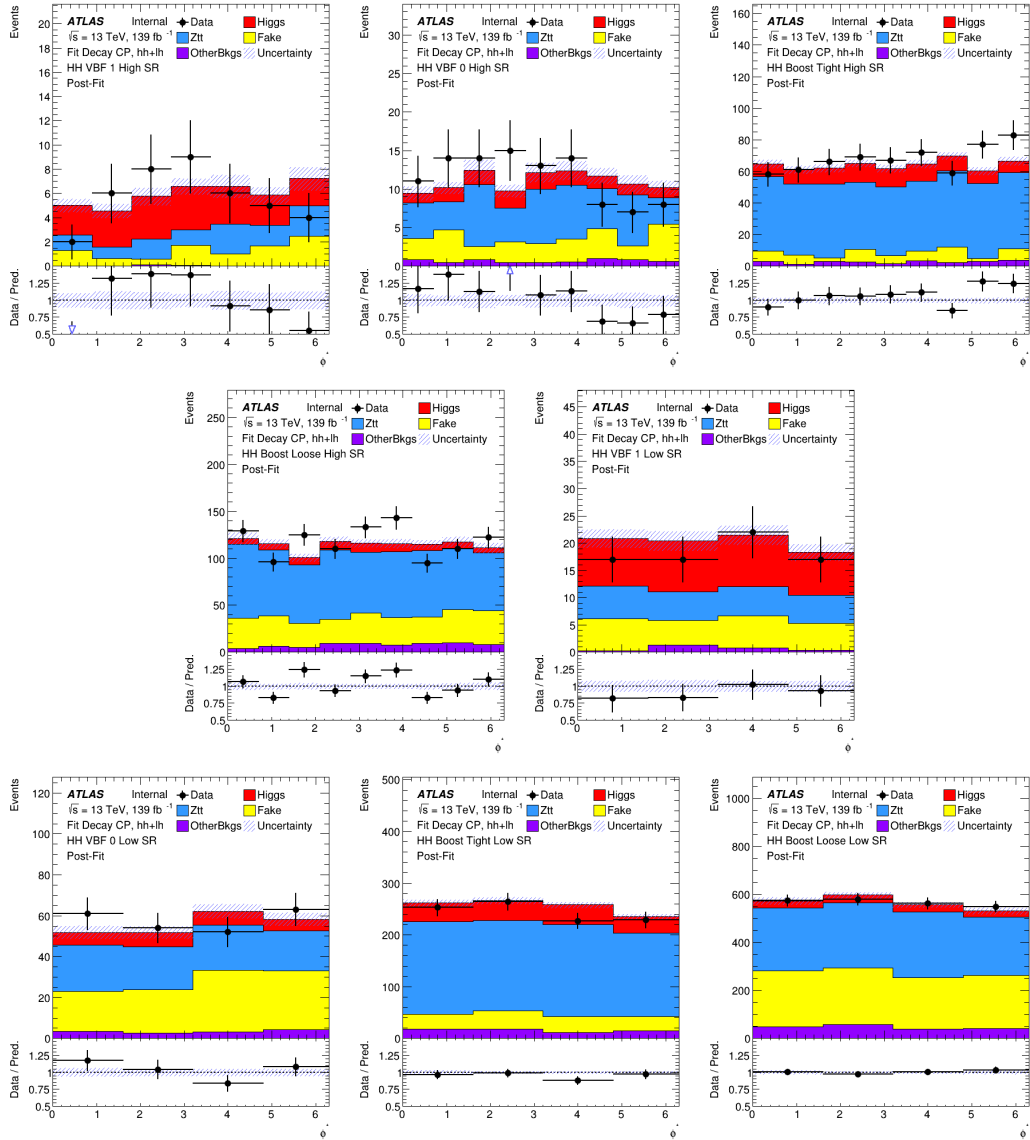


Figure 10.17: Observed post-fit distributions in the  $H \rightarrow \tau\tau$  signal regions in  $\mathcal{T}_{had}\mathcal{T}_{had}$  channels, which will be included in the likelihood model. ‘OtherBkgs’ includes  $W$ , diboson, top,  $Z \rightarrow ll$ ,  $H \rightarrow WW$  samples.



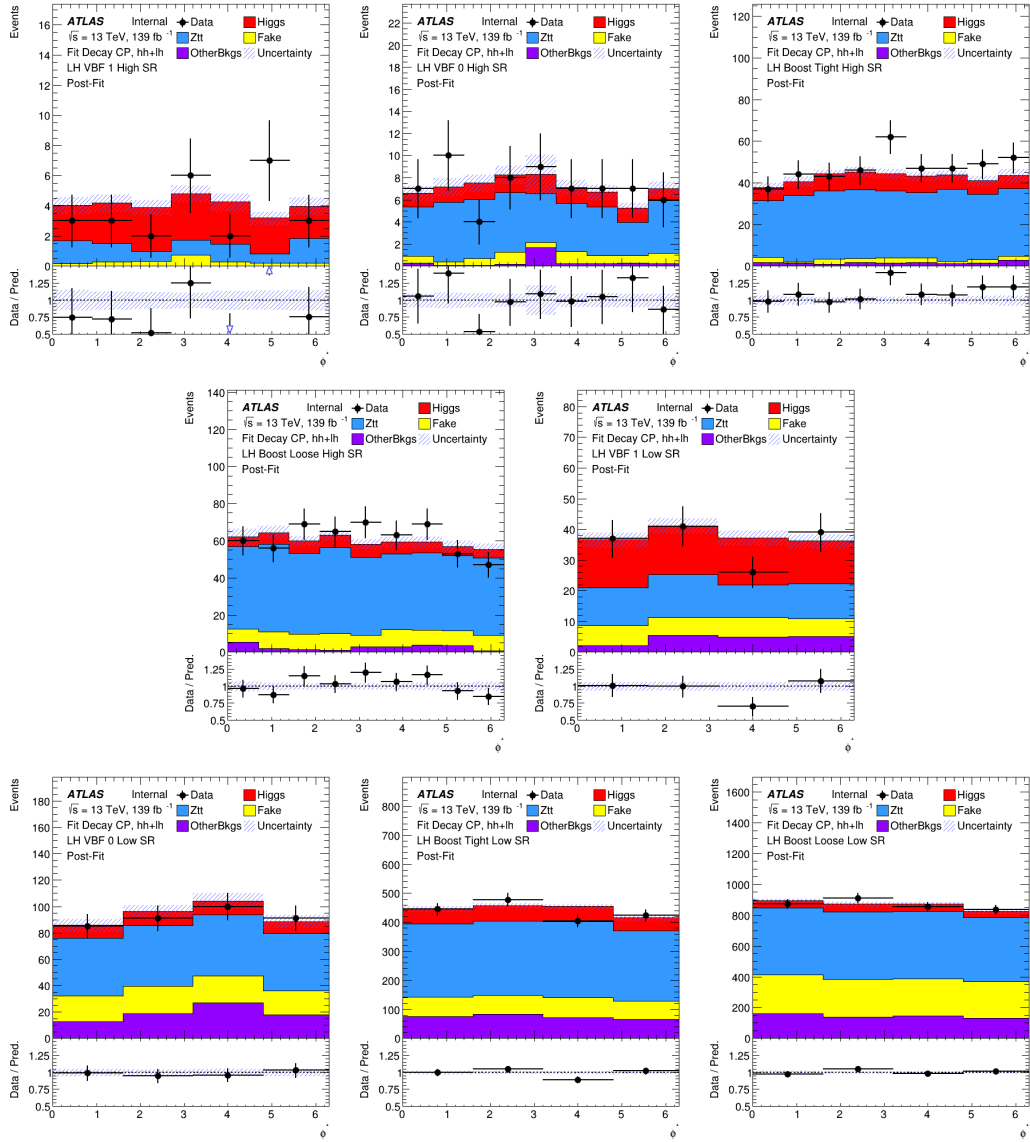


Figure 10.18: Observed post-fit distributions in the  $H \rightarrow \tau\tau$  signal regions in  $\pi_{lep}\tau_{had}$  channels, which will be included in the likelihood model. ‘OtherBkgs’ includes  $W$ , diboson, top,  $Z \rightarrow ll$ ,  $H \rightarrow WW$  samples.

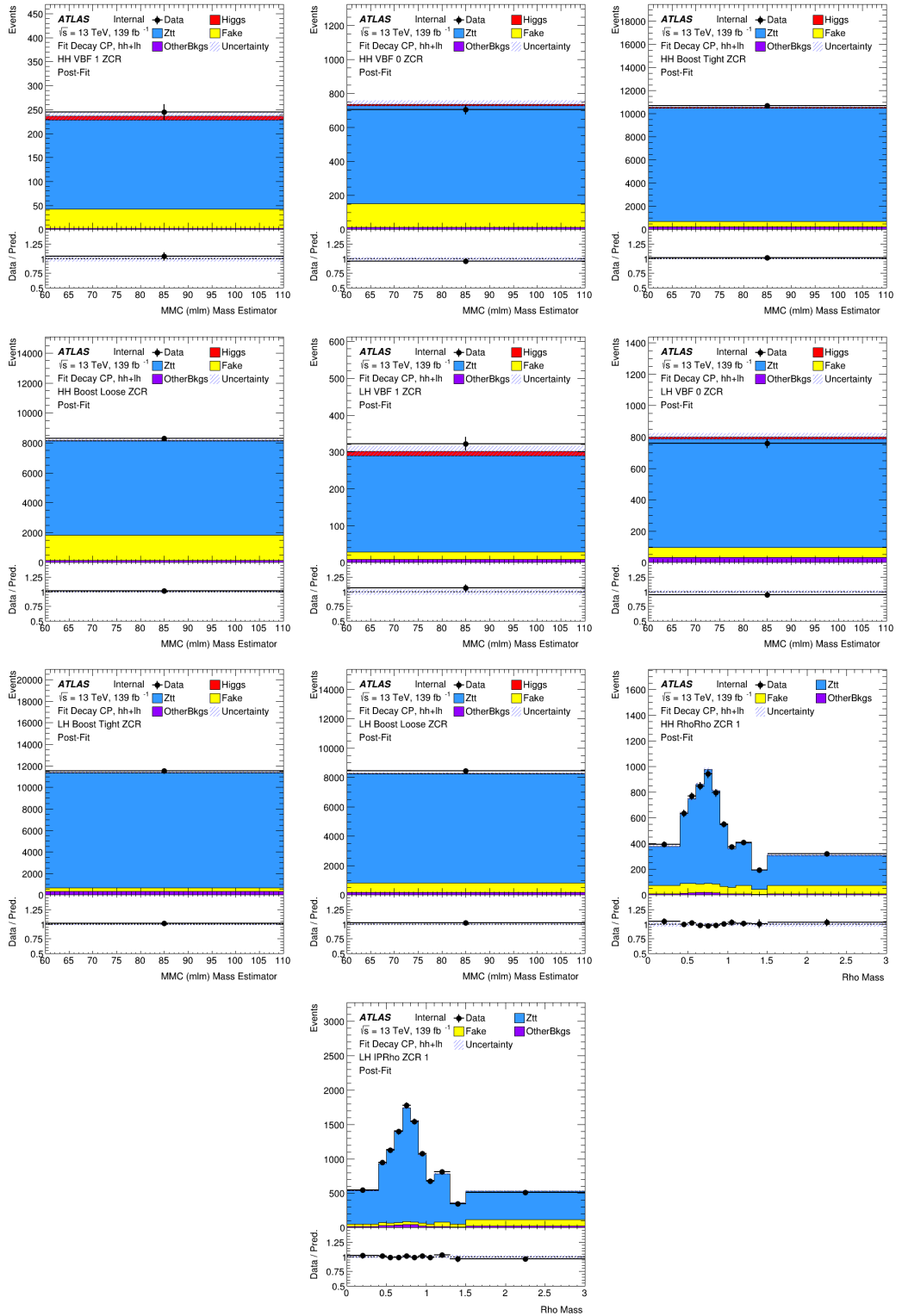


Figure 10.19: Observed post-fit distributions in the  $Z(\rightarrow \tau\tau)$  control regions in  $\tau_{had}\tau_{had}$  and  $\tau_{lep}\tau_{had}$  channels, which will be included in the likelihood model. ‘OtherBkgs’ includes  $W$ , diboson, top,  $Z \rightarrow ll, H \rightarrow WW$  samples.

# APPENDIX B

# 11

## 11.1 Lepton systematic uncertainties list

<b>Muon Resolution</b>	MUON-ID, MUON-SAGITTA-RHO, MUON-SAGITTA-RESBIAS, MUON-MS
<b>Energy scale</b>	MUON-SCALE
<b>Reconstruction/ Identification</b>	MUON-EFF-RECO-STAT, MUON-EFF-RECO-STAT-LOWPT, MUON-EFF-RECO-SYS, MUON-EFF-RECO-SYS-LOWPT
<b>Isolation</b>	MUON-EFF-ISO-STAT MUON-EFF-ISO-SYS
<b>Trigger</b>	MUON-EFF-TrigStatUncertainty, UON-EFF-TrigSystUncertainty

Table 11.1: Muon uncertainties list.

<b>Resolution</b>	EG-RESOLUTION-ALL
<b>Energy scale</b>	EG-SCALE-ALL
<b>Reconstruction</b>	EL-EFF-Reco-TOTAL-1NPCOR-PLUS-UNCOR
<b>Identification</b>	EL-EFF-ID-TOTAL-1NPCOR-PLUS-UNCOR
<b>Isolation</b>	EL-EFF-Iso-TOTAL-1NPCOR-PLUS-UNCOR
<b>Charge Identification</b>	EL-CHARGEID-STAT, EL-CHARGEID-SYS
<b>Trigger</b>	EL-EFF-Trigger-TOTAL-1NPCOR-PLUS-UNCOR

Table 11.2: Electron uncertainties list.

<b>Energy scale</b>	TAUS-TRUEHADTAU-SME-TES-INSITUEXP TAUS-TRUEHADTAU-SME-TES-INSITUFIT AUS-TRUEHADTAU-SME-TES-DETECTOR TAUS-TRUEHADTAU-SME-TES-MODEL-CLOSURE TAUS-TRUEHADTAU-SME-TES-PHYSICSLIST
<b>Reconstruction</b>	TAUS-TRUEHADTAU-EFF-RECO-TOTAL
<b>Identification</b>	TAUS-TRUEHADTAU-EFF-RNNID-[1]3]PRONGSTATSYSTPT[2025]2530]3040]GE40] TAUS-TRUEHADTAU-EFF-RNNID-SYST TAUS-TRUEHADTAU-EFF-RNNID-HIGHPT
<b>Trigger(2015-2017)</b>	AUS-TRUEHADTAU-EFF-TRIGGER[STATDATA161718]STATMC161718] TAUS-TRUEHADTAU-EFF-TRIGGER[SYST161718]SYSTMU161718]
<b>Trigger 2018</b>	TAUS-TRUEHADTAU-EFF-TRIGGER-[STATDATA2018]STATMC2018] TAUS-TRUEHADTAU-EFF-TRIGGER-[SYST2018]SYSTMU2018]
<b>eVeto on fake <math>\tau_{had-vis}</math></b>	TAUS-TRUEELECTRON-EFF-ELEBDT-STAT TAUS-TRUEELECTRON-EFF-ELEBDT-SYST

Table 11.3: Tau uncertainties list.

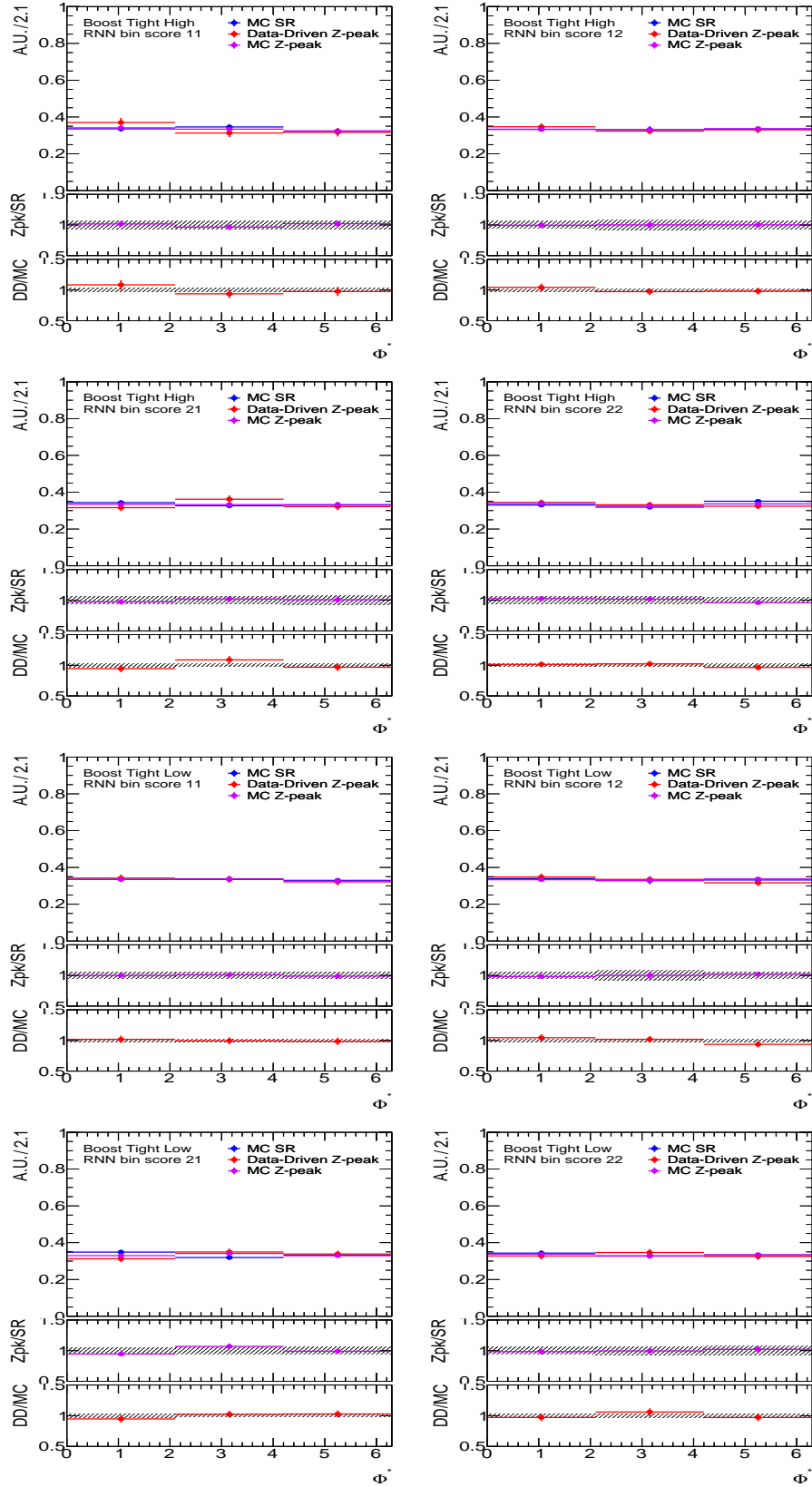
## 11.2 Jet systematic uncertainties list

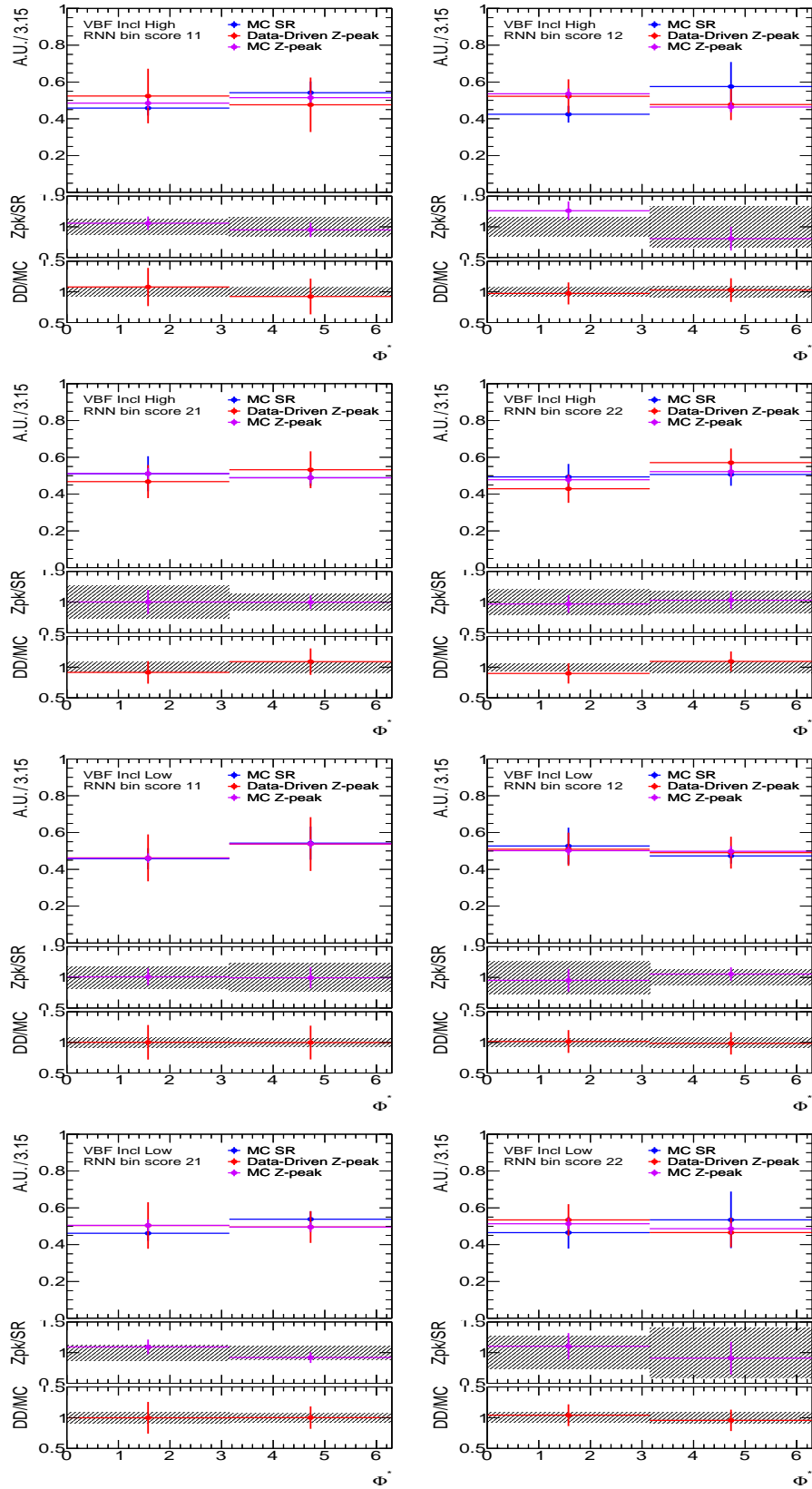
<b>Energy scale</b>	JET-EffectiveNP-Statistical[1-6], JET-PunchThrough-MC16 JET-BJES-Response JET-EtaIntercalibration-[Modelling Modelling-2018data TotalStat], JET-EtaIntercalibration-[NonClosure-highE NonClosure-negEta NonClosure-posEta], JET-SingleParticle-HighPt, JET-EffectiveNP-Mixed[1 2 3], JET-Pileup-[OffsetMu OffsetNPV PtTerm RhoTopology], JET-EffectiveNP-Modelling[1-4], JET-EffectiveNP-Detector[1-2]
<b>Jet Energy resolution</b>	JET-JER-EffectiveNP-[1-12restTerm], JET-JER-DataVsMC
<b>Jet Flavor</b>	JET-Flavor-Composition, JET-Flavor-Response
<b>Jet Vertex Tagger</b>	JET-JvtEfficiency (central jets), JET-fJvtEfficiency (forward jets)
<b>b-tagging</b>	FT-EFF-Eigen-B-[0-2], FT-EFF-Eigen-C-[0-2], FT-EFF-Eigen-Light-[0-3], FT-EFF-extrapolation, FT-EFF-extrapolation-from-charm

Table 11.4: JET uncertainties list.

### 11.3 $\phi^*$ distributions for the RNN shape uncertainties studies

The  $\phi^*$  distributions in the  $Z_{MC} \rightarrow \tau\tau$  SR and ZCR and the data-driven ZCR are shown in figures 11.1 and 11.2 for the Boost tight and VBF channels respectively.

Figure 11.1: Boost Tight  $\phi^*$  shape distributions shown for the different RNN bins.

Figure 11.2: VBF  $\phi^*$  shape distributions shown for the different RNN bins.



## 11.4 Primary vertex distributions studies

Primary vertex distribution compared among the different BS constraint for VBF odd sample 11.3, ggH even sample 11.4 and ggH odd sample 11.5.

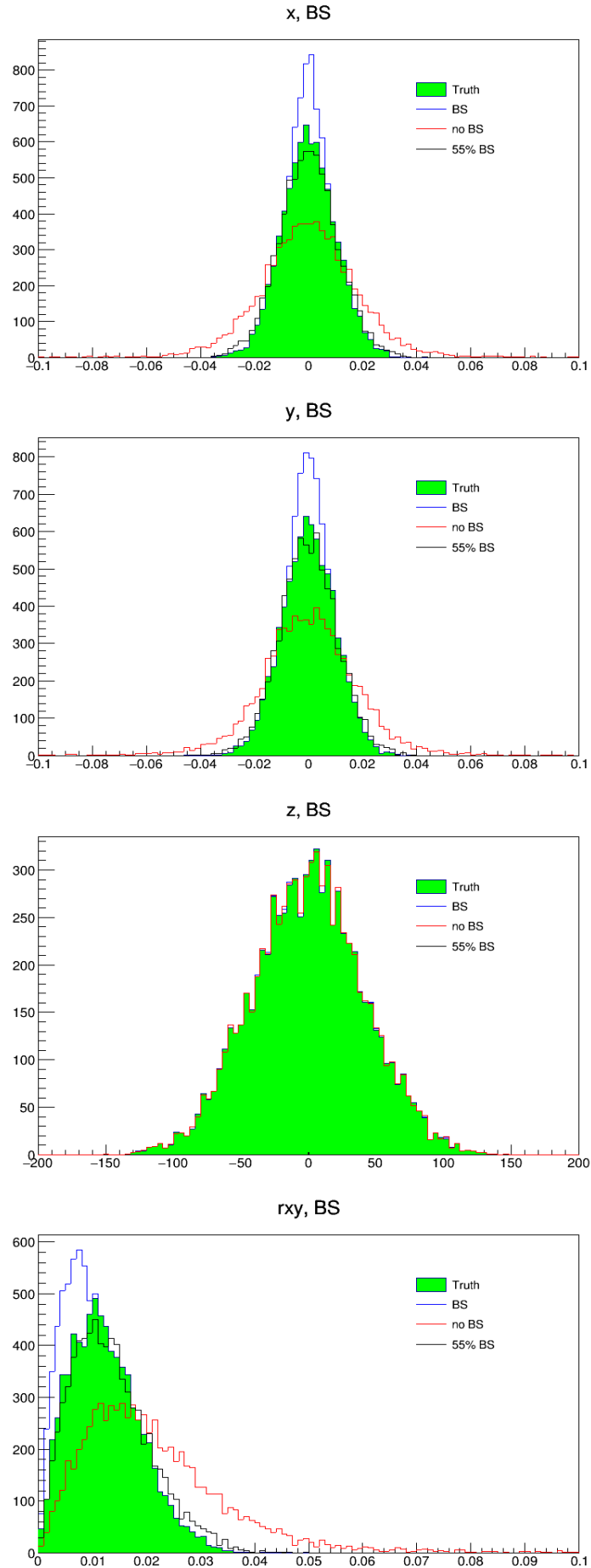


Figure 11.3: PV coordinates ( $x$ ,  $y$ ,  $z$ ) and the distance in the detector transverse plane ( $R_{xy}$ ) for the VBF CP-odd sample. Comparison between the MC values and the default BS constraint, no BS constraint and 55% of the BS constraint is performed.

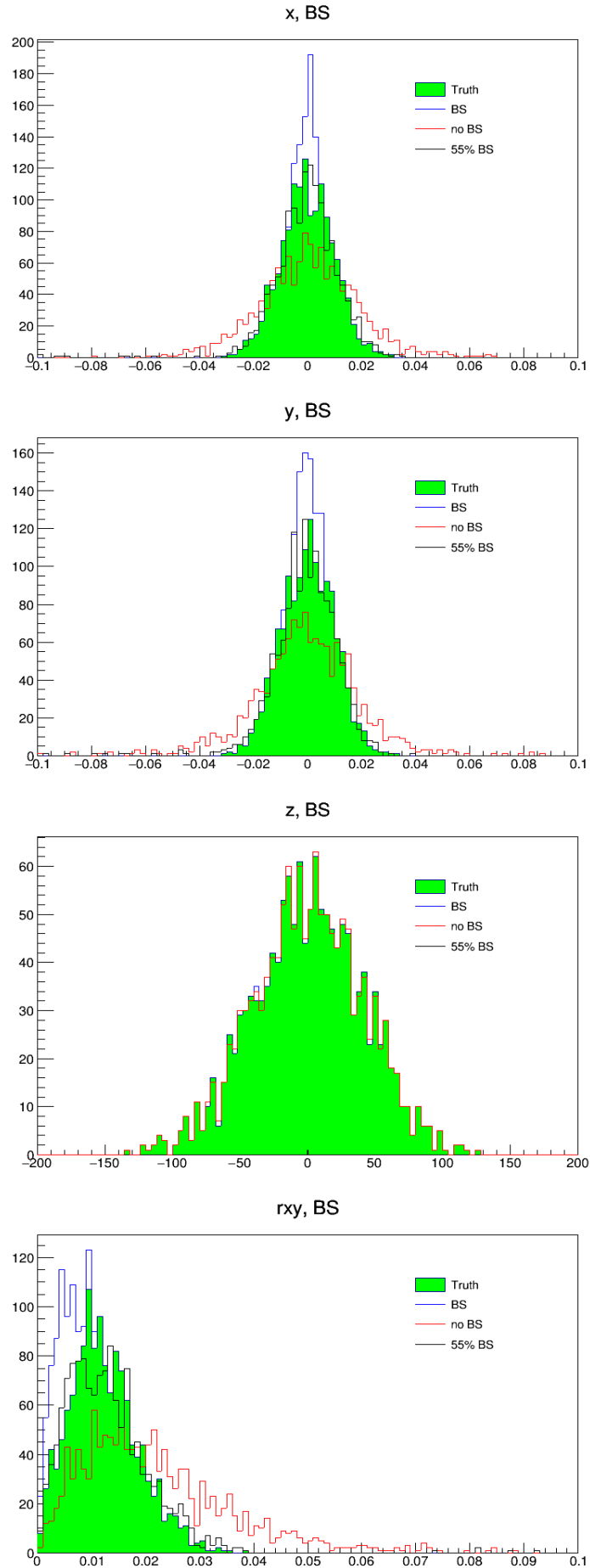


Figure 11.4: PV coordinates ( $x$ ,  $y$ ,  $z$ ) and the distance in the detector transverse plane ( $R_{xy}$ ) for the ggH CP-even sample. Comparison between the MC values and the default BS constraint, no BS constraint and 55% of the BS constraint is performed.

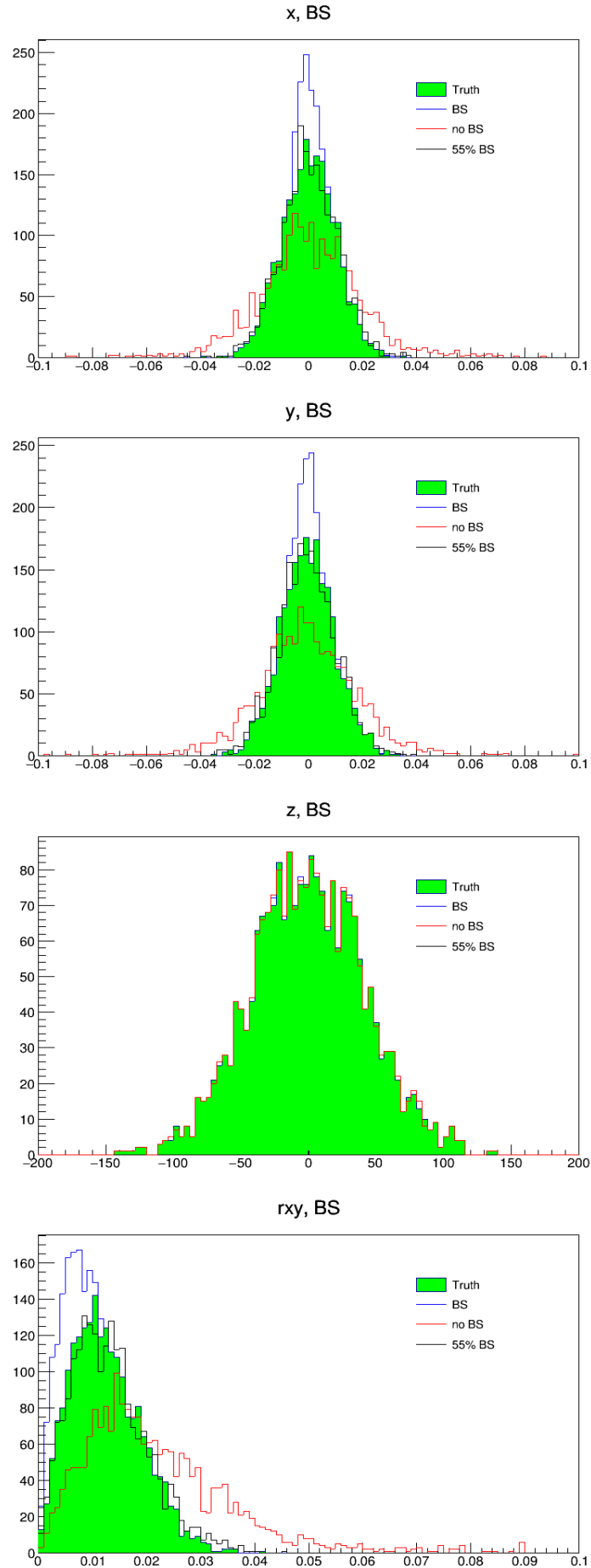


Figure 11.5: PV coordinates ( $x$ ,  $y$ ,  $z$ ) and the distance in the detector transverse plane ( $R_{xy}$ ) for the ggH CP-odd sample. Comparison between the MC values and the default BS constraint, no BS constraint and 55% of the BS constraint is performed.

## 11.5 Coordinate residual plots

Coordinate residual plots for the VBF odd sample in figure 11.6, ggH even sample in figure 11.7 and ggH odd sample in figure 11.8.

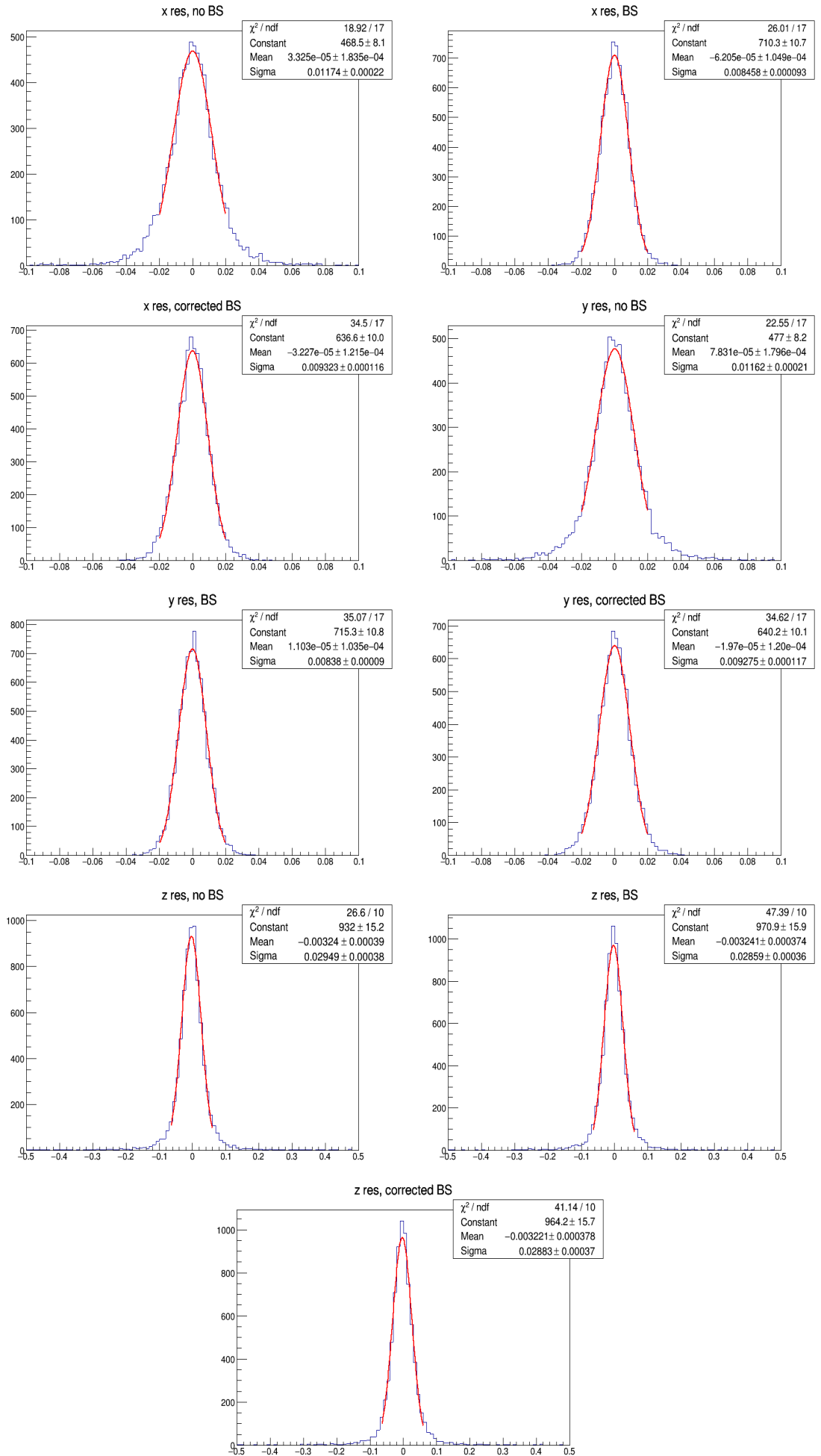


Figure 11.6: Coordinate residual plots of the VBF CP-odd sample.

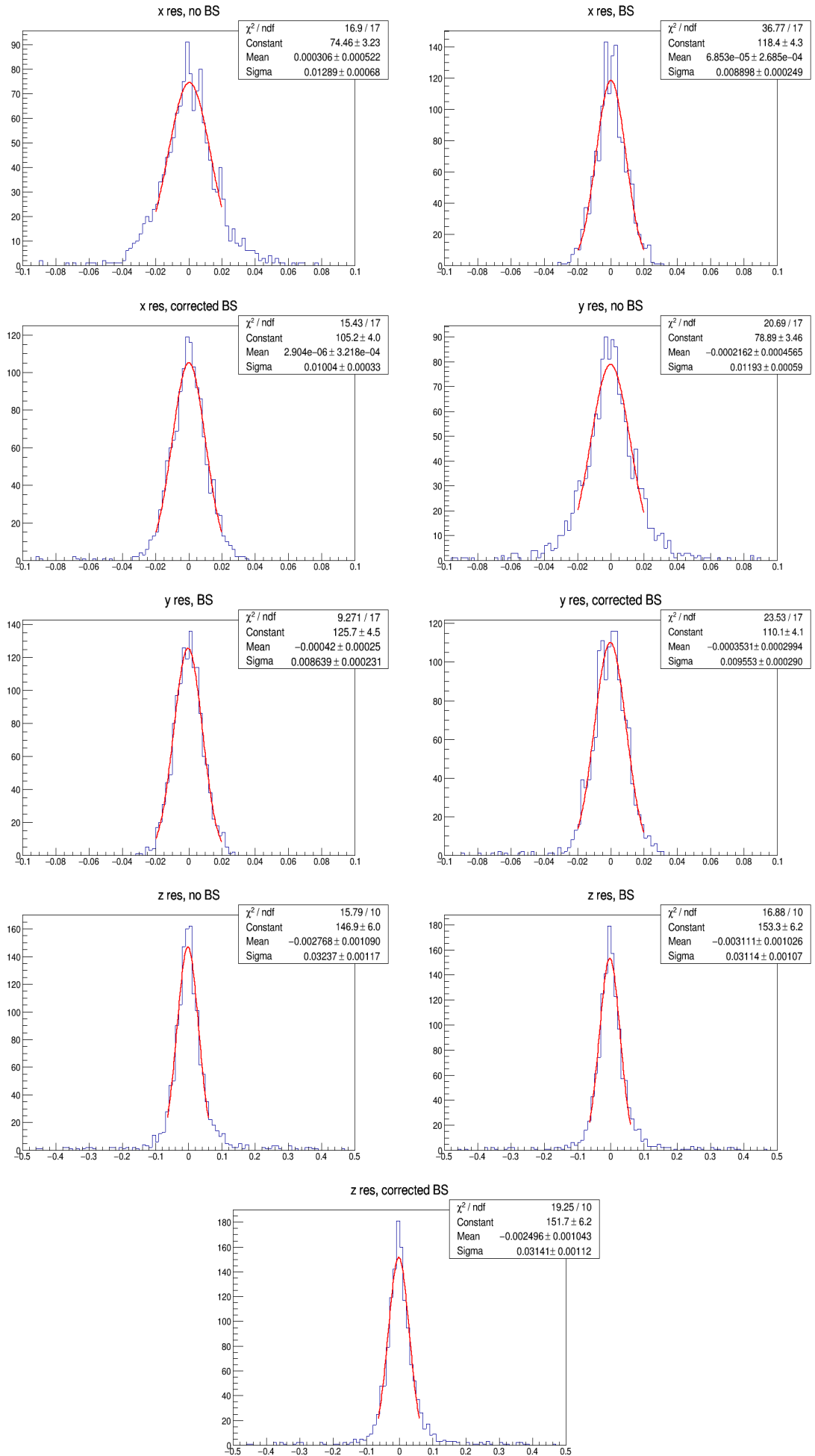


Figure 11.7: Coordinate residual plots of the ggH CP-even sample.

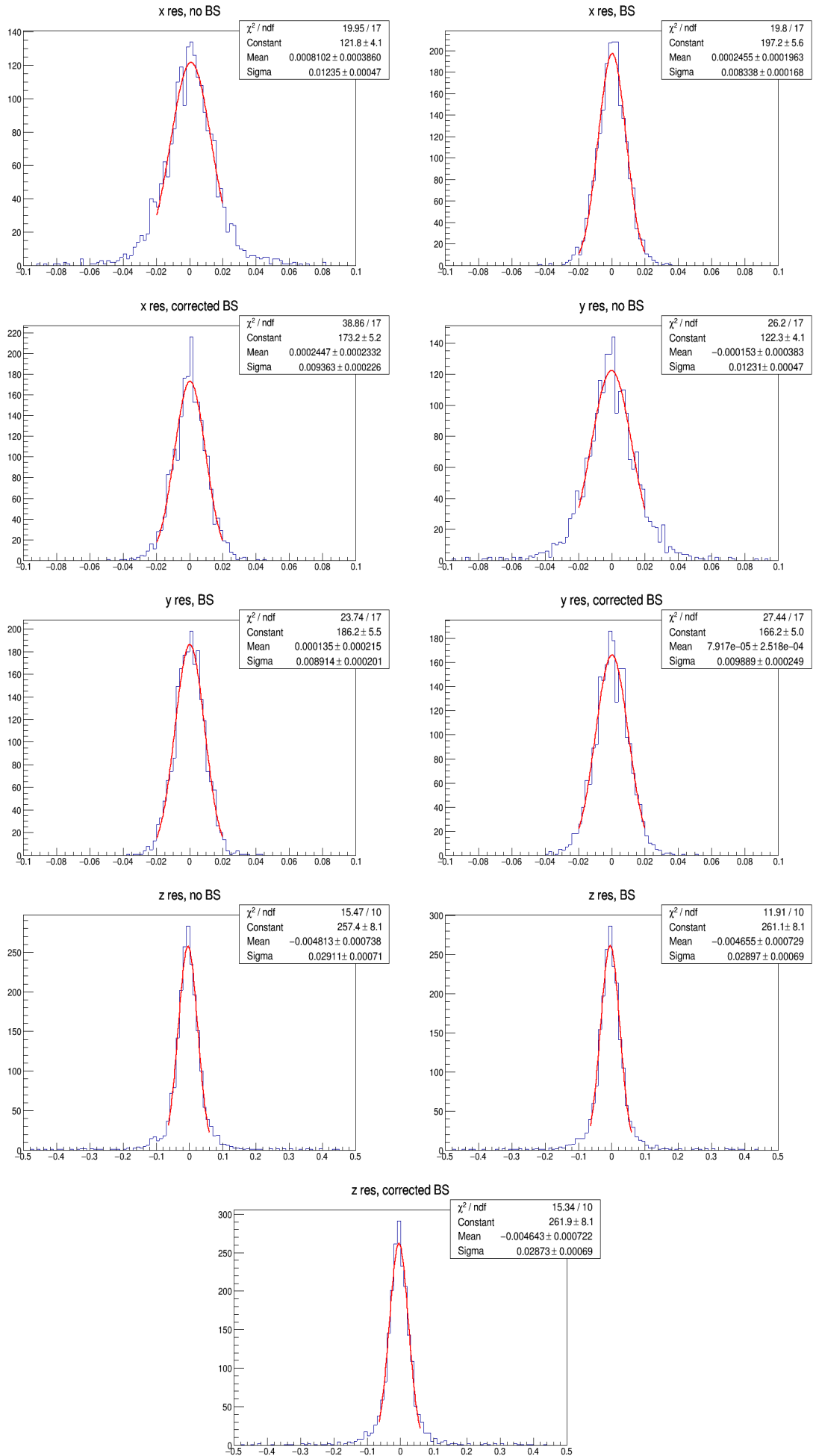


Figure 11.8: Coordinate residual plots of the ggH CP-odd sample.



## 11.6 2D Reconstructed vertex coordinates

2D reconstructed vertex coordinates are shown for the VBF odd sample in figure 11.9, for the ggH even sample in figure 11.10 and for the ggH odd sample in figure 11.11

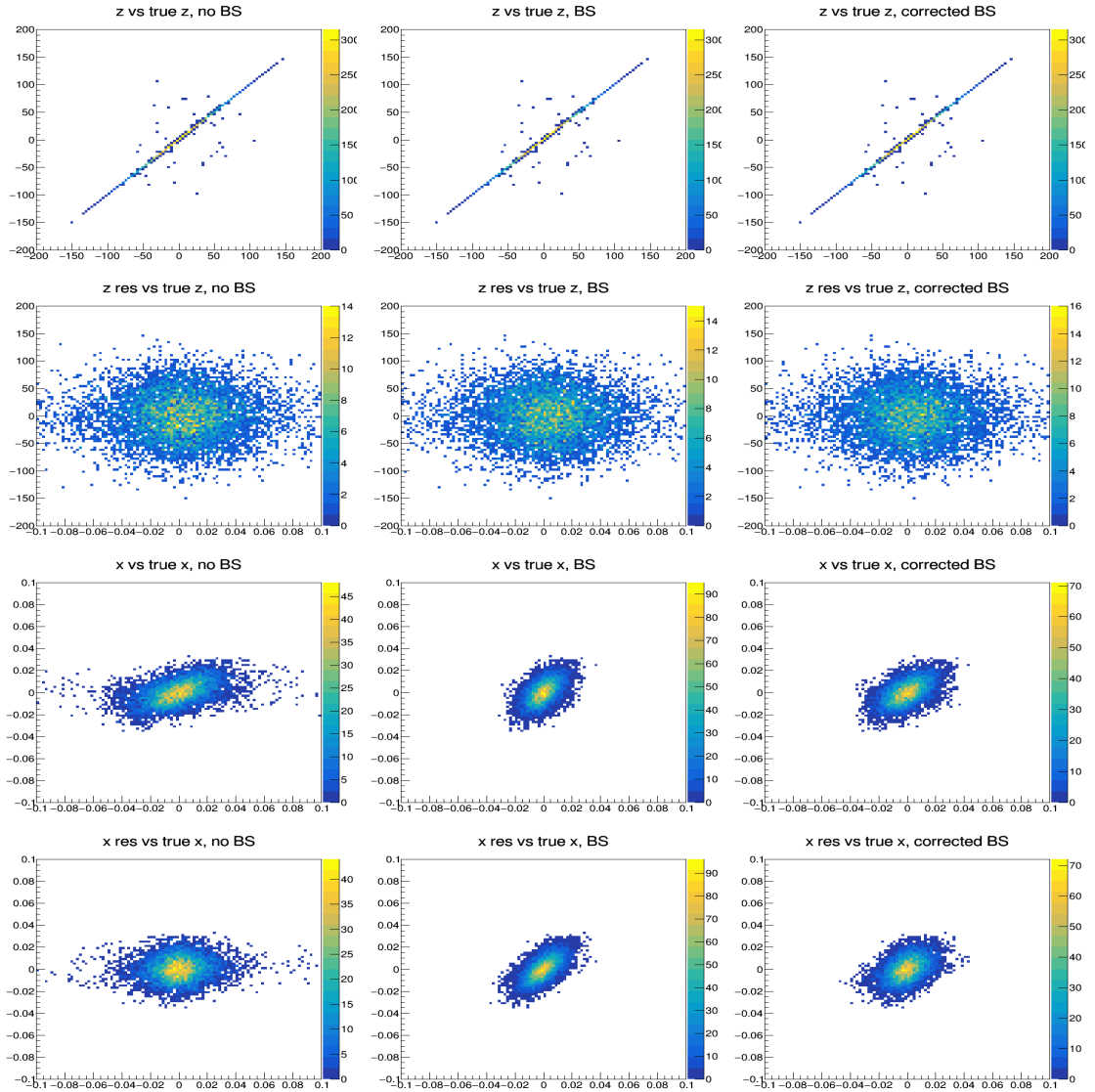


Figure 11.9: First row shows true z position (y-axis) vs z position (x-axis) plots, second row shows row true z position (y-axis) vs z residual (x-axis) plots, third row shows true x position (y-axis) vs x position (x-axis) plots and fourth row shows true x position (y-axis) vs x residual (x-axis) plots. From left to right the plots correspond to the noBS case, default case and soft BS case. All of them are calculated using the VBF odd sample.

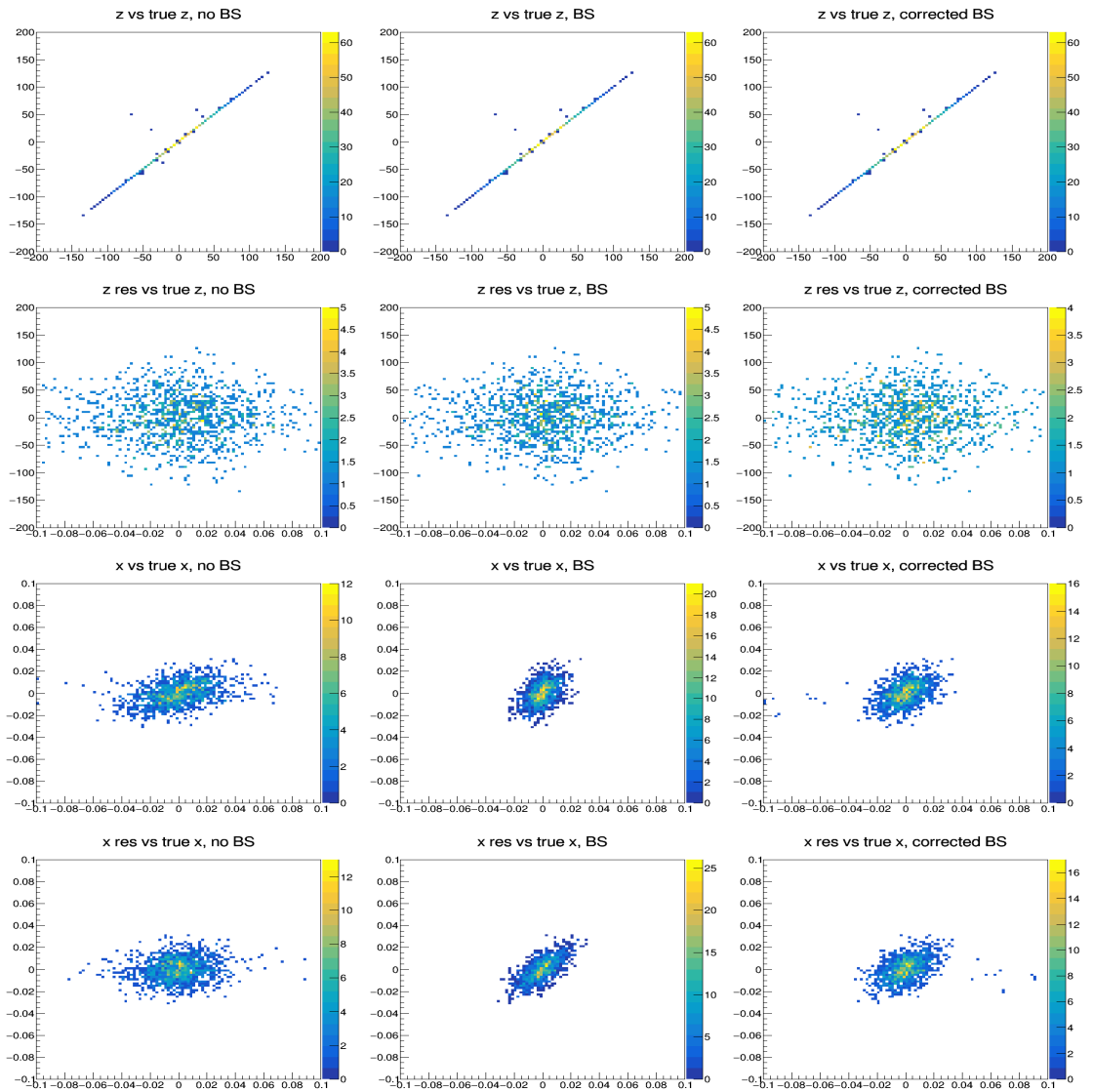


Figure 11.10: First row shows true  $z$  position (y-axis) vs  $z$  position (x-axis) plots, second row shows true  $z$  position (y-axis) vs  $z$  residual (x-axis) plots, third row shows true  $x$  position (y-axis) vs  $x$  position (x-axis) plots and fourth row shows true  $x$  position (y-axis) vs  $x$  residual (x-axis) plots. From left to right the plots correspond to the noBS case, default case and soft BS case. All of them are calculated using the ggH even sample.

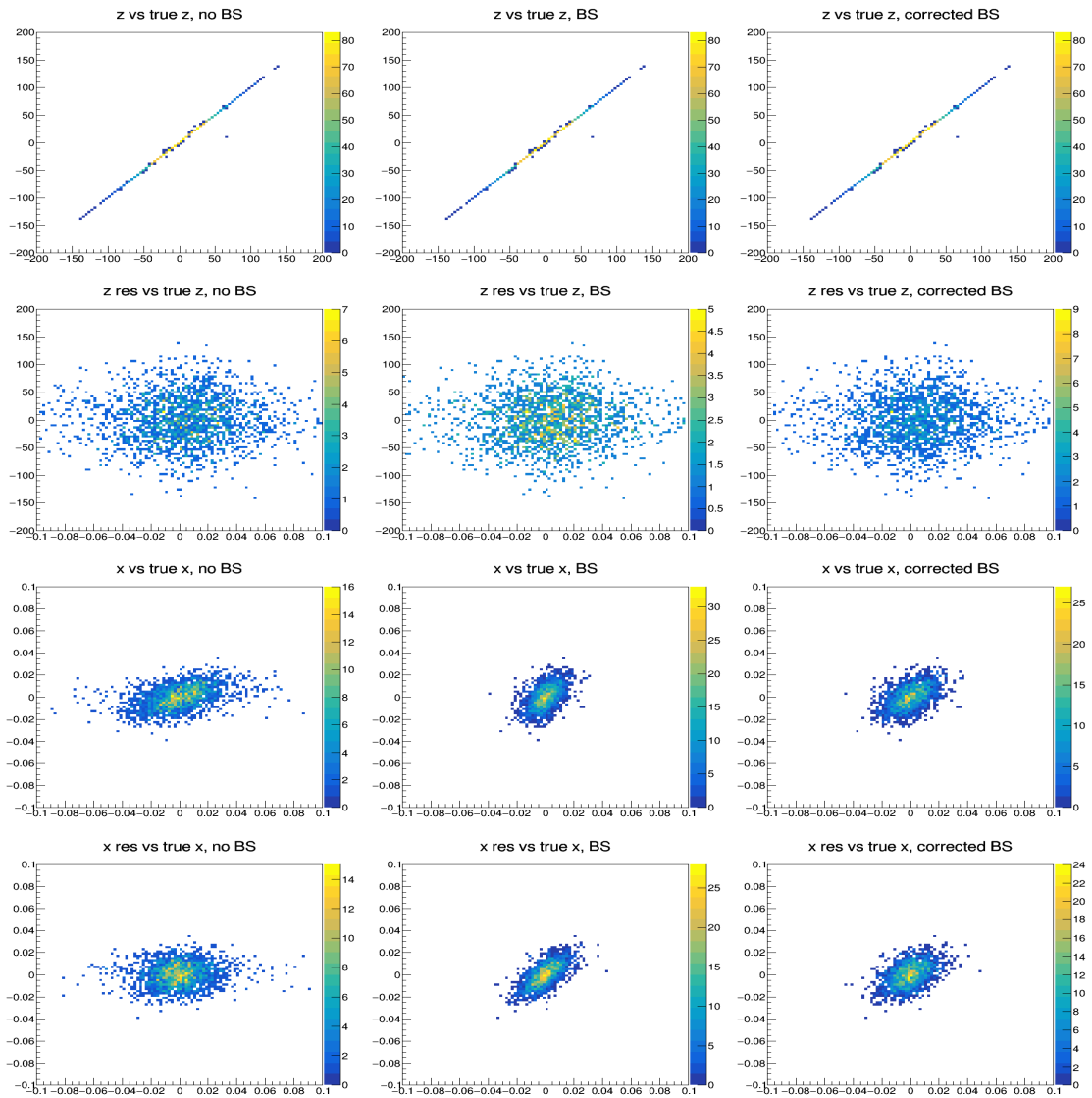


Figure 11.11: First row shows true  $z$  position (y-axis) vs  $z$  position (x-axis) plots, second row shows true  $z$  position (y-axis) vs  $z$  residual (x-axis) plots, third row shows true  $x$  position (y-axis) vs  $x$  position (x-axis) plots and fourth row shows true  $x$  position (y-axis) vs  $x$  residual (x-axis) plots. From left to right the plots correspond to the noBS case, default case and soft BS case. All of them are calculated using the ggH odd sample.

## 11.7 $\varphi^*$ distributions

The  $\varphi^*$  distribution obtained, plotted using 7 bins between  $[0, 2\pi]$ , is shown for the hybrid method. In figure 11.12 is calculated using the VBF odd sample, while figures 11.13 and 11.14 shown the The  $\varphi^*$  distribution using respectively the ggH even and odd samples.

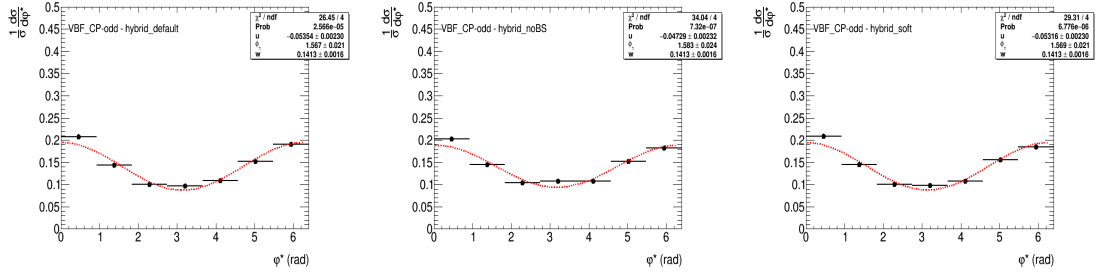


Figure 11.12: From left to right the VBF CP-odd  $\varphi^*$  hybrid method distributions for default, no BS and soft BS.

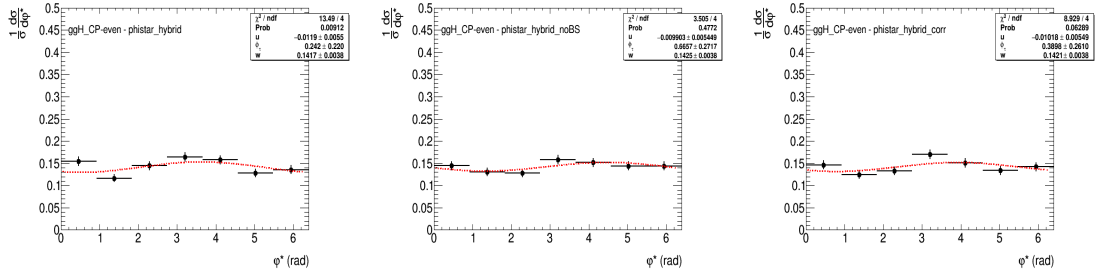


Figure 11.13: From left to right the ggH CP-even  $\varphi^*$  hybrid method distributions for default, no BS and soft BS.

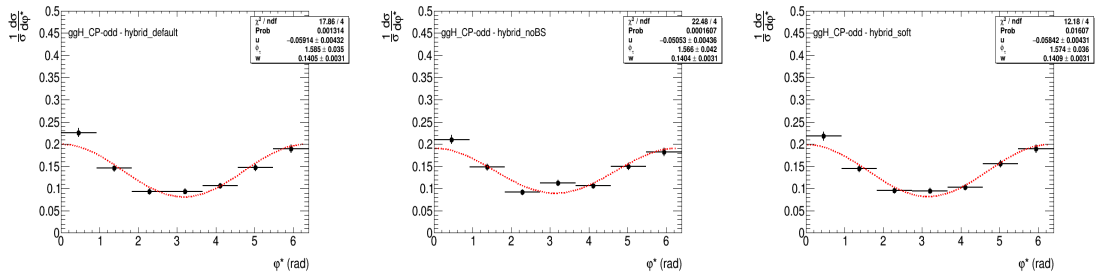


Figure 11.14: From left to right the ggH CP-odd  $\varphi^*$  hybrid methods distributions for default, no BS and soft BS.

## 11.8 $\varphi^*$ residual comparison plots

Within the primary vertex study the  $\varphi^*$  residual comparison plots are produced between the three different scenarios and the MC distribution. The  $\varphi^*$  residual plots for the MC method are shown in figure 11.15 for the VBF odd sample, in figure 11.16 for the ggH even sample and in figure 11.17 for the ggH odd sample, while, the Hybrid method residual plots are shown in figure 11.18 for the VBF odd sample, in figure 11.19 for the ggH even sample and in figure 11.20 for the ggH odd sample.

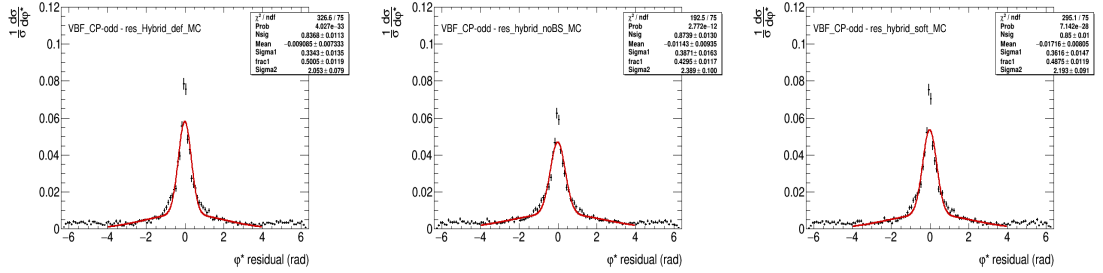


Figure 11.15: From left to right the VBF CP-odd  $\varphi^*$  residual plot when considering (hybrid default - MC), (hybrid noBS - MC) and (hybrid soft - MC) respectively.

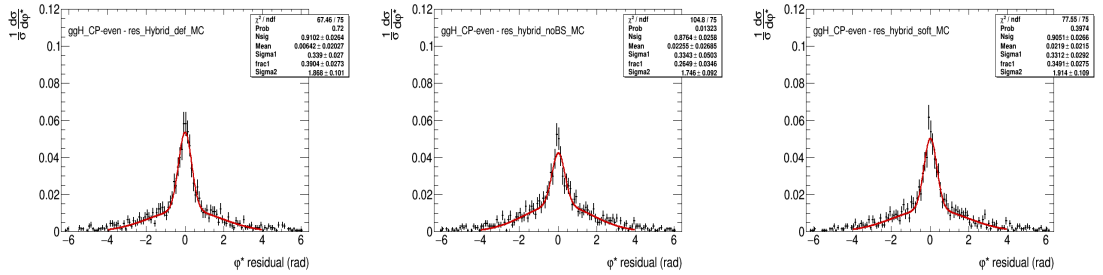


Figure 11.16: From left to right the ggH CP-even  $\varphi^*$  residual plot when considering (hybrid default - MC), (hybrid noBS - MC) and (hybrid soft - MC) respectively.

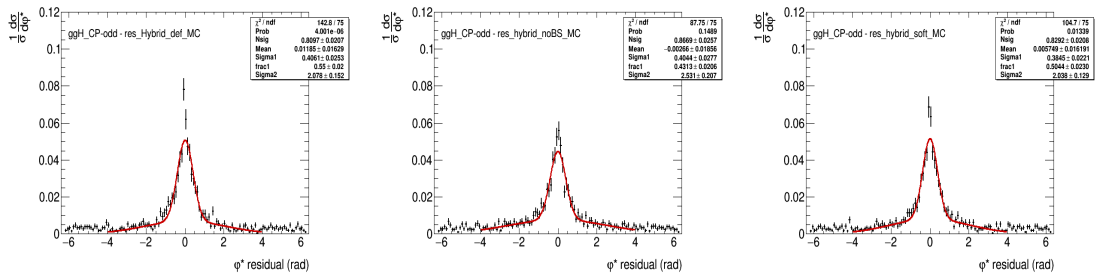


Figure 11.17: From left to right the ggH CP-odd  $\varphi^*$  residual plot when considering (hybrid default - MC), (hybrid noBS - MC) and (hybrid soft - MC) respectively.

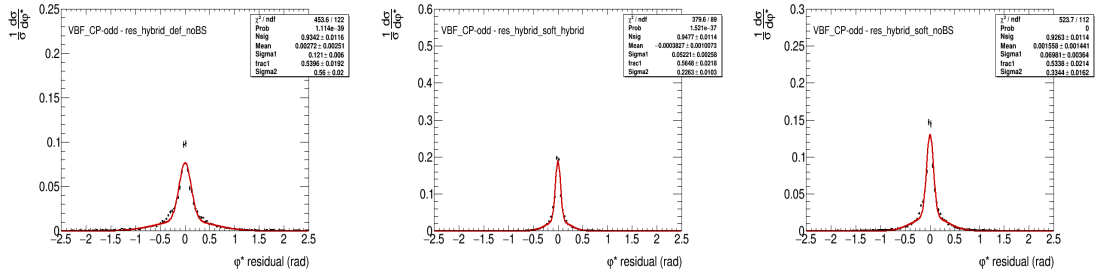


Figure 11.18: From left to right the VBF CP-odd  $\varphi^*$  residual plot when considering (hybrid default - hybrid noBS), (hybrid soft - hybrid default) and (hybrid soft - hybrid noBS) respectively.

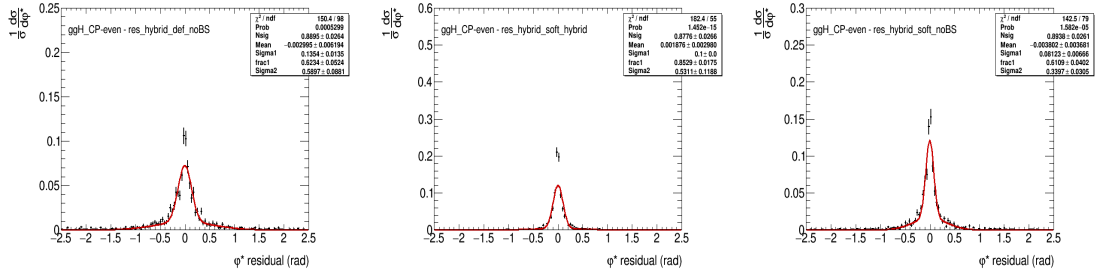


Figure 11.19: From left to right the ggH CP-even  $\varphi^*$  residual plot when considering (hybrid default - hybrid noBS), (hybrid soft - hybrid default) and (hybrid soft - hybrid noBS) respectively.

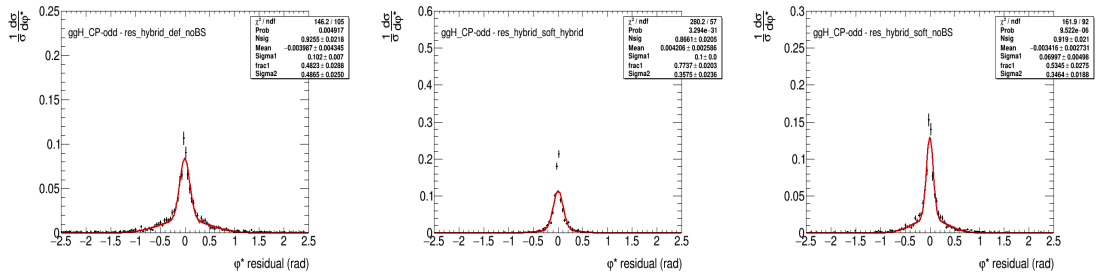


Figure 11.20: From left to right the ggH CP-odd  $\varphi^*$  residual plot when considering (hybrid default - hybrid noBS), (hybrid soft - hybrid default) and (hybrid soft - hybrid noBS) respectively.

---

## 11.9 Smoothing Options studies

The effects of the different smoothing options on selected systematics are illustrated in figures 11.21 11.22 11.23 11.24 for `TAUS_TRUEHADTAU_EFF_RNNID_HIGHPT` and `TAUS_TRUEHADTAU_EFF_JETID` respectively. For each systematic, the boost tight high and the boost loose high regions are shown.

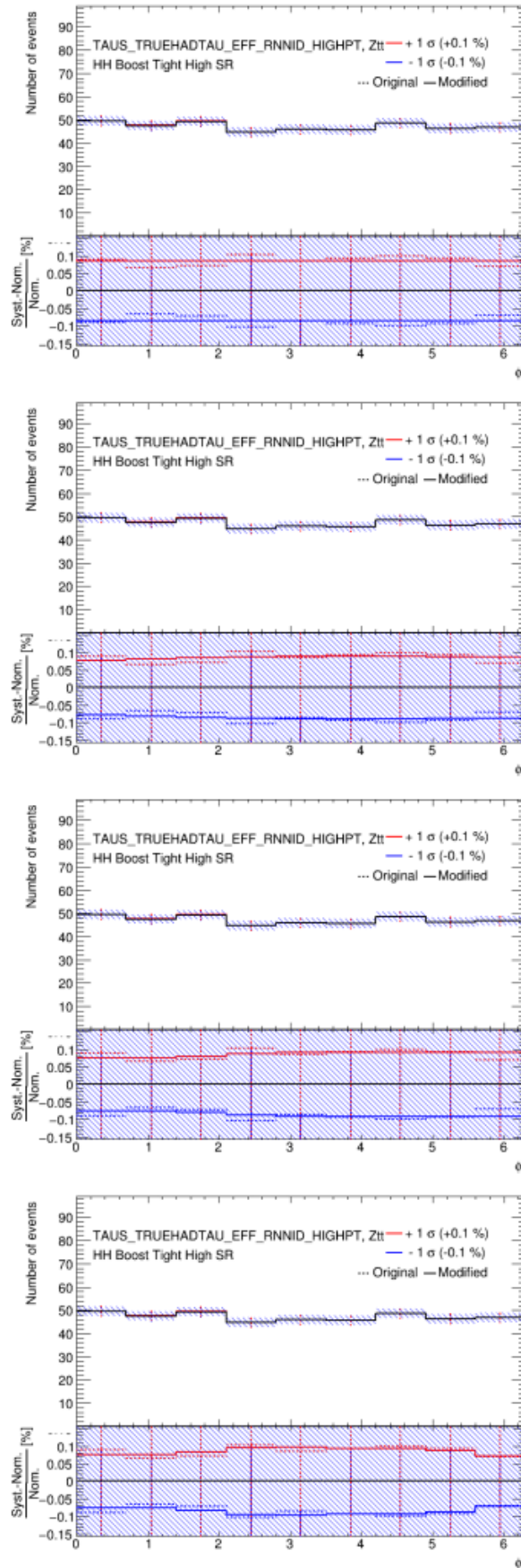


Figure 11.21: The TAUS\_TRUEHADTAU\_EFF\_RNNID\_HIGHPT is shown, for the  $\tau_{\text{had}}\tau_{\text{had}}$  channel, in the boost tight high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic.



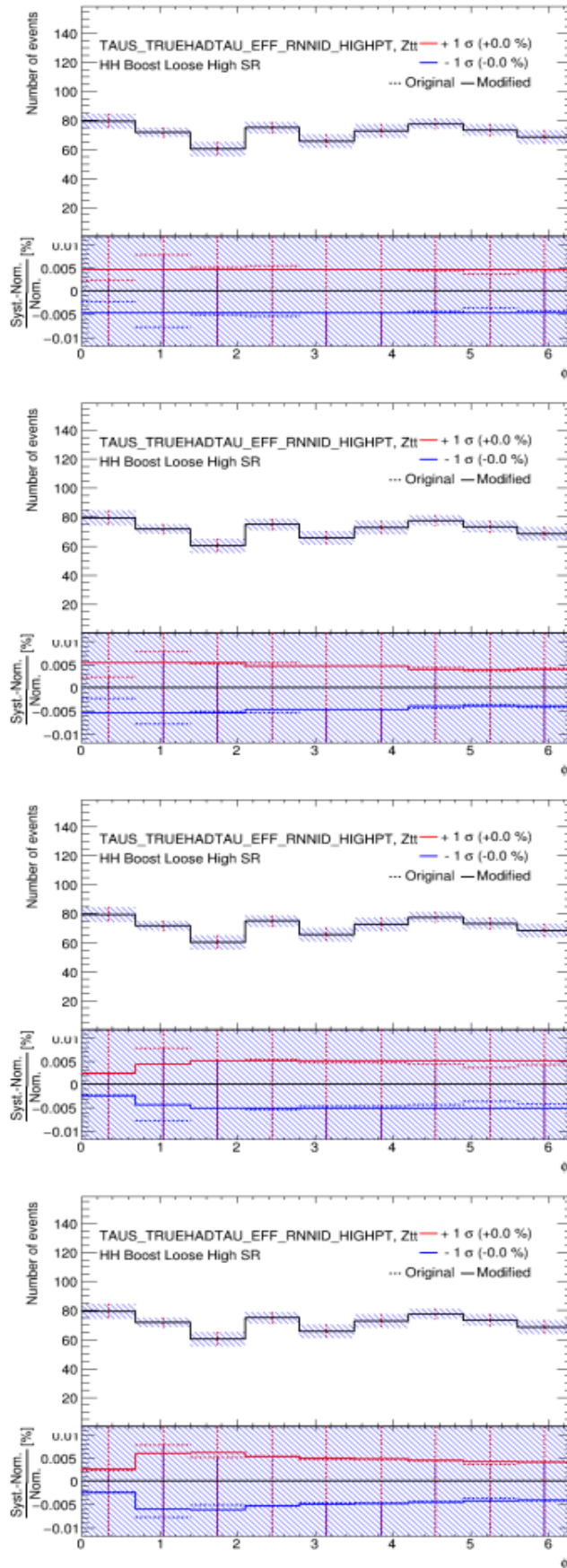


Figure 11.22: The TAUS\_TRUEHADTAU\_EFF\_RNNID\_HIGHPT is shown, for the  $\tau_{\text{had}}\tau_{\text{had}}$  channel, in the boost loose high signal region for the  $Z_{\text{tt}}\text{QCD}$  envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic.

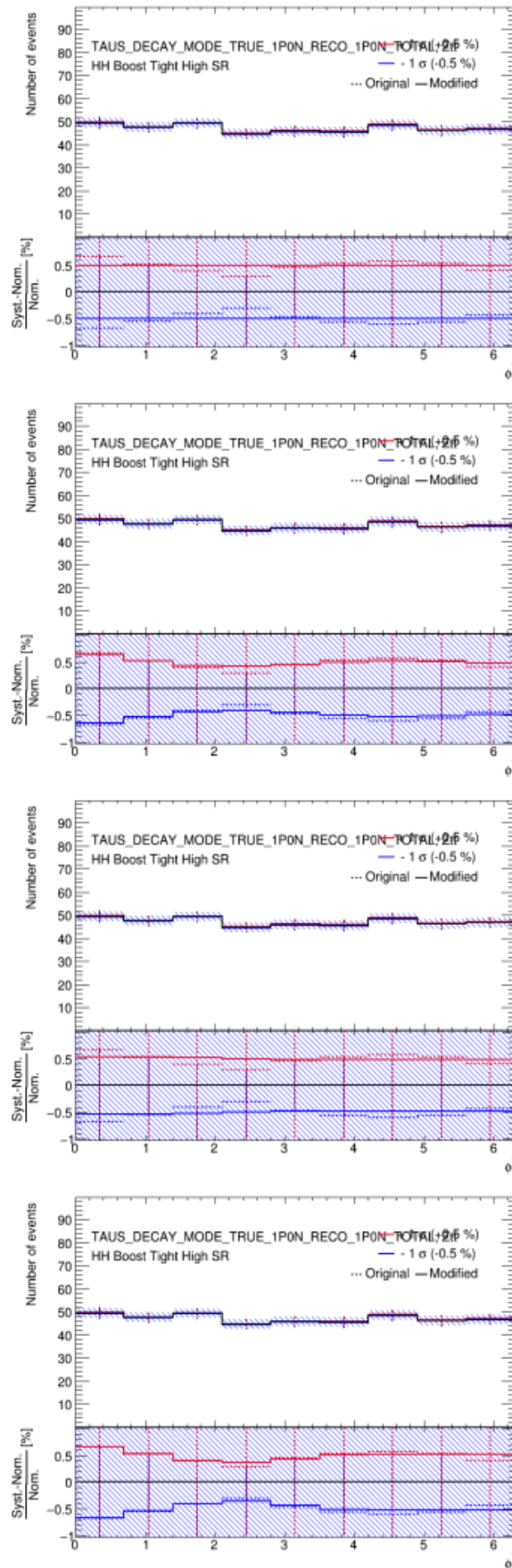


Figure 11.23: The  $\text{TAUS\_TRUEHADTAU\_EFF\_JETID\_DECAY\_MODE\_TRUE\_1P0N\_RECO\_1P0N\_TOTAL}$  is shown, for the  $\tau_{\text{had}}\tau_{\text{had}}$  channel, in the boost tight high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TtBar, MaxVariation, Monotonic and Parabolic.

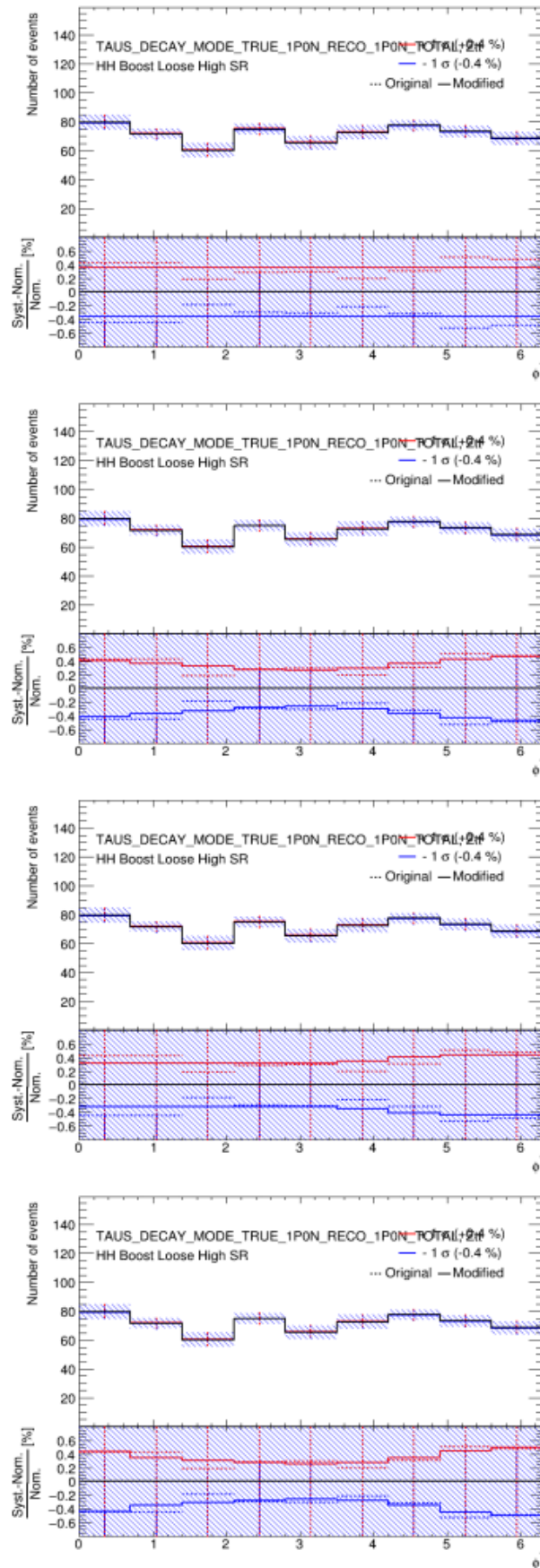


Figure 11.24: The TAUS\_TRUEHADTAU\_EFF\_JETID\_DECAY\_MODE\_TRUE\_1P0N\_RECO\_1P0N\_TOTAL is shown, for the  $\tau_{\text{had}}\tau_{\text{had}}$  channel, in the boost loose high signal region for the ZttQCD envelopes. The Smoothing options applied, from up to down, are TTBar, MaxVariation, Monotonic and Parabolic.

# GLOSSARY

# 12

<b>ALICE</b>	A Large Ion Collider Experiment .....	20
<b>ATLAS</b>	A Toroidal LHC ApparatuS .....	10
<b>BDT</b>	Boosted Decisions Tree .....	39
<b>BEH</b>	Brout-Englert-Higgs mechanism .....	8
<b>BS</b>	Beam Spot .....	74
<b>BSM</b>	Beyond Standard Model .....	23
<b>CERN</b>	European Organization for Nuclear Research .....	5
<b>CKM</b>	Cabibbo–Kobayashi–Maskawa .....	9
<b>CL</b>	Confidence Level .....	10
<b>CMS</b>	Compact Muon Solenoid .....	10
<b>CR</b>	Control Region .....	77
<b>CSC</b>	Cathode Strip Chamber .....	30
<b>eBDT</b>	Electronic Boosted Decision Tree .....	37
<b>ECAL</b>	Electromagnetic Calorimeter .....	25
<b>EM</b>	electromagnetic .....	14
<b>EPJC</b>	The European Physical Journal C .....	133
<b>EW</b>	Electroweak .....	7
<b>EWSSB</b>	Electroweak symmetry breaking .....	14
$E_T^{miss}$	Missing Transverse Energy .....	76
<b>FTK</b>	Fast TracKer .....	32
<b>ggF</b>	Gluon Gluon Fusion .....	62

---

<b>GRL</b>	Good Run List	60
<b>GSC</b>	Global Sequential Calibration	76
<b>HCAL</b>	Hadronic Calorimeter	29
<b>HLT</b>	High Level Trigger	31
<b>IBL</b>	Insertable B-Layer	25
<b>ID</b>	Inner Detector	23
<b>IP</b>	Interaction Point	24
<b>IP</b>	Impact Parameter	24
$I_W$	Weak Isospin	7
<b>JVT</b>	Jet vertex Tagger	38
<b>L1</b>	Level-1	31
<b>LAr</b>	Liquid Argon	29
<b>LHC</b>	Large Hadron Collider	10
<b>LHCb</b>	Large Hadron Collider beauty	9
<b>LHCf</b>	Large Hadron Collider forward	22
<b>LINAC2</b>	Linear Accelerator 2	21
<b>LO</b>	Leading Order	65
<b>MC</b>	Monte Carlo	67
<b>MDT</b>	Monitored Drift Tube	30
<b>ME</b>	Matrix Element	65
<b>MET</b>	Missing Transverse Energy	38
<b>ML</b>	Maximum Likelihood	109
<b>MMC</b>	Missing Mass Calculator	122
<b>MJB</b>	Multi-jet Balance	76
<b>MoEDAL</b>	Monopole & Exotics Detector At the LHC	22
<b>MS</b>	Muon Spectrometer	30
<b>MVA</b>	Multivariate analysis	71
<b>N3LO</b>	Next-to-Next-to-Next-Leading Order	63
<b>NbTi</b>	Alloy of niobium and titanium	25
<b>NLL</b>	Negative Log Likelihood	121

---

<b>NLO</b>	Next-to-Leading Order .....	63
<b>NNLO</b>	Next-to-Next-Leading Order .....	62
<b>NP</b>	Nuisance Parameter .....	79
<b>PDF</b>	Parton Distribution Function .....	63
<b>PMNS</b>	Pontecorvo–Maki–Nakagawa–Sakata .....	9
<b>POI</b>	Parameter of interest .....	108
<b>PV</b>	Primary Vertex .....	26
<b>QCD</b>	Quantum Chromodynamics .....	4
<b>QFT</b>	Quantum Field Theory .....	7
<b>RF</b>	radiofrequency .....	20
<b>RNN</b>	Recurrent Neural Network .....	3
<b>RPC</b>	Resistive-Plate Chamber .....	31
<b>SCT</b>	SemiConductor Tracker .....	23
<b>SM</b>	Standard Model .....	iv
<b>SPS</b>	Super Proton Synchrotron .....	21
<b>SR</b>	Signal Region .....	72
<b>sys</b>	Systematic .....	101
<b>T2K</b>	Tokai to Kamioka .....	9
<b>TGC</b>	Thin-Gap Chamber .....	31
<b>TOTEM</b>	TOTAL cross section, Elastic scattering and diffraction dissociation Measurement at the LHC .....	22
<b>TRT</b>	Transition Radiation Tracker .....	23
<b>VBF</b>	Vector Boson Fusion .....	10
<b>VEV</b>	Vacuum expectation value .....	12
<b>ZCR</b>	Z Control Region .....	73
<b>ZMF</b>	Zero Momentum Frame .....	46

# BIBLIOGRAPHY

- [1] C. Lefèvre, *The CERN accelerator complex*,  
<http://cds.cern.ch/record/1260465>.
  
- [2] The ATLAS Collaboration, “Atlas experiment - photos.”  
<http://atlasexperiment.org/photos/inner-detector-combined.html>.
  
- [3] The ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** no. 08, (2008) S08003.  
<http://stacks.iop.org/1748-0221/3/i=08/a=S08003>.
  
- [4] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, M. Abolins, O. S. AbouZeid, H. Abramowicz, H. Abreu, R. Abreu, Y. Abulaiti, B. S. Acharya, and et al., *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, The European Physical Journal C **76** no. 5, (2016).  
<http://dx.doi.org/10.1140/epjc/s10052-016-4110-0>.
  
- [5] ATLAS and CMS Collaborations Collaboration, C. Mills, *Higgs couplings and properties measurements*, tech. rep., CERN, Geneva, Jan, 2022.  
<https://cds.cern.ch/record/2801634>.
  
- [6] K. Bai, E. V. Bouhova-Thacker, H. C. Cheng, R. A. M. Ferguson, H. Fox, C. Grefe, H. Li, K. Lie, A. C. Manthei, A. De Maria, K. Prokofiev, N. P. Readoff, S. Spinali, P. Wagner, J. Xiang, L. Zhang, P. Tong, T. Vickey, E. Richter-Was, S. Maeland, M. C. Hansen, M. Hubner, P. Bechtle, K. Desch, D. Zanzi, K. Becker, P.-O. Deviveiros, D. B. Ta, S. Hirose, and Q. Buat, *Test of CP Invariance in Higgs Boson Decays to Tau Leptons with the ATLAS Detector*, tech. rep., CERN, Geneva, Nov, 2020. <https://cds.cern.ch/record/2744109>.

- 
- [7] S. Berge and W. Bernreuther, *Determining the CP parity of Higgs bosons at the LHC in the tau to 1-prong decay channels*, Physics Letters B **671** no. 4–5, (2009) 470–476. <http://dx.doi.org/10.1016/j.physletb.2008.12.065>.
- [8] S. Berge, W. Bernreuther, and S. Kirchner, *Prospects of constraining the Higgs boson's CP-nature in the tau decay channel at the LHC*, Physical Review D **92** no. 9, (2015). <http://dx.doi.org/10.1103/PhysRevD.92.096012>.
- [9] A. Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Physical Review D **96** no. 7, (2017). <https://doi.org/10.1103/PhysRevD.96.072002>.
- [10] ATLAS Collaboration Collaboration,  *$E_T^{miss}$  performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, tech. rep., CERN, Geneva, Jun, 2018. <https://cds.cern.ch/record/2625233>. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2018-023>.
- [11] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** no. 4, (1961) 579–588. <https://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [12] A. Salam and J. Ward, *Electromagnetic and weak interactions*, Physics Letters **13** no. 2, (1964) 168–171. <https://www.sciencedirect.com/science/article/pii/0031916364907115>.
- [13] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266. <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [14] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields*, Nuclear Physics B **44** no. 1, (1972) 189–213. <https://www.sciencedirect.com/science/article/pii/0550321372902799>.
- [15] F. J. Dyson, *The Radiation Theories of Tomonaga, Schwinger, and Feynman*, Phys. Rev. **75** (1949) 486–502. <https://link.aps.org/doi/10.1103/PhysRev.75.486>.
- [16] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413–1415. <https://link.aps.org/doi/10.1103/PhysRev.105.1413>.
- [17] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.



- 
- [18] H. et al., *Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment*, Nucl. Phys. B **73** (1974) 1–22.  
<http://cds.cern.ch/record/203096>.
- [19] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587.  
<https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [20] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.  
<https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [21] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156–1163.  
<https://link.aps.org/doi/10.1103/PhysRev.145.1156>.
- [22] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132–133. <https://cds.cern.ch/record/641590>.
- [23] T. D. Lee and C. N. Yang, *Question of Parity Conservation in Weak Interactions*, Phys. Rev. **104** (1956) 254–258.  
<https://link.aps.org/doi/10.1103/PhysRev.104.254>.
- [24] J. W. Cronin, *CP symmetry violation—the search for its origin*, Reviews of Modern Physics **53** no. 3, (1981) 373.
- [25] KTeV Collaboration Collaboration, e. a. Alavi-Harati, *Observation of Direct CP Violation in  $K_{S,L} \rightarrow \pi\pi$  Decays*, Phys. Rev. Lett. **83** (1999) 22–27.  
<https://link.aps.org/doi/10.1103/PhysRevLett.83.22>.
- [26] V. F. et al., *A new measurement of direct CP violation in two pion decays of the neutral kaon*, Physics Letters B **465** no. 1-4, (1999) 335–348.  
<https://doi.org/10.1016%2Fs0370-2693%2899%2901030-8>.
- [27] R. A. et al., *First Observation of CP Violation in the Decays of  $B_s^0$  Mesons*, Physical Review Letters **110** no. 22, (2013).  
<https://doi.org/10.1103%2Fphysrevlett.110.221601>.
- [28] F. Betti, *Observation of CP violation in charm decays at LHCb*, 2019.  
<https://arxiv.org/abs/1905.05428>.
- [29] K. A. et al., *Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations*, Nature **580** no. 7803, (2020) 339–344.  
<https://doi.org/10.1038%2Fs41586-020-2177-0>.

- 
- [30] M. Fukugita and T. Yanagida, *Baryogenesis without grand unification*, Physics Letters B **174** no. 1, (1986) 45–47.  
<https://www.sciencedirect.com/science/article/pii/0370269386911263>.
- [31] I. Affleck and M. Dine, *A new mechanism for baryogenesis*, Nuclear Physics B **249** no. 2, (1985) 361–380.  
<https://www.sciencedirect.com/science/article/pii/0550321385900215>.
- [32] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32–35.
- [33] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, Physics Reports **516** no. 1–2, (2012) 1–102. <http://dx.doi.org/10.1016/j.physrep.2012.02.002>.
- [34] G. R. Farrar and M. E. Shaposhnikov, *Baryon asymmetry of the Universe in the minimal standard model*, Physical Review Letters **70** no. 19, (1993) 2833–2836.  
<http://dx.doi.org/10.1103/PhysRevLett.70.2833>.
- [35] W. Bernreuther, *CP violation and baryogenesis*, 2002.
- [36] A. Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Physics Letters B **726** no. 1-3, (2013) 120–144.  
<https://doi.org/10.1016%2Fj.physletb.2013.08.026>.
- [37] A. Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, The European Physical Journal C **75** no. 10, (2015).  
<https://doi.org/10.1140%2Fepjc%2Fs10052-015-3685-1>.
- [38] C. Collaboration, *Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs*, Physical Review Letters **110** no. 8, (2013).  
<https://doi.org/10.1103%2Fphysrevlett.110.081803>.
- [39] C. Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, Physical Review D **92** no. 1, (2015). <https://doi.org/10.1103%2Fphysrevd.92.012004>.
- [40] *Test of CP invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector*, The European Physical Journal C **76** no. 12, (2016).  
<https://doi.org/10.1140%2Fepjc%2Fs10052-016-4499-5>.

- 
- [41] A. Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson in the  $H \rightarrow \tau \tau$  channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector.*, Physics Letters B **805** (2020) 135426.  
<https://doi.org/10.1016/j.physletb.2020.135426>.
- [42] ATLAS Collaboration, *CP Properties of Higgs Boson Interactions with Top Quarks in the  $t\bar{t}H$  and  $tH$  Processes Using  $H \rightarrow \gamma\gamma$  with the ATLAS Detector*, Phys. Rev. Lett. **125** (2020) 061802.  
<https://link.aps.org/doi/10.1103/PhysRevLett.125.061802>.
- [43] A. Collaboration, *Measurement of the CP properties of Higgs boson interactions with  $\tau$ -leptons with the ATLAS detector*, 2022.
- [44] A. Collaboration, *Probing the CP nature of the top-Higgs Yukawa coupling in  $t\bar{t}H$  and  $tH$  events with  $H \rightarrow b\bar{b}$  decays using the ATLAS detector at the LHC*, 2023.
- [45] J. R. A. et al, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, arXiv:1307.1347.
- [46] M. Spira, A. Djouadi, D. Graudenz, and R. Zerwas, *Higgs boson production at the LHC*, Nuclear Physics B **453** no. 1-2, (1995) 17–82.  
[http://dx.doi.org/10.1016/0550-3213\(95\)00379-7](http://dx.doi.org/10.1016/0550-3213(95)00379-7).
- [47] D. Graudenz, M. Spira, and P. M. Zerwas, *QCD corrections to Higgs boson production at proton proton colliders*, Phys. Rev. Lett. **70** (1993) 1372–1375.
- [48] E. Laenen and L. Magnea, *Threshold resummation for electroweak annihilation from DIS data*, Physics Letters B **632** no. 2-3, (2006) 270–276.  
<http://dx.doi.org/10.1016/j.physletb.2005.10.038>.
- [49] S. Catani, D. d. Florian, M. Grazzini, and P. Nason, *Soft-gluon resummation for Higgs boson production at hadron colliders*, Journal of High Energy Physics **2003** no. 07, (2003) 028–028. <http://dx.doi.org/10.1088/1126-6708/2003/07/028>.
- [50] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos, and B. Mistlberger, *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, 2016.
- [51] R. V. Harlander and W. B. Kilgore, *Next-to-Next-to-Leading Order Higgs Production at Hadron Colliders*, Physical Review Letters **88** no. 20, (2002).  
<http://dx.doi.org/10.1103/PhysRevLett.88.201801>.

- 
- [52] V. Ravindran, J. Smith, and W. van Neerven, *NNLO corrections to the total cross section for Higgs boson production in hadron-hadron collisions*, Nuclear Physics B **665** (2003) 325–366. [http://dx.doi.org/10.1016/S0550-3213\(03\)00457-7](http://dx.doi.org/10.1016/S0550-3213(03)00457-7).
- [53] S. Dawson and R. P. Kauffman, *QCD corrections to Higgs boson production: Nonleading terms in the heavy quark limit*, Physical Review D **49** no. 5, (1994) 2298–2309. <http://dx.doi.org/10.1103/PhysRevD.49.2298>.
- [54] C. Anastasiou, V. Del Duca, E. Furlan, B. Mistlberger, F. Moriello, A. Schweitzer, and C. Specchia, *Mixed QCD-electroweak corrections to Higgs production via gluon fusion in the small mass approximation*, Journal of High Energy Physics **2019** no. 3, (2019). [http://dx.doi.org/10.1007/JHEP03\(2019\)162](http://dx.doi.org/10.1007/JHEP03(2019)162).
- [55] V. Ahrens, T. Becher, M. Neubert, and L. L. Yang, *Updated predictions for Higgs production at the Tevatron and the LHC*, Physics Letters B **698** no. 4, (2011) 271–274. <http://dx.doi.org/10.1016/j.physletb.2010.12.072>.
- [56] M. Ciccolini, A. Denner, and S. Dittmaier, *Strong and Electroweak Corrections to the Production of a Higgs Boson plus 2 Jets via Weak Interactions at the Large Hadron Collider*, Physical Review Letters **99** no. 16, (2007). <http://dx.doi.org/10.1103/PhysRevLett.99.161803>.
- [57] M. Ciccolini, D. A., and D. S., *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC*, Physical Review D **77** no. 1, (2008). <http://dx.doi.org/10.1103/PhysRevD.77.013002>.
- [58] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, *Higgs Boson Production via Vector-Boson Fusion at Next-to-Next-to-Leading Order in QCD*, Physical Review Letters **105** no. 1, (2010). <http://dx.doi.org/10.1103/PhysRevLett.105.011801>.
- [59] O. Brein, R. V. Harlander, and T. J. Zirke,  *$v@hnnlo$ —Higgs Strahlung at hadron colliders*, Computer Physics Communications **184** no. 3, (2013) 998–1003. <http://dx.doi.org/10.1016/j.cpc.2012.11.002>.
- [60] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira, and P. M. Zerwas, *Higgs Radiation Off Top Quarks at the Tevatron and the LHC*, Physical Review Letters **87** no. 20, (2001). <http://dx.doi.org/10.1103/PhysRevLett.87.201805>.
- [61] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira, and P. Zerwas, *NLO QCD corrections to production in hadron collisions*, Nuclear Physics B **653** no. 1-2, (2003) 151–203. [http://dx.doi.org/10.1016/S0550-3213\(03\)00044-0](http://dx.doi.org/10.1016/S0550-3213(03)00044-0).

- 
- [62] L. Reina and S. Dawson, *Next-to-leading order results for  $t$  anti- $t$   $h$  production at the Tevatron*, Phys. Rev. Lett. **87** (2001) 201804, [arXiv:hep-ph/0107101](https://arxiv.org/abs/hep-ph/0107101).
- [63] S. Dawson, L. Orr, L. Reina, and D. Wackeroth, *Next-to-leading order QCD corrections to  $pp \rightarrow t\bar{t}h$  at the CERN Large Hadron Collider*, Phys. Rev. D **67** (2003) .
- [64] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** no. 1, (2012) 1–29.  
<http://dx.doi.org/10.1016/j.physletb.2012.08.020>.
- [65] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, and et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** no. 1, (2012) 30–61.  
<http://dx.doi.org/10.1016/j.physletb.2012.08.021>.
- [66] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole and P. Proudlock, *LHC Design Report*, <https://cds.cern.ch/record/782076>.
- [67] W. Herr and B. Muratori, *Concept of luminosity*,  
<https://cds.cern.ch/record/941318>.
- [68] J.M. Jowett, M. Schaumann, R. Alemany, R. Bruce, M. Giovannozzi, P. Hermes, W. Hofle, M. Lamont, T. Mertens, S. Redaelli, J. Uythoven, J. Wenninger, *The 2015 Heavy-Ion Run of the LHC*,  
<http://accelconf.web.cern.ch/accelconf/ipac2016/papers/tupmw027.pdf>.
- [69] CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation **3** no. 08, (2008) S08004.  
<http://stacks.iop.org/1748-0221/3/i=08/a=S08004>.
- [70] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST (2008).
- [71] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3** no. 08, (2008) S08002.  
<http://stacks.iop.org/1748-0221/3/i=08/a=S08002>.
- [72] TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** no. 08, (2008) S08007.  
<http://stacks.iop.org/1748-0221/3/i=08/a=S08007>.

- 
- [73] LHCf Collaboration, *LHCf experiment: Technical Design Report*. Technical Design Report LHCf. CERN, Geneva, 2006. <http://cds.cern.ch/record/926196>.
- [74] J. Pinfold, *The MoEDAL experiment at the LHC*, EPJ Web Conf. **145** (2017) 12002.
- [75] A. Yamamoto, Y. Makida, R. Ruber, Y. Doi, T. Haruyama, F. Haug, H. ten Kate, M. Kawai, T. Kondo, Y. Kondo, J. Metselaar, S. Mizumaki, G. Olesen, O. Pavlov, S. Ravat, E. Sbrissa, K. Tanaka, T. Taylor, and H. Yamaoka, *The ATLAS central solenoid*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **584** no. 1, (2008) 53 – 74.  
<http://www.sciencedirect.com/science/article/pii/S0168900207020414>.
- [76] The ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. **JC70** , [arXiv:1004.5293](https://arxiv.org/abs/1004.5293).
- [77] The ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*,. <https://cds.cern.ch/record/1291633>.
- [78] G. A. et al., *ATLAS pixel detector electronics and sensors*, Journal of Instrumentation **3** no. 07, (2008) P07007.  
<http://stacks.iop.org/1748-0221/3/i=07/a=P07007>.
- [79] A. Bingül, *The ATLAS TRT and its Performance at LHC*, Journal of Physics: Conference Series **347** no. 1, (2012) 012025.  
<http://iopscience.iop.org/article/10.1088/1742-6596/347/1/012025/meta>.
- [80] The ATLAS Collaboration, *Studies of the performance of the ATLAS detector using cosmic-ray muons*, The European Physical Journal C **71** no. 3, (2011) 1593.  
<https://doi.org/10.1140/epjc/s10052-011-1593-6>.
- [81] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1996.  
<https://cds.cern.ch/record/331062>.
- [82] D. Mahon, *ATLAS LAr calorimeter performance in LHC Run 2*, Journal of Instrumentation **15** no. 06, (2020) C06045.  
<https://dx.doi.org/10.1088/1748-0221/15/06/C06045>.
- [83] The ATLAS Collaboration, *ATLAS muon spectrometer: Technical design report*,. <https://cds.cern.ch/record/331068>.

- 
- [84] The ATLAS Collaboration, *Fast TracKer (FTK) Technical Design Report*,  
<https://cds.cern.ch/record/1552953>.
- [85] M. Aaboud, G. Aad, B. Abbott, D. C. Abbott, O. Abdinov, B. Abeloos, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, H. Abreu, and et al., *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at  $\sqrt{s} = 13$  TeV*, The European Physical Journal C **79** no. 8, (2019).  
<http://dx.doi.org/10.1140/epjc/s10052-019-7140-6>.
- [86] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, tech. rep., CERN, Geneva, 2012. <https://cds.cern.ch/record/1449796>. All figures including auxiliary figures are available at  
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2012-047>.
- [87] A. Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV*, The European Physical Journal C **76** no. 5, (2016).  
<https://doi.org/10.1140/epjc/s10052-016-4120-y>.
- [88] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $K_t$  jet clustering algorithm*, Journal of High Energy Physics **2008** no. 04, (2008) 063–063.  
<https://doi.org/10.1088/1126-6708/2008/04/063>.
- [89] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, tech. rep., CERN, Geneva, Aug, 2019. <https://cds.cern.ch/record/2688062>. All figures including auxiliary figures are available at  
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-033>.
- [90] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at  $\sqrt{s} = 13$  TeV*, tech. rep., CERN, Geneva, 2017. <https://cds.cern.ch/record/2261772>. All figures including auxiliary figures are available at  
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2017-029>.

- 
- [91] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, tech. rep., CERN, Geneva, 2022. [arXiv:2211.16345](https://arxiv.org/abs/2211.16345).  
<https://cds.cern.ch/record/2842028>. 34 pages in total, 19 figures, 2 tables, submitted to EPJC. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/FTAG-2019-07/>.
- [92] B. et al., *Measurement of the Higgs boson coupling properties in the  $H \rightarrow \tau\tau$  decay channel at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, tech. rep., CERN, Geneva, Oct, 2020. <https://cds.cern.ch/record/2741326>.
- [93] CMS Collaboration, *Measurements of  $t\bar{t}H$  Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel*, Phys. Rev. Lett. **125** (2020) 061801.  
<https://link.aps.org/doi/10.1103/PhysRevLett.125.061801>.
- [94] Journal of High Energy Physics (2022).  
<https://doi.org/10.1007%2Fjhep06%282022%29012>.
- [95] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159–177.  
<http://dx.doi.org/10.1016/j.cpc.2015.01.024>.
- [96] Particle Data Group Collaboration, C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. C **40** no. 10, (2016) 100001.
- [97] S. Berge, W. Bernreuther, and S. Kirchner, *Determination of the Higgs CP-mixing angle in the tau decay channels*, 2014.
- [98] A. Collaboration, *Measurements of Higgs boson production cross-sections in the  $H \rightarrow \tau\tau$ - decay channel in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Journal of High Energy Physics **2022** no. 8, (2022).  
<https://doi.org/10.1007%2Fjhep08%282022%29175>.
- [99] A. Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of tau-leptons in proton-proton collisions at  $s=13$ TeV with the ATLAS detector*, Physical Review D **99** no. 7, (2019).  
<http://dx.doi.org/10.1103/PhysRevD.99.072001>.
- [100] T. Przedzinski, E. Richter-Was, and Z. Was, *TauSpinner: a tool for simulating CP effects in  $H \rightarrow \tau\tau$  decays at LHC*, The European Physical Journal C **74** no. 11, (2014). <http://dx.doi.org/10.1140/epjc/s10052-014-3177-8>.



- 
- [101] Z. Czerwinski, T. Przedziński, and Z. Was, *TauSpinner program for studies on spin effect in tau production at the LHC*, The European Physical Journal C **72** no. 4, (2012). <http://dx.doi.org/10.1140/epjc/s10052-012-1988-z>.
- [102] A. Kaczmarska, J. Piatlicki, T. Przedziński, E. Richter-Was, and Z. Was, *Application of TauSpinner for Studies on  $\tau$ -Lepton Polarization and Spin Correlations in Z, W and H Decays at the LHC*, Acta Physica Polonica B **45** no. 10, (2014) 1921. <http://dx.doi.org/10.5506/APhysPolB.45.1921>.
- [103] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, Journal of High Energy Physics **2013** no. 10, (2013). [http://dx.doi.org/10.1007/JHEP10\(2013\)222](http://dx.doi.org/10.1007/JHEP10(2013)222).
- [104] K. Hamilton, P. Nason, and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, 2015.
- [105] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics **2010** no. 6, (2010). [http://dx.doi.org/10.1007/JHEP06\(2010\)043](http://dx.doi.org/10.1007/JHEP06(2010)043).
- [106] P. Nason, *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*, Journal of High Energy Physics **2004** no. 11, (2004) 040–040. <http://dx.doi.org/10.1088/1126-6708/2004/11/040>.
- [107] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **0711** (2007) 070, [arXiv:0709.2092 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0709.2092).
- [108] K. Hamilton, P. Nason, and G. Zanderighi, *MINLO: multi-scale improved NLO*, Journal of High Energy Physics **2012** no. 10, (2012). [http://dx.doi.org/10.1007/JHEP10\(2012\)155](http://dx.doi.org/10.1007/JHEP10(2012)155).
- [109] J. M. Campbell, R. K. Ellis, R. Frederix, P. Nason, C. Oleari, and C. Williams, *NLO Higgs boson production plus one and two jets using the POWHEG BOX, MadGraph4 and MCFM*, Journal of High Energy Physics **2012** no. 7, (2012). [http://dx.doi.org/10.1007/JHEP07\(2012\)092](http://dx.doi.org/10.1007/JHEP07(2012)092).
- [110] K. Hamilton, P. Nason, C. Oleari, and G. Zanderighi, *Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, Journal of High Energy Physics **2013** no. 5, (2013). [http://dx.doi.org/10.1007/JHEP05\(2013\)082](http://dx.doi.org/10.1007/JHEP05(2013)082).

- [111] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. D. Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, and et al., *PDF<sub>4</sub>LHC recommendations for LHC Run II*, Journal of Physics G: Nuclear and Particle Physics **43** no. 2, (2016) 023001. <http://dx.doi.org/10.1088/0954-3899/43/2/023001>.
- [112] G. Aad, B. Abbott, J. Abdallah, S. Abdel Khalek, O. Abdinov, R. Aben, B. Abi, M. Abolins, O. S. AbouZeid, and et al., *Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in  $pp$  collisions at  $s = 7 \sqrt{s} = 7$  TeV with the ATLAS detector*, Journal of High Energy Physics **2014** no. 9, (2014). [http://dx.doi.org/10.1007/JHEP09\(2014\)145](http://dx.doi.org/10.1007/JHEP09(2014)145).
- [113] LHC Higgs Cross Section Working Group Collaboration, de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. CERN, Geneva, Oct, 2016. <https://cds.cern.ch/record/2227475>. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- [114] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger, *Higgs Boson Gluon-Fusion Production in QCD at Three Loops*, Physical Review Letters **114** no. 21, (2015). <http://dx.doi.org/10.1103/PhysRevLett.114.212001>.
- [115] F. Dulat, A. Lazopoulos, and B. Mistlberger, *iHixs 2 — Inclusive Higgs cross sections*, Computer Physics Communications **233** (2018) 243–260. <http://dx.doi.org/10.1016/j.cpc.2018.06.025>.
- [116] R. V. Harlander and K. J. Ozeren, *Finite top mass effects for hadronic Higgs production at next-to-next-to-leading order*, Journal of High Energy Physics **2009** no. 11, (2009) 088–088. <http://dx.doi.org/10.1088/1126-6708/2009/11/088>.
- [117] R. V. Harlander and K. J. Ozeren, *Top mass effects in Higgs production at next-to-next-to-leading order QCD: Virtual corrections*, Physics Letters B **679** no. 5, (2009) 467–472. <http://dx.doi.org/10.1016/j.physletb.2009.08.012>.
- [118] R. V. Harlander, H. Mantler, S. Marzani, and K. J. Ozeren, *Higgs production in gluon fusion at next-to-next-to-leading order QCD for finite top mass*, The European Physical Journal C **66** no. 3-4, (2010) 359–372. <http://dx.doi.org/10.1140/epjc/s10052-010-1258-x>.
- [119] A. Pak, M. Rogal, and M. Steinhauser, *Finite top quark mass effects in NNLO Higgs boson production at LHC*, Journal of High Energy Physics **2010** no. 2, (2010). [http://dx.doi.org/10.1007/JHEP02\(2010\)025](http://dx.doi.org/10.1007/JHEP02(2010)025).

- 
- [120] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, *NLO electroweak corrections to Higgs boson production at hadron colliders*, Physics Letters B **670** no. 1, (2008) 12–17. <http://dx.doi.org/10.1016/j.physletb.2008.10.018>.
- [121] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, *NNLO computational techniques: The cases and*, Nuclear Physics B **811** no. 1-2, (2009) 182–273. <http://dx.doi.org/10.1016/j.nuclphysb.2008.11.024>.
- [122] M. Bonetti, K. Melnikov, and L. Tancredi, *Higher order corrections to mixed QCD-EW contributions to Higgs boson production in gluon fusion*, Physical Review D **97** no. 5, (2018). <http://dx.doi.org/10.1103/PhysRevD.97.056017>.
- [123] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Computer Physics Communications **108** no. 1, (1998) 56–74. [http://dx.doi.org/10.1016/S0010-4655\(97\)00123-9](http://dx.doi.org/10.1016/S0010-4655(97)00123-9).
- [124] M. Spira, *QCD Effects in Higgs Physics*, Fortschritte der Physik **46** no. 3, (1998) 203–284. [http://dx.doi.org/10.1002/\(SICI\)1521-3978\(199804\)46:3<203::AID-PROP203>3.0.CO;2-4](http://dx.doi.org/10.1002/(SICI)1521-3978(199804)46:3<203::AID-PROP203>3.0.CO;2-4).
- [125] A. Djouadi, M. M. Muhlleitner, and M. Spira, *Decays of Supersymmetric Particles: the program SUSY-HIT (SUSpect-SdecaY-Hdecay-InTerface)*, 2006.
- [126] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, *Radiative corrections to the semileptonic and hadronic Higgs-boson decays  $H \rightarrow WW/ZZ \rightarrow 4$  fermions*, Journal of High Energy Physics **2007** no. 02, (2007) 080–080. <http://dx.doi.org/10.1088/1126-6708/2007/02/080>.
- [127] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, *Precise predictions for the Higgs-boson decay  $H \rightarrow WW/ZZ \rightarrow 4$  leptons*, Physical Review D **74** no. 1, (2006). <http://dx.doi.org/10.1103/PhysRevD.74.013004>.
- [128] A. Bredenstein, A. Denner, S. Dittmaier, and M. Weber, *Precision calculations for the Higgs decays  $H \rightarrow ZZ/WW \rightarrow 4$  leptons*, Nuclear Physics B - Proceedings Supplements **160** (2006) 131–135. <http://dx.doi.org/10.1016/j.nuclphysbps.2006.09.104>.
- [129] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, Journal of High Energy Physics **2010** no. 2, (2010). [http://dx.doi.org/10.1007/JHEP02\(2010\)037](http://dx.doi.org/10.1007/JHEP02(2010)037).

- 
- [130] M. L. Ciccolini, S. Dittmaier, and M. Krämer, *Electroweak radiative corrections to associated  $WH$  and  $ZH$  production at hadron colliders*, Physical Review D **68** no. 7, (2003). <http://dx.doi.org/10.1103/PhysRevD.68.073003>.
- [131] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Physics Letters B **579** no. 1-2, (2004) 149–156. <http://dx.doi.org/10.1016/j.physletb.2003.10.112>.
- [132] O. Brein, R. V. Harlander, M. Wiesemann, and T. Zirke, *Top-quark mediated effects in hadronic Higgs-Strahlung*, The European Physical Journal C **72** no. 2, (2012). <http://dx.doi.org/10.1140/epjc/s10052-012-1868-6>.
- [133] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak, and T. J. E. Zirke, *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, Journal of High Energy Physics **2013** no. 2, (2013). [http://dx.doi.org/10.1007/JHEP02\(2013\)078](http://dx.doi.org/10.1007/JHEP02(2013)078).
- [134] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, Computer Physics Communications **195** (2015) 161–171. <http://dx.doi.org/10.1016/j.cpc.2015.04.021>.
- [135] R. V. Harlander, A. Kulesza, V. Theeuwes, and T. Zirke, *Soft gluon resummation for gluon-induced Higgs Strahlung*, Journal of High Energy Physics **2014** no. 11, (2014). [http://dx.doi.org/10.1007/JHEP11\(2014\)082](http://dx.doi.org/10.1007/JHEP11(2014)082).
- [136] S. Frixione, G. Ridolfi, and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, Journal of High Energy Physics **2007** no. 09, (2007) 126–126. <http://dx.doi.org/10.1088/1126-6708/2007/09/126>.
- [137] L. D. W. H. Hartanto, B. Hager, *Higgs boson production in association with top quarks in the POWHEG BOX*, Physical Review D **91** no. 9, (2015). <http://dx.doi.org/10.1103/PhysRevD.91.094003>.
- [138] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, and et al., *Parton distributions for the LHC run II*, Journal of High Energy Physics **2015** no. 4, (2015). [http://dx.doi.org/10.1007/JHEP04\(2015\)040](http://dx.doi.org/10.1007/JHEP04(2015)040).
- [139] *ATLAS Pythia 8 tunes to 7 TeV data*, tech. rep., CERN, Geneva, Nov, 2014. <https://cds.cern.ch/record/1966419>. All figures including auxiliary figures are available at

- <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-021>.
- [140] E. Bothmann, G. Singh Chahal, S. Höche, J. Krause, F. Krauss, S. Kuttimalai, S. Liebschner, D. Napoletano, M. Schönherr, H. Schulz, and et al., *Event generation with Sherpa 2.2*, SciPost Physics **7** no. 3, (2019).  
<http://dx.doi.org/10.21468/SciPostPhys.7.3.034>.
- [141] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, Journal of High Energy Physics **2008** no. 12, (2008) 039–039.  
<http://dx.doi.org/10.1088/1126-6708/2008/12/039>.
- [142] F. Buccioni, J.-N. Lang, J. M. Lindert, P. Maierhöfer, S. Pozzorini, H. Zhang, and M. F. Zoller, *OpenLoops 2*, The European Physical Journal C **79** no. 10, (2019).  
<http://dx.doi.org/10.1140/epjc/s10052-019-7306-2>.
- [143] F. Cascioli, P. Maierhöfer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Physical Review Letters **108** no. 11, (2012).  
<http://dx.doi.org/10.1103/PhysRevLett.108.111601>.
- [144] A. Denner, S. Dittmaier, and L. Hofer, *Collier: A fortran-based complex one-loop library in extended regularizations*, Computer Physics Communications **212** (2017) 220–238. <http://dx.doi.org/10.1016/j.cpc.2016.10.013>.
- [145] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani-Seymour dipole factorisation*, Journal of High Energy Physics **2008** no. 03, (2008) 038–038.  
<http://dx.doi.org/10.1088/1126-6708/2008/03/038>.
- [146] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *A critical appraisal of NLO+PS matching methods*, Journal of High Energy Physics **2012** no. 9, (2012).  
[http://dx.doi.org/10.1007/JHEP09\(2012\)049](http://dx.doi.org/10.1007/JHEP09(2012)049).
- [147] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, Journal of High Energy Physics **2013** no. 4, (2013).  
[http://dx.doi.org/10.1007/JHEP04\(2013\)027](http://dx.doi.org/10.1007/JHEP04(2013)027).
- [148] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, *QCD Matrix Elements + Parton Showers*, Journal of High Energy Physics **2001** no. 11, (2001) 063–063.  
<http://dx.doi.org/10.1088/1126-6708/2001/11/063>.
- [149] S. Hoche, F. Krauss, S. Schumann, and F. Siegert, *QCD matrix elements and truncated showers*, Journal of High Energy Physics **2009** no. 05, (2009) 053–053.  
<http://dx.doi.org/10.1088/1126-6708/2009/05/053>.

- 
- [150] NNPDF Collaboration, NNPDF Collaboration, R.D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph].
- [151] C. Anastasiou, L. Dixon, K. Melnikov, and F. Petriello, *High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, Physical Review D **69** no. 9, (2004).  
<http://dx.doi.org/10.1103/PhysRevD.69.094008>.
- [152] C. O. Alioli, P. Nason and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, tech. rep., 2009. arXiv:0907.4076.
- [153] ATLAS Collaboration Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, tech. rep., CERN, Geneva, Sep, 2016.  
<https://cds.cern.ch/record/2216168>. All figures including auxiliary figures are available at  
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-020>.
- [154] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Computer Physics Communications **178** no. 11, (2008) 852–867.  
<http://dx.doi.org/10.1016/j.cpc.2008.01.036>.
- [155] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and et al., *Parton distributions with LHC data*, Nuclear Physics B **867** no. 2, (2013) 244–289.  
<http://dx.doi.org/10.1016/j.nuclphysb.2012.10.003>.
- [156] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152–155.
- [157] R. Frederix, E. Re, and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, Journal of High Energy Physics **2012** no. 9, (2012). [http://dx.doi.org/10.1007/JHEP09\(2012\)130](http://dx.doi.org/10.1007/JHEP09(2012)130).
- [158] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, The European Physical Journal C **71** no. 2, (2011).  
<http://dx.doi.org/10.1140/epjc/s10052-011-1547-z>.
- [159] S. Frixione, E. Laenen, P. Motylinski, C. White, and B. R. Webber, *Single-top hadroproduction in association with a Wboson*, Journal of High Energy Physics **2008** no. 07, (2008) 029–029. <http://dx.doi.org/10.1088/1126-6708/2008/07/029>.

- 
- [160] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038, [arXiv:0709.1027](https://arxiv.org/abs/0709.1027) [hep-ph].
- [161] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, [arXiv:0808.3674](https://arxiv.org/abs/0808.3674) [hep-ph].
- [162] ATLAS Collaboration Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model*, tech. rep., CERN, Geneva, Aug, 2016.  
<https://cds.cern.ch/record/2206965>. All figures including auxiliary figures are available at  
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-017>.
- [163] e. a. Andreazza, *Measurement of the  $\rightarrow \tau^+\tau^-$  cross-section in 13 TeV Collisions with the ATLAS Detector*, tech. rep., CERN, Geneva, Apr, 2017.  
<https://cds.cern.ch/record/2261605>.
- [164] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A new mass reconstruction technique for resonances decaying to*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** no. 1, (2011) 481–489.  
<http://dx.doi.org/10.1016/j.nima.2011.07.009>.
- [165] A. Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, The European Physical Journal C **75** no. 1, (2015).  
<https://doi.org/10.1140/epjc%2Fs10052-014-3190-y>.
- [166] e. a. Avoni, *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017. 33 p. <http://cds.cern.ch/record/2633501>.
- [167] ATLAS Collaboration Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*, tech. rep., CERN, Geneva, Jun, 2019. <https://cds.cern.ch/record/2677054>. All figures including auxiliary figures are available at  
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-021>.
- [168] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, The European Physical Journal C **75** no. 5, (2015). <https://doi.org/10.1140/epjc%2Fs10052-015-3397-6>.

- 
- [169] *Final Tracking CP Recommendations for Run 2*. 2021. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TrackingCPRecsRun2Final>.
- [170] M. e. a. Aaboud, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, The European Physical Journal C **77** no. 5, (2017). <http://dx.doi.org/10.1140/epjc/s10052-017-4887-5>.
- [171] *TRExFitter v4.07, 2020*,  
<https://gitlab.cern.ch/ATauLeptonAnalysis/TRExFitter>.
- [172] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, tech. rep., Nucl. Inst. and Meth. in Phys. Res. A **389** (1997) 81-86, 1996.
- [173] ROOT Collaboration Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep., New York U., New York, Jan, 2012.  
<https://cds.cern.ch/record/1456844>.
- [174] D. K. Wouter Verkerke, *The RooFit toolkit for data modeling*, tech. rep., 2003.  
[arXiv:physics/0306116](https://arxiv.org/abs/physics/0306116).
- [175] G. S. (for the RooStats Team), *RooStats for Searches*, tech. rep., 2012.  
[arXiv:physics/0306116](https://arxiv.org/abs/physics/0306116).
- [176] A. Read, *Linear interpolation of histograms*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **425** no. 1, (1999) 357–360.  
<https://www.sciencedirect.com/science/article/pii/S0168900298013473>.
- [177] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** no. 2, (2011). <http://dx.doi.org/10.1140/epjc/s10052-011-1554-0>.
- [178] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **236** no. 767, (1937) 333–380.  
<http://www.jstor.org/stable/91337>.
- [179] CMS Collaboration, *Analysis of the CP structure of the Yukawa coupling between the Higgs boson and  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, 2021.  
<https://arxiv.org/abs/2110.04836>.



This thesis was typeset using the L<sup>A</sup>T<sub>E</sub>X typesetting system created by Leslie Lamport.

The body text size is set to 11 pt with *Utopia Regular* with *Fourier* font, part of T<sub>E</sub>X

Live.

The bibliography was typeset using the ATLAS-paper style.