Measurement of the Higgs boson properties using its decay to a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV at the Large Hadron Collider

by

Doyeong Kim

BS, Chonbuk National University, Republic of Korea, 2016

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

The Compact Muon Solenoid (CMS) detector has successfully operated with the Large Hadron Collider (LHC) collisions since 2010, leading to the Higgs boson discovery in 2012 by the CMS and ATLAS collaborations. The analyses of data taken during Run 2 (2016-2018) has already led to many interesting results about the properties of the Higgs boson and for the most part they are compatible in the context of the standard model (SM) so far. These studies are important because the SM appears to be an approximation of a more general theory. Precise studies of the Higgs boson's interactions with the known matter and force particles are essential to reveal the microscopic deviations from the SM prediction. One of such interactions is Yukawa couplings of the Higgs boson and fermions. The $H \to \tau\tau$ decay is the most sensitive fermionic decay mode that allows direct probing of this interaction. This thesis presents a study of the Higgs boson properties by observing its decay to a pair of tau leptons and testing evidence of non-SM interactions in the Higgs boson production. The results exploit the data collected by the CMS detector during LHC Run-2, in proton-proton collisions with a center-of-mass energy of 13 TeV.

Firstly, the Higgs bosons' properties were explored by two complementary approaches by measuring its cross sections. Thanks to the observation of the $H \to \tau\tau$ decay in 2016, we can measure the total cross section with higher precision. Furthermore, in the first analysis, the cross sections of different Higgs productions were scrutinized using fine selections within each Higgs production mode which are mutually exclusive using the Simplified Template Cross Section (STXS) framework, proposed by the LHC Higgs Cross Section Working Group. In the second analysis, the first-ever measurement of the inclusive and differential fiducial cross sections for the Higgs boson production decaying to a pair of tau leptons was presented.

In the third analysis, a search for anomalous couplings in the Higgs production vertex, including Charge-Parity (CP) violating couplings, is also performed targeting Higgs boson production via gluon fusion in association with two jets and vector boson fusion.

None of these analyses has found any hint of new physics beyond the SM, but stringent limits on anomalous Higgs boson couplings to data were set and cross sections of the Higgs bosons in smaller phase spaces were precisely measured.

Measurement of the Higgs boson properties using its decay to a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV at the Large Hadron Collider

by

Doyeong Kim

BS, Chonbuk National University, Republic of Korea, 2016

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Ketino Kaadze

Copyright

Abstract

The Compact Muon Solenoid (CMS) detector has successfully operated with the Large Hadron Collider (LHC) collisions since 2010, leading to the Higgs boson discovery in 2012 by the CMS and ATLAS collaborations. The analyses of data taken during Run 2 (2016-2018) has already led to many interesting results about the properties of the Higgs boson and for the most part they are compatible in the context of the standard model (SM) so far. These studies are important because the SM appears to be an approximation of a more general theory. Precise studies of the Higgs boson's interactions with the known matter and force particles are essential to reveal the microscopic deviations from the SM prediction. One of such interactions is Yukawa couplings of the Higgs boson and fermions. The $H \to \tau\tau$ decay is the most sensitive fermionic decay mode that allows direct probing of this interaction. This thesis presents a study of the Higgs boson properties by observing its decay to a pair of tau leptons and testing evidence of non-SM interactions in the Higgs boson production. The results exploit the data collected by the CMS detector during LHC Run-2, in proton-proton collisions with a center-of-mass energy of 13 TeV.

Firstly, the Higgs bosons' properties were explored by two complementary approaches by measuring its cross sections. Thanks to the observation of the $H \to \tau\tau$ decay in 2016, we can measure the total cross section with higher precision. Furthermore, in the first analysis, the cross sections of different Higgs productions were scrutinized using fine selections within each Higgs production mode which are mutually exclusive using the Simplified Template Cross Section (STXS) framework, proposed by the LHC Higgs Cross Section Working Group. In the second analysis, the first-ever measurement of the inclusive and differential fiducial cross sections for the Higgs boson production decaying to a pair of tau leptons was presented.

In the third analysis, a search for anomalous couplings in the Higgs production vertex, including Charge-Parity (CP) violating couplings, is also performed targeting Higgs boson production via gluon fusion in association with two jets and vector boson fusion.

None of these analyses has found any hint of new physics beyond the SM, but stringent limits on anomalous Higgs boson couplings to data were set and cross sections of the Higgs bosons in smaller phase spaces were precisely measured.

Table of Contents

List of Figures xiii

List of Tables xix

Acknowledgements xxi

1 Theory and motivation 1

1.1 Elementary particles and forces 1

1.1.1 Electromagnetic force 2

1.1.2 Weak force 3

1.1.3 Strong force 3

1.1.4 Gravitational Force 4

1.2 Quantum field theory and gauge symmetry 4

1.3 Motivation 9

2 Phenomenology of the Higgs boson interactions 10

2.1 Production modes 10

2.2 Decay modes 12

2.3 Anomalous couplings 13

2.3.1 Couplings to vector bosons 14

3 The Experiment 18

3.1 The Large Hadron Collider 18

3.2 The CMS Experiment 20

3.2.1 Coordinate system 22

3.2.2 Magnet . . . 22
3.2.3 Parts of the CMS Detector . . . 23
3.2.4 Trigger . . . 26
3.2.5 Particle identification and reconstruction . . . 26

4 Statistics . . . 28
4.1 Likelihood . . . 28
4.1.1 Basics . . . 28
4.1.2 Introducing systematic uncertainties . . . 29
4.2 Maximum likelihood fit . . . 33
4.3 Exclusion limits . . . 33
4.4 Goodness-of-fit test . . . 34
4.5 Chapter summary . . . 36

5 Object reconstruction and analysis tools . . . 37
5.1 Object reconstruction . . . 37
5.1.1 Tracks and vertices . . . 38
5.1.2 Particle flow . . . 39
5.1.3 Muon . . . 39
5.1.4 Electron . . . 41
5.1.5 Jets . . . 43
5.1.6 b-jets . . . 45
5.1.7 Missing transverse energy . . . 46
5.1.8 Hadronic taus . . . 47
5.2 Datasets and triggers . . . 50
5.3 Event simulation . . . 52
5.3.1 Monte Carlo samples . . . 53
5.3.2 Embedded samples . . . 54

5.4 Simulation corrections . . . 56
5.4.1 E_T^{miss} recoil corrections . . . 56
5.4.2 Pile-up reweighting . . . 56
5.4.3 Top p_T reweighting . . . 57
5.4.4 Z/$\gamma^* p_T$-mass reweighting . . . 58
5.4.5 Electron and muon efficiency corrections . . . 58
5.4.6 Hadronic tau efficiency corrections . . . 58
5.4.7 e$\rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ fake rate corrections . . . 58
5.4.8 Tau energy scale correction . . . 59
5.4.9 e$\rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ energy scale corrections . . . 59
5.5 Background methods . . . 59
5.5.1 Modelling backgrounds using MC . . . 59
5.5.2 Modelling Z/$\gamma^* \rightarrow \tau\tau$ using embedding . . . 60
5.5.3 The fake-factor method . . . 60
5.6 SVFit for Di-tau mass . . . 64

6 Systematic uncertainties . . . 67
6.1 Uncertainties in the object reconstruction . . . 67
6.1.1 τ_h identification . . . 67
6.1.2 τ_h energy scale . . . 68
6.1.3 $e \rightarrow \tau_h$ misidentification . . . 68
6.1.4 $e \rightarrow \tau_h$ energy scale . . . 68
6.1.5 $\mu \rightarrow \tau_h$ misidentification . . . 69
6.1.6 $\mu \rightarrow \tau_h$ energy scale . . . 69
6.1.7 Muon identification . . . 69
6.1.8 Muon energy scale . . . 69
6.1.9 Electron identification . . . 70
6.1.10 Electron energy scale . . . 70

6.1.11 Trigger . . . 70
6.1.12 Jet energy scale . . . 70
6.1.13 Jet energy resolution . . . 70
6.1.14 p_T^{miss} reconstruction . . . 71
6.2 Uncertainties in the the background estimation methods . . . 71
6.2.1 $jet \to \tau_h$ fake background . . . 71
6.3 Uncertainties in the background cross sections . . . 72
6.4 Theoretical uncertainties for the signal . . . 72
6.4.1 Uncertainties related to the embedded samples . . . 72
6.4.2 Other uncertainties . . . 73

7 Measurement of the Higgs boson cross section in STXS framework . . . 74
7.1 Introduction to the STXS framework . . . 74
7.2 Event selection . . . 76
7.3 Event categorization . . . 78
7.4 Systematic uncertainties . . . 78
7.5 Signal extraction and statistical interpretation . . . 79
7.6 Results . . . 80
7.6.1 Inclusive and Stage 0 best-fit signal strengths . . . 80
7.6.2 Stage 1 best-fit signal strengths . . . 81
7.6.3 Cross section measurements . . . 83
7.7 Chapter summary . . . 85

8 Measurement of the inclusive and differential Higgs boson production cross sections 93
8.1 Event selection . . . 93
8.2 Event categorization . . . 94
8.3 Higgs p_T resolution . . . 94
8.4 Systematic uncertainties . . . 95

8.5 Signal extraction and statistical interpretation . . . 95
8.5.1 Fiducial region . . . 95
8.5.2 Signal splitting at generated level . . . 97
8.5.3 Fit strategy - signal extraction and unfolding procedure . . . 98
8.6 Results . . . 105
8.7 Chapter summary . . . 109

9 Search for anomalous couplings of Higgs bosons . . . 110
9.1 Event selection . . . 110
9.2 Event categorization . . . 110
9.3 Multivariate observables . . . 111
9.3.1 Matrix Element Likelihood Approach (MELA) . . . 111
9.3.2 Neural network . . . 115
9.4 Systematic uncertainties . . . 120
9.5 Signal extraction and statistical interpretation . . . 120
9.6 Results . . . 125
9.6.1 Results of the HVV analysis . . . 125
9.6.2 Results of the ggH analysis . . . 126
9.7 Chapter summary . . . 128

10 Summary . . . 131

Bibliography . . . 132

List of Figures

1.1 The interactions of particles in the SM.[1] 3

2.1 Feynman diagram for the Higgs boson production modes 11

2.2 Higgs boson branching ratios 13

2.3 Feynman diagram for H$\rightarrow \bar{f}f$ decay 14

3.1 A picture of the LHC and interaction points 19

3.2 Delivered luminosity versus time for 2010-2012 and 2015-2018 20

3.3 A cutaway view of the CMS detector. 21

3.4 Magnetic field and field lines predicted on the superconducting solenoid magnet in the CMS detector 23

3.5 A cross sectional view of the CMS detector.[2] 27

4.1 Log-normal (left) and Gamma (right) distributions for various parameter values. 31

4.2 Example of goodness-of-fit test. 35

5.1 The muon p_T resolution 41

5.2 Efficiencies for simulated τ_h decays 51

5.3 Illustration of the four main steps of the τ-embedding technique 55

5.4 p_T^{miss} distribution before (left) and after (right) recoil correction in 2018 $Z \rightarrow \mu\mu$ events. 57

5.5 The invariant mass of Higgs boson candidate, $m_{\tau\tau}$, in VBF topology. All events selected in the $\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$ and $\tau_e\tau_\mu$ channels are included. Only statistical uncertainties are shown. 61

5.6 Fake factors determined in the QCD multijet sideband region with 0jet in the $\tau_h\tau_h$ final state in 2016 (left), 2017 (middle) and 2018 (right). They are fitted with the sum of a Landau function and polynomial as a function of the leading τ_h p_T. 63

5.7 Correction for the inversion of the sign requirement in the $\tau_h\tau_h$ final state of the QCD multijet sideband region in 2016 (left), 2017 (middle) and 2018 (right). 63

5.8 Parametrization of *tau* lepton decay in *tau* rest frame. 64

7.1 Stage 1.2 bins for ggH 75

7.2 Stage 1.2 bins for qqH 76

7.3 Observed and predicted distributions in the $\tau_h\tau_h$ channel 86

7.4 Observed and predicted distributions in the $\tau_h\tau_h$ channel 87

7.5 Observed and expected $m_\tau\tau$ distribution obtained by reweighting 88

7.6 Observed results of the fit to the signal strength modifiers 88

7.7 Correlation matrix between merged stage 1.2 STXS parameters.[3] 89

7.8 Observed stage 1.2 parameters obtained from the fit of all data taking years, categories, and final states. The left plot corresponds to the process-based merging scheme, and ther right plot to the topology-based merging scheme. The contributions from the statistical uncertainty to the total uncertainty in each parameter are shown as colored squares. The ggH processes are indicated in blue while qqH are indicated in orange. The green squares indicate parameters for the ggH and qqH combined productions in the topology-based merging scheme. All parameters are measured simultaneously.[3] 90

7.9 Production of the cross section and branching ratio measured for the inclusive, stage 0 and stage 1.2 parameters. The top plot corresponds to the process-based merging scheme, and the bottom plot to the topology-based merging scheme. The ggH processes are indicated in blue while qqH are indicated in orange. The green squares indicate parameters for the ggH and qqH combined productions in the topology-based merging scheme.[3] 91

7.10 The negative log-likelihood scans as a function of κ_V and κ_f (left) and the coupling modifiers for the ggF and VBF productions (right), for $m_H = 125.09$ GeV.[3] 92

8.1 2D distribution ratio of expected total Higgs signal yield over observed yield 95

8.2 The resolution of Higgs p_T in high p_T (> 200 GeV) region 96

8.3 Correction factor applied to the reconstructed Higgs p_T in $\tau_h\tau_h$ channel. . . . 96

8.4 Response matrices for the reconstructed vs. generated p_T^H (top), N_{jets} (middle), and p_T^{jet1} (bottom) in the $\tau_h\tau_h$ channel for 2018. The horizontal axis represents the generator-level bins and the vertical axis represents the reconstruction-level bins. Each column is normalized so that event yield in the column adds up to unity. 99

8.5 Correlation between signal strength modifiers in fiducial p_T^H bins obtained from fits to data without regularization. 102

8.6 Correlation between signal strength modifiers in fiducial p_T^H bins obtained from fits to data with regularization. 102

8.7 Correlation between signal strength modifiers in fiducial N_{jets} bins obtained from fits to data without regularization. 103

8.8 Correlation between signal strength modifiers in fiducial N_{jets} bins obtained from fits to data with regularization. 103

8.9 Correlation between signal strength modifiers in fiducial p_T^{jet1} bins obtained from fits to data without regularization. 104

8.10 Correlation between signal strength modifiers in fiducial p_T^{jet1} bins obtained from fits to data with regularization. 104

8.11 Optimization of the regularization strength δ for p_T^H (left), N_{jets} (middle), and p_T^{jet1} (right) fits, with respect to the mean of the global correlation coefficients. 105

8.12 Expected and observed $m_{\tau\tau}$ distributions in the different reconstructed ranges of the differential observable. Every $m_{\tau\tau}$ distribution of each category, year, and final state are reweighted by a factor proportional to the ratio between the signal and background yields. The distributions are normalized to conserve signal yield. The signal and background distributions result from a multidimensional maximum likelihood regularized fit. The green line in the ratio plot shows the expected SM signal, while the red line corresponds to the best-fit signal prediction. The uncertainty band includes systematic and statistical uncertainties.[4] 106

8.13 Observed signal strength modifiers in p_T^H (left), N_{jets} (middle), and p_T^{jet1} (right) bins. 107

8.14 Observed and expected differential fiducial cross section in bins of p_T^H (upper left), N_{jets} (upper right), and p_T^{jet1} (lower). Both regularized (full markers) and unregularized (hollow markers) are shown. The first bin in the p_T^{jet1} plot contains all events without a jet. The shaded bands in the theoretical predictions include uncertainties from the following sources: PDF, renormalization and factorization scale, underlying event and parton showering, and branching fraction of the Higgs boson to τ leptons. Overflow is included in the last bins of each distribution.[4] 108

9.1 Illustrations of Higgs production in VBF ($qq' \to qq'H$) (left) and VH ($q\bar{q'} \to V^* \to VH \to qq'H$) (right) in the rest frame of the Higgs. Higgs decays to a τ lepton pair is shown without illustrating the further decay chain. Brown shows the incoming partons and fermions in the V decay. The intermediate or final-state particles are shown in red and green. Five angles characterizing kinematic distributions are shown in blue and are defined in the respective rest frames[5]. The illustration for Higgs production via ggH in association with two jets is identical to the VBF diagram, except with $V = g$. 112

9.2 The distribution of five angles characterizing kinematics of the Higgs + 2 jet system in the $\tau_h\tau_h$ channel, 2018. Only statistical uncertainties are included in the uncertainty band. 113

9.3 The distribution of two momenta characterizing kinematics of the Higgs + 2 jet system in the $\tau_h\tau_h$ channel, 2018. Only statistical uncertainties are included in the uncertainty band. 114

9.4 The distribution of the p_T of the Higgs boson (left) and di-tau invariant mass (right) in the $\tau_h\tau_h$ channel, 2018. Only statistical uncertainties are included in the uncertainty band. 117

9.5 Plots shows correlation between all neural network training input variables in the $\tau_h\tau_h$ channel, 2018. 117

9.6 Training and validation loss for 2018, the $\tau_h\tau_h$ channel. The training and validation loss converge to a similar value. 118

9.7 $\mathcal{D}_{NN}$ distribution in the $\tau_h\tau_h$ channel, 2018. Only statistical uncertainties are included in the uncertainty band. 118

9.8 Distribution comparison between different generator and hypotheses (POWHEG SM (black), JHUGEN SM (red), and JHUGEN PS (green)) of VBF process in the $\tau_h\tau_h$ channel in VBF category for 2018. 121

9.9 Examples of data and signal and background predictions for MELA and neural network discriminants in the $\tau_h\tau_h$ ($\tau_h\tau_h$) and $\tau_\mu\tau_h$ ($\mu\tau_h$) channels. Statistical uncertainties and systematic uncertainties that affect the normalization of the background distribution are included in the uncertainty band. The expectation in the ratio panel is the sum of the estimated the SM Higgs signal and backgrounds. For the $\mathcal{D}_{0-}^{\text{ggH}}$ discriminant the distribution expected for a pseudoscalar Higgs hypothesis (labeled "PS" in the legend) is overlaid to be compared to the SM signal. Similarly, for the $\mathcal{D}_{CP}^{\text{ggH}}$ discriminant the distribution for a CP-violating scenario with the maximum-mixing between CP-even and CP-odd couplings (labeled "MM" in the legend) is shown.[6] 123

9.10 The observed and predicted 2D distribution of ($\mathcal{D}_{0-}$, $\mathcal{D}_{NN}$, $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$) before the fit to data in the most sensitive VBF category in the $\tau_h\tau_h$ channel. The total Higgs signal is shown stacked on top of the background in the solid red histogram. The VBF+VH signal for the CP-even (CP-odd) scenario is shown by the red (blue) line. The uncertainty band include only the statistical uncertainties.[6] 127

9.11 Observed (solid) and expected (dashed) likelihood scans of f_{a3} (upper left), f_{a2} (upper right), $f_{\Lambda1}$ (lower left), and $f_{\Lambda1}^{Z\gamma}$ (lower right) in Approach 1 ($a_i^{WW} = a_i^{ZZ}$).[6] 128

9.12 Observed (solid) and expected (dashed) likelihood scans of f_{a3} in Approach 2.[6] 129

9.13 The observed and predicted 2D distribution of ($\mathcal{D}_{0-}^{ggH}$, $\mathcal{D}_{NN}$) before the fit to data in the most sensitive VBF category region with $0.3 < \mathcal{D}_{\text{2jet}}^{\text{VBF}} < 0.7$ in the $\tau_h\tau_h$ channel. The total Higgs signal is shown stacked on top of the background in the solid red histogram. The ggH signal for the CP-even (CP-odd) scenario is shown by the red (blue) line. The uncertainty band include only the statistical uncertainties.[6] 129

9.14 Observed (solid) and expected (dashed) likelihood scans of f_{a3}^{ggH} obtained.[6] . 130

List of Tables

1.1 The twelve fundamental fermions grouped into quarks and leptons. 2

1.2 The four known forces of particle physics. The relative strengths are indicative values for two fundamental particles at a distance of $1fm = 10^{-15}m$ which is approximately the radius of a proton. 4

1.3 Quantum numbers of fermions singlets and doublets. 7

1.4 Quantum numbers of Higgs doublets. 7

2.1 Summary of the Higgs boson production modes 12

2.2 Cross sections for the anomalous contributions (σ_i) used to define the fractional cross sections[6]. The σ_i values are defined as the cross section computed with $a_i = 1$ and all other couplings set to zero. All cross sections are given relative to the SM value (σ_1). In the case of the κ_1 and $\kappa_2^{Z\gamma}$ couplings, the numerical values $\Lambda_1 = \Lambda_1^{Z\gamma} = 100$ GeV are considered so as to keep all coefficients of similar order of magnitude. 16

5.1 Summary of the most common hadronic tau decay modes 48

6.1 Uncertainty in the muon energy scale 69

7.1 Kinematic selection requirements for the four di-tau decay channels. 77

7.2 Event categorization in STXS measurements 79

7.3 Observed inclusive and stage 0 signal strengths per year with all final states combined, per final state with all years combined, and with all final states and years combined.[3] 81

7.4 Product of the production cross section σ and branching fraction $\mathcal{B}(H \to \tau\tau)$ measured for the inclusive and stage 0 processes.[3] 83

7.5 Product of the production cross section σ and branching fraction $\mathcal{B}(H \to \tau\tau)$ measured for the stage 1.2 processes in the process-based merging scheme.[3] . 84

7.6 Product of the production cross section σ and branching fraction $\mathcal{B}(H \to \tau\tau)$ measured for the stage 1.2 processes in the topology-based merging scheme.[3] 84

8.1 Fiducial region definition 97

8.2 Observed and expected fiducial cross sections in N_{jets}, p_T^H, and p_T^{jet1} bins. The signal strengths, which do not include yield uncertainties in the predictions of the SM Higgs boson production cross sections nor branching fractions, are indicated for the regularized and unregularized fits. The observed cross section corresponds to the result of the regularized fit. Regularization biases in the signal strength[7] are given in the last column and not included in the total uncertainties. They are negligible in the p_T^{jet1} differential measurement. Results for $N_{jets} = 0$ appear twice: the first (second) occurence corresponds to the fit of N_{jets}-based (p_T^{jet1}-based) categories.[4] 107

9.1 Event categorization in search for anomalous couplings. 111

9.2 List of discriminants for separating anomalous couplings (BSM) from the SM contribution.[6] 114

9.3 List of observables used in each category. 124

9.4 Allowed 68% (central values with uncertainties) and 95% CL (in square brackets) intervals on anomalous HVV coupling parameters. Approaches 1 and 2 refer to the choice of the relationship between the a_i^{WW} and a_i^{ZZ} couplings.[6] 126

Acknowledgments

Words cannot express gratitude to countless people that supported my efforts throughout my Ph.D. I'd like to express my sincere gratitude to my Professors and Advisor Keti Kaadze for your patience, guidance, and support. I am extremely grateful that she took me on as a student and continued to have faith in me over the years. For the extraordinary experiences she arranged for me and for providing opportunities to grow professionally. I would like to offer my special thanks to Professor. Yurii Maravin for his invaluable support and guidance throughout my Ph.D.

I consider myself fortunate to have worked with Dr. Cecile Caillol and Professor. Isobel Ojalvo. Alongside, Dr. Adollah Mohammadi, Andrew Loelige, Dr. Senka Duric, and Dr. Kyungyook Nam.

I would also like to thank Dr. Ted Liu and Professor. Zhenyu Ye for their guidance and support through my stay at Fermilab.

Furthermore, I could not have undertaken this journey without my defense committee, who generously provided knowledge and expertise. Thank you, Professor Glen Horton-Smith, Professor. Tim Bolton, and Professor. Lizaveta Inhatsyeva. Thank you, to everyone in the K-state physics department. Additionally, I want to thank Felisa Osburn.

Finally, I want to thank my family members and close friends who have been by my side through this endeavor. And for Joanna Sanchez, thanks for all your support.

Chapter 1

Theory and motivation

1.1 Elementary particles and forces

Since the early 20th century, the theories and discoveries of a number of physicists have resulted in a marvelous insight into the fundamental structure of matter: All in the Universe is found to be made from a few basic building blocks known as fundamental particles, governed by four fundamental forces. Our current understanding is embodied in the SM of particle physics, which provides a unified picture of particles. Notably, the SM gives a successful description of all current experimental data and represents one of the triumphs of modern physics.

In the SM, the fundamental particles that make up ordinary matter contains fermions, particles with half-integer spin that obey Fermi-Dirac statistics, while the force carriers that facilitate interaction between particles is comprised of bosons, particles with integer spin that obey Bose-Einstein statistics. Fermions are grouped as leptons and quarks depending on the fundamental interactions they possessed.

Almost all commonly faced physical phenomena can be described in terms of the electron, the electron neutrino, the up quark and down quark, interacting by the electromagnetic, strong and weak forces. They are known collectively as the first generation particles. However, further complexity is revealed when particle interactions are studied at the energy

scales encountered in high-energy particle colliders. There are exactly two other copies that differ from each of the first generation particles by their masses, known as the second and third generations. Besides the differences in masses, the properties of the particles in each generation are identical because they hold the same fundamental interactions. Table 1.1 summarizes the three generations of fundamental fermions and their properties.

	Leptons			Quarks		
	Particle	Charge	Mass (GeV)	Particle	Charge	Mass (GeV)
1st	electron (e)	-1	0.0005	down (d)	-1/3	0.003
	electron neutrino (ν_e)	0	$<10^{-9}$	up (u)	+2/3	0.005
2nd	muon (μ)	-1	0.106	strange (s)	-1/3	0.1
	muon neutrino (ν_μ)	0	$<10^{-9}$	charm (c)	+2/3	1.3
3rd	tau (τ)	-1	1.78	bottom (b)	-1/3	4.5
	tau neutrino (ν_τ)	0	$<10^{-9}$	top (t)	+2/3	174

Table 1.1: *The twelve fundamental fermions grouped into quarks and leptons.*

The particles interact with each other through the four fundamental forces: the gravitational force, the electromagnetic force, the strong force and the weak force. The gravitational force between fundamental particles is extremely small and is not included in the SM. The types of interaction categorize the properties of the twelve fundamental fermions that they experience, which is shown in Figure 1.1. A brief description of these forces is given in the following sections. The Table 1.2 summarizes the relative strengths of the forces associated with the different gauge bosons. These numbers are only indicative as the strengths of the forces depend on the distance and energy scale being considered.

1.1.1 Electromagnetic force

Except for the neutrinos, which are electrically neutral, the other nine particles are electrically charged and undergo electromagnetic interaction. The interactions between charged particles are mediated by the exchange of virtual photons. Electromagnetic force is a long range force causing attraction and repulsion. It is responsible for attraction between atomic nuclei and electrons that holds atoms together or chemical bonds between atoms that create molecules.

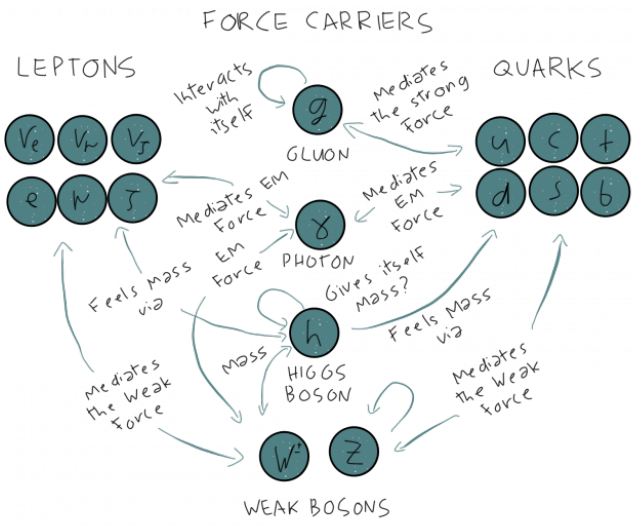

Figure 1.1: *The interactions of particles in the SM.*[1]

1.1.2 Weak force

Every twelve fundamental fermions "feel" the weak force and undergo weak interaction. The weak charged-current interaction is mediated by charged W^+ and W^- bosons which are roughly eighty times more massive than the proton. There is also weak neutral-current interaction which is mediated by the electrically neutral Z boson. Weak force is a short range force and is responsible for neuclear β-decay and nuclear fusion.

1.1.3 Strong force

Unlike leptons, the quarks carry color charges binding the quarks inside protons and neutrons, and other hadrons via strong interaction. Because the quarks feel the strong force, their properties are quite different from those of the electron, muon, τ lepton, and neutrinos, together known as the leptons. The strong force is carried by a massless and electrically neutral boson called a gluon. The uniqueness of the strong force is that its strength increase with the distance, which means the quarks inside hadrons are bound in spite of repulsion they fell due to their electric charge. This unique feature of the strong force is known as

confinement.

1.1.4 Gravitational Force

Objects with masses attract each other via gravitational force. This long-range force is responsible for many of the large-scale structures in the Universe, although it is the weakest among all forces at atomic scale. In many beyond standard model (BSM) theories, "Graviton" is considered as force carrying particle of the gravitational force.

Force	Strength	Boson		Spin	Mass/GeV
Strong	1	Gluon	g	1	0
Electromagnetism	10^{-3}	Photon	gamma	1	0
Weak	10^{-8}	W boson	W	1	80.4
		Z boson	Z	1	91.2
Gravity	10^{-37}	Graviton?	G	2	0

Table 1.2: *The four known forces of particle physics. The relative strengths are indicative values for two fundamental particles at a distance of* $1fm = 10^{-15}m$ *which is approximately the radius of a proton.*

1.2 Quantum field theory and gauge symmetry

Quantum field theory (QFT) is a natural starting point for describing the SM. QFT was started as an attempt to establish a quantum mechanics consistent with Einstein's theory of special relativity. One of the very first ideas was trying to formulate a relativistic version of the Schrödinger equation keeping the relativistic dispersion relation $E^2 = p^2c^2 + m^2c^4$ which is made by Dirac. He used matrices to make a linear equation, and this enlightenment resulted in the derivation of the Dirac equation:

$$i\gamma^\mu \partial_\mu \psi - m\psi = 0 \,, \tag{1.1}$$

where γ^μ is the Dirac matrices and ∂_μ is the gradient generalized to four dimensions. The Dirac equation is now accepted as the equation of motion for all free spin-$\frac{1}{2}$ massive fermions,

such as electrons and quarks. Dirac successfully derived the relativistic quantum theory of the electromagnetic interaction, which is now known as quantum electrodynamics (QED).

QED can be derived using the concept of gauge invariance, which describes the interaction of fermions by means of exchange of massless $U(1)$ gauge boson, or photons. The Lagrangian for a fermionic field interacting with the electromagnetic field is given as

$$\mathcal{L} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \ , \tag{1.2}$$

where $B^{\mu\nu}$ is the field strength tensor. The Lagrangian in Equation 1.2 is invariant under global phase transformation on ψ, or global $U(1)$ transformations leading the conservation of electric charge. However, the local symmetry is broken, which can be recovered by adding a new gauge field "photon field". Despite the establish of QED, still there were phenomena which cannot be explained by QED, such as β-decays. This implied the existence of a new fundamental force and led to the theory of the weak interaction which is characterized by $SU(2)$ gauge group.

In order to explain massive fermions and gauge bosons, which had been already discovered, a more elegant mechanism is introduced in the SM in order to describe both the particle masses and gauge invariance in a consistent way. This mechanism is now called spontaneous symmetry breaking, also known as the Higgs mechanism in consideration of the electroweak theory. It begin with the introduction of a scalar field Φ that couples to $U(1)$ and $SU(2)$ gauge bosons, constructing $SU(2) \times U(1)$. The scalar field can be expanded about this minimum by writing

$$\Phi = \langle\Phi\rangle + \eta = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \ . \tag{1.3}$$

The physical field $h(x)$ is named the Higgs field. With the expansion of Higgs field term, the entire Lagrangian can be rewritten breaking the symmetry of $SU(2) \times U(1)$. There are four gauge bosons emerged from spontaneous symmetry breaking. Two charged gaube bosons

are identified as the W^+ and W^- bosons. Also, there is a massless neutral gauge boson and a massive neutral gauge boson named as the photon and the Z boson, respectively. When the electroweak theory was first proposed, there were no experimental evidences for the Z boson's existence. Thus, the discovery of the Z and $W^\pm$ bosons in 1983 is now regarded as one of the triumph of the SM.

Remarkably, it can be also used to generate the masses of the fermions. The fermions acquire mass via the Higgs mechanism through Yukawa interaction. Weak isospins are the eigenvalues of $SU(2)$ and have the analogy to spins in the rotational symmetry. In the SM, left-handed chiral fermions are placed in $SU(2)$ doublets, while right-handed fermions must be singlets. Left-handed chiral fermions have the total weak isospin $I_W = 1/2$ and can be grouped into doublets with the third component of weak isospin $I_W^3 = \pm 1/2$:

$$l_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \quad q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \tag{1.4}$$

On the other hand, the representation of the right-handed singlets are

$$e_R,\ u_R,\ d_R, \tag{1.5}$$

with trivial eigenvalues $I_W = I_W^3 = 0$. The right-handed neutrinos have never been observed experimentally so far. For this reason, neutrinos are assumed to be massless in the SM, which means there are no Yukawa interactions to neutrinos, which can generate neutrino mass term including the right-handed term intrinsically. Therefore, $SU(2) \times U(1)$ can be reduced to $SU(2)_L \times U(1)$.

There is a remaining local $U(1)$ gauge symmetry related to the physical photon field after the symmetry breaking of $SU(2) \times U(1)$, which will be denoted by $U(1)_{em}$ to distinguish from $U(1)$ in $SU(2) \times U(1)$. The electric charge of particles that we can observe is indeed the quantum number of $U(1)_{em}$ (Q). On the other hand, the quantum number of the original

$U(1)$ is called the weak hypercharge $(Y/2)$. The relation between the conserved charges is:

$$Y = 2(Q - I_W^3) \ . \tag{1.6}$$

All the quantum numbers of fermions in the $SU(2) \times U(1)$ electroweak symmetry group are summarized in Table 1.2.

Particle	Hypercharge(Y)	Weak isospin charge(I_W^3)	Electric Charge(Q)
u_L	+1/3	+1/2	+2/3
d_L	+1/3	-1/2	-1/3
u_R	+4/3	0	+2/3
d_R	-2/3	0	-1/3
ν_L	-1	+1/2	0
e_L	-1	-1/2	-1
e_R	-2.	0	-1

Table 1.3: *Quantum numbers of fermions singlets and doublets.*

In the same way, the quantum numbers of the Higgs doublet can be obtained, which are summarized in Table 1.2.

Particle	Hypercharge(Y)	Weak isospin charge(I_W^3)	Electric Charge(Q)
Φ^+	+1	+1/2	+1
Φ^-	+1	-1/2	0

Table 1.4: *Quantum numbers of Higgs doublets.*

Yukawa couplings between the Higgs field and charged leptons are

$$\mathcal{L}_{Yukawa,l} = -y_l \bar{l}_L \Phi e_R - y_l \bar{e}_R \Phi^\dagger l_L \ , \tag{1.7}$$

and between the Higgs field and down-type quarks are

$$\mathcal{L}_{Yukawa,d} = -y_d \bar{q}_L \Phi d_R - y_d \bar{d}_R \Phi^\dagger q_L \ , \tag{1.8}$$

where y_l and y_d are the Yukawa coupling strength for each fermion. After expanding the Higgs

term about the minimum perturbatively, we get terms of the form $y_f \frac{v}{\sqrt{2}} \bar{f}_L f_R$ corresponding to mass term with $m_f = y_f \frac{v}{\sqrt{2}}$.

The dynamics of quarks and gluons are named the quantum chromodynamics (QCD). The symmetry of QCD belongs to $SU(3)$ of which the fundamental representatoin is a triplet. There are three conserved "color" charges, r, b and g where color is simply label for the orthogonal states in the $SU(3)$ color space. In addition, there are eight gluon states corresponding to the $SU(3)$ generators. From the renormalization procedure in QCD, running coupling constant α_s is

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + B\alpha_s(\mu^2) ln(\frac{q^2}{\mu^2})} \ . \tag{1.9}$$

where the B depends on the number of fermionic and bosonic loops. For n_f quark flavors and n_c colors,

$$B = \frac{11n_c - 2n_f}{12\pi} \ . \tag{1.10}$$

For $n_c = 3$ colors and $n_f = 6$ quarks, B is greater than zero. Taking the asymptotic limits, this leads to some strange features of QCD. In the high momentum transfer limit ($\mu \to \infty$), the coupling constant approaches zero and perturbation theory provides accurate description of asymptotically free quarks and weakly interacting gluons (asymptotic freedom). On the other hand, at the low momentum transfer limit ($\mu \to 0$), interaction between quark and gluon is too strong and they must be confined in color singlet bound state, hadrons.

Finally, we have a complete theory of the SM. In the SM, all forces are derived from gauge symmetries of the interacting particles based on $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge group, where $SU(3)_C$ describe the strong interaction while $SU(2)_L \times U(1)_Y$ describes electroweak interaction. The subscript C refers to the color, L to the left chiral nature of the $SU(2)$ coupling and Y to the weak hypercharge.

1.3 Motivation

This thesis will describe three analyses measuring the properties of the Higgs boson using the $H \to \tau\tau$ decay channel. With the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments at the LHC, all of the particles in the standard model (SM) have been observed. Despite a huge success of the SM, it is not a complete description of the Universe as there are a number of fundamental questions that are not answered yet. After its discovery, the Higgs boson became an important target for higher-precision experimental test of the predictions of the SM. Any deviations from the SM would indicate beyond SM (BSM) physics, such as new particles or interactions. Perhaps, a deeper understanding of the properties of the Higgs boson may give us hint to what lies beyond the SM.

Chapter 2

Phenomenology of the Higgs boson interactions

In the SM, once the mass of the Higgs boson is determined, the properties of the Higgs boson, decay widths, and the production mechanisms can be predicted. This chapter will discuss properties of the Higgs boson with a mass of 125 GeV and parameterize possible deviations that could be produced by beyond the standard model (BSM) effects. This sets the stage for analyses described in Chapters 7 - 9, which include precision measurement of the properties of the SM Higgs boson and search for deviations from the SM expectation.

2.1 Production modes

The Higgs bosons are produced through several different mechanism at the LHC. The kinematic of the Higgs boson and any associated particles are distinct depending on production modes, and as a result each one provides a unique handle on the Higgs boson's couplings and properties.

The four most common Higgs boson production modes at the LHC with a centre-of-mass energy of $\sqrt{s}$=13 TeV are shown in Fig. 2.1. The most dominant mechanism for Higgs boson production is gluon-gluon fusion through a heavy quark loop that is expected under

the SM to account for 87% of Higgs boson events[8]. Gluons do not couple directly to the Higgs boson because they are massless. Because the Higgs boson couples preferentially to mass, the primary contribution has a top quark in the loop, while the bottom quark also contributes a good amount to affect the p_T spectrum. At the leading order gluon-gluon-fusion production of the Higgs boson there are no association particles. At higher order, however, jets can be produced from the initial state gluons through QCD interactions which are typically soft.

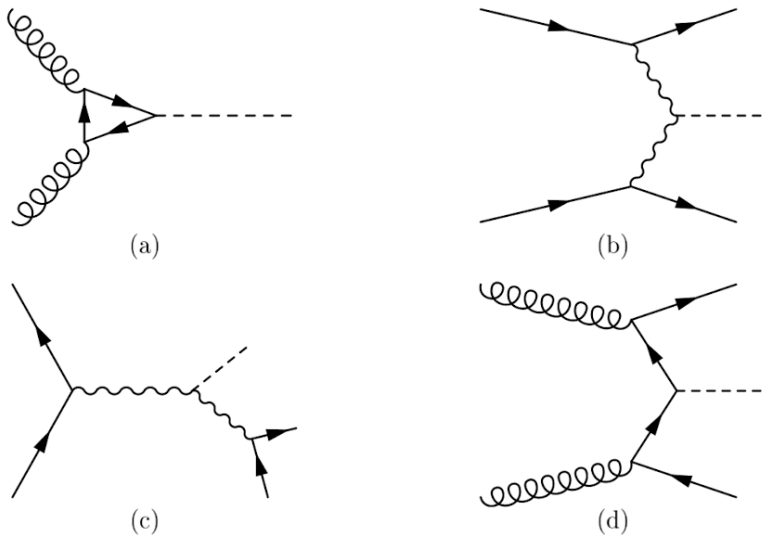

Figure 2.1: *Feynman diagram for the most common Higgs boson production modes at the LHC. (a) gluon fusion, (b) vector boson fusion, (c) associated VH productions, and (d) $t\bar{t}H$ or $b\bar{b}H$ production.*

Although the gluon-gluon fusion process has the largest cross section, from the experimental perspective the vector boson fusion, or VBF, process is also crucial which are accounts for 7% of Higgs boson events at 13 TeV. In VBF production, it results in more readily identifiable final states consisting of just the decay products of the Higgs boson and two forward jets from the break-up of the two incoming protons.

Besides that, 4% of Higgs boson events are produced through associated VH production, which also involves the interaction between the Higgs boson and vector bosons. During the process, two quarks from the incoming protons produce an offshell W or Z boson which then

emits a Higgs boson. Hence, we can observe the decay products of the W or Z boson together with the Higgs boson, which could be quarks (identified as jets), leptons, or neutrinos (identified as missing transverse energy). The invariant di-fermion mass of the associated particles is around the mass of W (80.4 GeV) or Z (91.2 GeV).

The remaining 2% is largely comprised of $t\bar{t}H$ and $b\bar{b}H$ production. In this caces, either a pair of top or bottom quark is in the final state, which can be detected by typical methods for identifying top quark decay products or b-jets. Although each of production mode contributes similarly to the cross sections, $t\bar{t}H$ is more straightforward to detect at the LHC because the top quark can provide the associated particles a large momentum due to its heavier mass. The b-jets produced in $b\bar{b}H$ are usually too soft to detect, and the p_T spectrum of $b\bar{b}H$ is very close to that of gluon-gluon fusion events.

Other production modes, such as gg→ZH or tqH, contribute less than 0.2% to the total Higgs boson cross section. Table 2.1 has the summary of the production modes mentioned above.

Production mode	Couplings	Extra particles	Contribution at 13 TeV
ggH	fermion	None or QCD jets	87 %
VBF	vector boson	two quarks	7 %
VH	vector boson	Z or W	4 %
$t\bar{t}H$	fermion	two top quarks	0.9 %
$b\bar{b}H$	fermion	two bottom quarks	0.9 %
gg→ZH	both	Z	0.2 %
tqH	both	top and light quark	0.1 %

Table 2.1: *Summary of the Higgs boson production modes discussed in this section.*

2.2 Decay modes

The Higgs boson can decay to any light enough massive particle and, through loops, to massless particles as well. The branching fractions for the primary decay modes, for a mass range around 125 GeV are shown in Fig. 2.2.

One of the basic decay of the Higgs boson is simple decays to two fermions which is shown

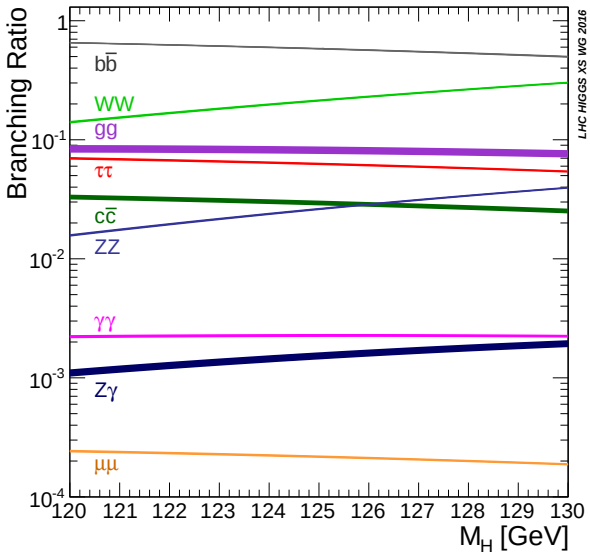

Figure 2.2: *Higgs boson branching ratios and their uncertainties for the mass range around 125 GeV*[9]

in Fig. 2.3. The Higgs boson couples to other particles with a strength proportional to their mass, hence the strongest couplings is to the heaviest fermions. However, the mass of the top quark is much higher than that of the Higgs boson which means this decay is kinematically forbidden. Accordingly, the next heaviest fermions are the most crucial fermionic decays, $b\bar{b}$ or $\tau^+\tau^-$. The H$\rightarrow b\bar{b}$ decay is the dominant Higgs boson decay channel for a mass of 125 GeV, but because this decay produces b jets, it has huge backgrounds from QCD processes which is relatively hard to isolate. On the other hand, H$\rightarrow \tau^+\tau^-$ decay has a lower branching ratio and lower background. This channel has been observed by CMS[10] and ATLAS[11]. The τ leptons may decay either to lighter lepton (e or μ) with two neutrinos or one or more mesons with a neutrino. In the chapter 7, analyses using H$\rightarrow \tau^+\tau^-$ decay channel will be discussed further.

2.3 Anomalous couplings

The SM predicts how the Higgs boson interacts with other particles. Thus far, analyses by CMS and ATLAS confirmed that the properties of the Higgs boson are consistent with the

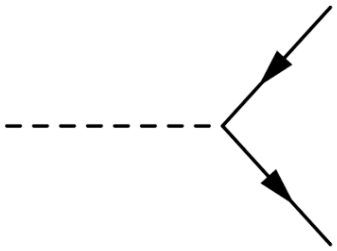

Figure 2.3: *Feynman diagram for H→ $f\bar{f}$ decay (such as $b\bar{b}$ or $\tau^+\tau^-$)*

SM. In particular, the Higgs boson was found to be a spin-0 particle[5 12] and it predominantly interacts with vector bosons through the tree level, scalar tensor structure predicted by the SM. However, additional contributions to the interaction of the Higgs boson are still possible due to the limited precision of current studies. These anomalous couplings are tiny but nonzero even in the SM via loops, where the top quarks dominates, and BSM effects can strengthen anomalous couplings. Measurements of these anomalous couplings will be discussed in detail in Chapter 9.

2.3.1 Couplings to vector bosons

Lorentz invariance limits the possible interaction between the Higgs boson and other particles. A scattering amplitude that parameterizes interactions of a spin-0 Higgs boson and two spin-1 gauge bosons VV, such as WW, ZZ, Zγ, $\gamma\gamma$, and gg, up to $\mathcal{O}\left(q^2\right)$ is written as

$$\mathcal{A}(HVV) \sim \left[a_1^{VV} + \frac{\kappa_1^{VV} q_1^2 + \kappa_2^{VV} q_2^2}{\left(\Lambda_1^{VV}\right)^2}\right] m_{V1}^2 \epsilon_{V1}^* \epsilon_{V2}^* + a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2)\mu\nu} + a_3^{VV} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2)\mu\nu} \quad (2.1)$$

where q_i, ϵ_{Vi}, and m_{V1} are the four-momentum, polarization vector, and pole mass of the gauge boson, indexed by $i = 1, 2$. The field strength tensor and the dual field strength tensor of the gauge boson are $f^{(i)\mu\nu} = \epsilon_{Vi}^{\mu} q_i^{\nu} - \epsilon_{Vi}^{\nu} q_i^{\mu}$ and $\tilde{f}_{\mu\nu}^{(i)} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} f^{(i)\rho\sigma}$ respectively[13].

In the SM, the only nonzero tree level coupling is $a_1^{ZZ} = a_1^{WW} = 2$ under custodial symmetry. The rest of the ZZ and WW couplings are considered anomalous contributions,

which are either small contributions arising in the SM due to loop effects or new BSM contributions. The Λ_1^{VV} term represents modification to this term as a function of the momenta of the involved particles and two other coupling coefficients a_2^{VV} and a_3^{VV} multiply rest two tensor structures. The a_1^{VV} and a_2^{VV} terms are CP-even couplings while a_3^{VV} is CP-odd coupling.

Generic parameterization in Equation 2.1 does not assume any specific mechanism for producing anomalous behavior. In other words, any model that predicts anomalous HVV couplings can be described by Equation 2.1. For example, a new heavy t' quark in a loop can contribute to the term with the second tensor structure. However, this term can be arisen in the SM from t quark loop that cannot be detected with the current precision.

Instead of measuring the anomalous couplings a_i directly, it is convenient to measure the effective cross section ratios f_{ai} because most uncertainties cancel in the ratio and the parameter range is conveniently bounded between -1 and 1. Furthermore, these the effective cross section ratios are invariant with respect to the coupling convention. The effective fractional cross sections f_{ai} are defined as follows.

$$
\begin{aligned}
f_{a3} &= \frac{|a_3|^2\sigma_3}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\Lambda 1} + |\kappa_1^{Z\gamma}|^2\sigma_{\Lambda 1}^{Z\gamma}} \operatorname{sgn}\left(\frac{a_3}{a_1}\right), \\
f_{a2} &= \frac{|a_2|^2\sigma_2}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\Lambda 1} + |\kappa_1^{Z\gamma}|^2\sigma_{\Lambda 1}^{Z\gamma}} \operatorname{sgn}\left(\frac{a_2}{a_1}\right), \\
f_{\Lambda 1} &= \frac{|\kappa_1|^2\sigma_{\Lambda 1}}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\Lambda 1} + |\kappa_1^{Z\gamma}|^2\sigma_{\Lambda 1}^{Z\gamma}} \operatorname{sgn}\left(\frac{-\kappa_1}{a_1}\right), \\
f_{\Lambda 1}^{Z\gamma} &= \frac{|\kappa_1^{Z\gamma}|^2\sigma_{\Lambda 1}^{Z\gamma}}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\Lambda 1} + |\kappa_1^{Z\gamma}|^2\sigma_{\Lambda 1}^{Z\gamma}} \operatorname{sgn}\left(\frac{-\kappa_2^{Z\gamma}}{a_1}\right),
\end{aligned}
\tag{2.2}
$$

where σ_i is the cross section for the process corresponding to $a_i = 1$ while all other couplings set to zero ($a_{i\neq j} = 0$)[6].

In the case of Hgg couplings, it is a purely loop-induced process through a t quark and a smaller contribution from the b quark that gives rise to the CP-even interaction based on the SM. Nevertheless, a contribution of the CP-odd interaction is not excluded. Under the assumption that only SM particles can contribute to the gluon fusion loop, a CP-structure

Coupling	σ_i/σ_1
a_3	0.153
a_2	0.361
κ_1	0.682
$\kappa_2^{Z\gamma}$	1.746

Table 2.2: *Cross sections for the anomalous contributions (σ_i) used to define the fractional cross sections*[6]*. The σ_i values are defined as the cross section computed with $a_i = 1$ and all other couplings set to zero. All cross sections are given relative to the SM value (σ_1). In the case of the κ_1 and $\kappa_2^{Z\gamma}$ couplings, the numerical values $\Lambda_1 = \Lambda_1^{Z\gamma} = 100$ GeV are considered so as to keep all coefficients of similar order of magnitude.*

measurement in the ggH process is equivalent to the measurement of the CP structure in Yukawa couplings, which can be parameterized with the amplitude

$$\mathcal{A}(\mathrm{Hff}) = -\frac{m_f}{v}\bar{\psi}_f\left(\kappa_f + \mathrm{i}\,\tilde{\kappa_f}\gamma_5\right)\psi_f. \tag{2.3}$$

The effective fractional cross section for Hff couplings is then can be written as

$$f_{CP}^{\mathrm{Hff}} = \frac{|\tilde{\kappa_\mathrm{f}}|^2}{|\kappa_\mathrm{f}|^2 + |\tilde{\kappa_\mathrm{f}}|^2}\ \mathrm{sgn}\left(\frac{\tilde{\kappa_\mathrm{f}}}{\kappa_\mathrm{f}}\right). \tag{2.4}$$

The ggH loop also can be generated by unknown heavy BSM particles as well as the SM fermions, and gauge invariance prohibits the first tensor structure in Equation 2.1. Remaining possibilities are then a_2^{gg} and a_3^{gg}. In the effective field theory (EFT) approach, a_2^{gg} and a_3^{gg} couplings correspond to two EFT couplings in the Higgs basis

$$\begin{aligned} c_{gg} &= -\frac{1}{2\pi\alpha_S}a_2^{gg}, \\ \tilde{c}_{gg} &= -\frac{1}{2\pi\alpha_S}a_3^{gg}, \end{aligned} \tag{2.5}$$

where α_S is the running strong coupling constant[9]. Therefore, there are four parameters to consider $(\kappa_t, \tilde{\kappa}_t, c_{gg}, \tilde{c}_{gg})$, where we have $(\kappa_t, \tilde{\kappa}_t, c_{gg}, \tilde{c}_{gg}) = (1, 0, 0, 0)$ in the SM. It is difficult to distinguish between the κ_f and a_2^{gg}, or between $\tilde{\kappa}_f$ and a_3^{gg} although there are small differences in the Higgs transverse momentum, p_T^H, distributions, and the effects in the off-

shell H production can be observed[14]. Therefore, in the analysis discussed in this thesis, the SM fermion loop contribution is absorbed into the overall a_2^{gg} and a_3^{gg} couplings. There is only one remaining effective fractional cross section for the Hgg interaction, which is given as

$$f_{a3}^{\mathrm{ggH}} = \frac{|a_3^{gg}|^2}{|a_2^{gg}|^2 + |a_3^{gg}|^2} \,\mathrm{sgn}\left(\frac{a_3^{gg}}{a_2^{gg}}\right) . \tag{2.6}$$

Under the assumption that only the top and b quarks contribute to the ggH loop with $\kappa_t = \kappa_b$ and $\tilde{\kappa}_t = \tilde{\kappa}_b$, the following relationship holds.

$$\left|f_{CP}^{\mathrm{ggH}}\right| = \left(1 + 2.38\left[\frac{1}{\left|f_{a3}^{\mathrm{ggH}}\right|} - 1\right]\right)^{-1} \tag{2.7}$$

The tree-level gluon fusion process shown in Figure 2.1 (a) cannot provide enough information to separate a_2^{gg} from a_3^{gg}. However, if there are two or more associated jets, their kinematics carry information that can be exploited to measure the tensor structure of the Hgg interactions. The search for such a CP-violating interaction can be performed in $t\bar{t}H$ production and $H \to \tau\tau$ decay, and the amount of data collected is just sufficient for the first measurements of this process using the $H \to \tau\tau$ channel, which will be discussed later in Chapter 9.

Chapter 3

The Experiment

In order to produce heavy particles, such as top quark or Z boson, the center of mass energy must be large enough to cover the mass. However, typically it is hard to find a given targeted process because only a tiny fraction of collisions will actually produce that process. Therefore, a large number of collisions, or high luminosity, is essential to study interesting physic processes. The Large Hadron Collideer (LHC) is designed to have the capacity to explore the entire predicted energy range where the Higgs boson could appear. The LHC accelerate particles to the record-breakingly high energies and accumulated the highest luminosity of any collider to date. Furthermore, it is also necessary to have a detector capable of measuring the energy, and momentum of the particle produced. The Compact Muon Solenoid (CMS) experiment is one of two general-purpose particle detectors at the LHC and is designed to measure particle decays to exceptional precision. The primary goals of the LHC and its detectors was to provide a definitive answer on the existence of the Higgs boson which was achieved in 2012.

3.1 The Large Hadron Collider

The Large Hadron Collider is located 100 m under the French and Swiss countryside just outside Geneva, Switzerland, shown in Fig. 3.1. It is designed to accelerate protons around

a 27km tunnel and collide them together. Several smaller accelerators are used to accelerate the protons to consecutively higher energies until the Super Proton Synchrotron (SPS, shown in Fig. 3.1) takes them to provide beams for the LHC at an energy of 450 GeV. Two high-energy particle beams travel in opposite directions inside the accelerator. All around the accelerator ring, as series of 1232 superconducting dipole magnets are accommodated and steer the protons through the curves of the tunnel, then 392 quadrupole magnets are used to focus the beams. Lastly, the beams are more squeezed by another type of quadrupole magnets in order to increase the chance of collisions just prior to collision. The beams inside the LHC are brought together to collide at four locations around the accelerator ring with a center of mass energy of 13 TeV.

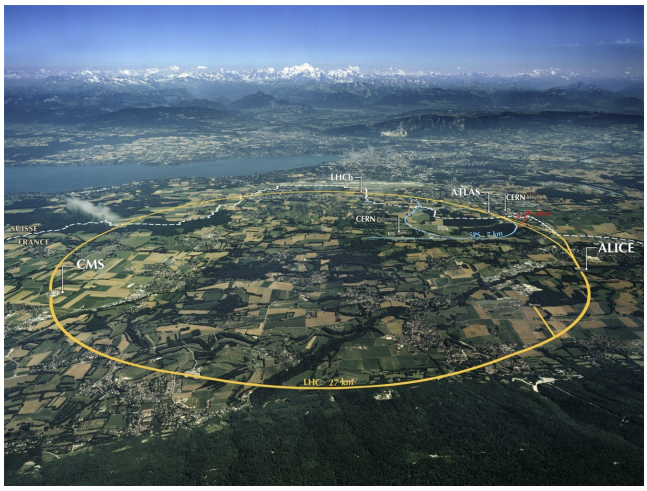

Figure 3.1: *A picture of the LHC with its path shown in a circle and the interaction points labeled with the experiments located there. (Image: Maximilien Brice/CERN)*[15]

The four particle detectors (ATLAS, CMS, ALICE and LHCb) gather clues about the particles at each of the four interaction points. ATLAS and CMS are the two general-purpose detectors at the LHC to study broad range of processes. LHCb experiment is a detector dedicated to quark flavor physics. In particular, it is optimized for investing the slight differences between matter and antimatter by studying b quark. ALICE is specialized in studying processes in heavy-ion collisions.

During norminal operations, the bunches cross every 25 ns. Approximately 2800 bunches

circulate in the tunnel at a time, and each of beam of a protons consists of bunches that contains around 10^{11} protons. Around 37-38 collisions occurs in each bunch crossing, which results in a total of 1.5 billion collisions per second.

Luminosity is a key indicator of an accelerator's performance. It is proportional to the number of collisions that occur in a given time meaning the higher the luminosity, the more data the experiments can gather. The delivered luminosity has increased due to accelerator upgrades, as shown in Fig. 3.2.

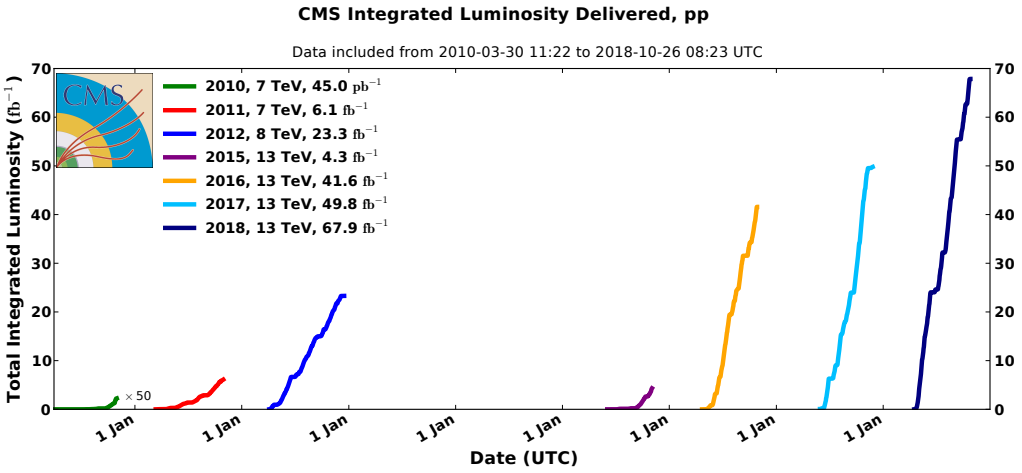

Figure 3.2: *Delivered luminosity versus day delivered to CMS during stable beam for pp collision at nominal center-of-mass energy. This shown for data-taking in 2010 (green), 2011 (red), 2012 (blue), 2015 (purple), 2016 (orange), 2017 (light blue), and 2018 (navy blue). These plots use the best available offline calibrations for each year.*[16]

3.2 The CMS Experiment

The CMS detector is designed to cover as much angular space as possible in order to detect almost all of the interactions that take place within its volume. The CMS detector gets its name from the fact that:

- CMS really is quite *compact*. Despite CMS has a volume less than a sixth that of ATLAS, it weighs nearly twice as much.

- CMS is optimized to detect *muons* very accurately.

- CMS has the most powerful *solenoid magnet* ever made (3.8 T).

The detector has several concentric layers of components, each sensitive to different types of particles depending on their properties. First of all, the trajectory of all charged particles is identified by the innermost layer, the tracker. The electromagnetic calorimeter (ECAL) is the inner layer of the two calorimeters and measures the energy of electrons and photons. Hadrons fly through the ECAL and are absorbed by the outer layer of the two calorimeters, called the Hadron Calorimeter (HCAL).

Muons pass through both tracker and calorimeters then are detected by muon system, special sub-detectors for muons. The muon system provides a second measurement of momentum of muons, which can match to tracks in the tracker. Fig. 3.3 shows a cutaway view of the CMS detector and its parts.

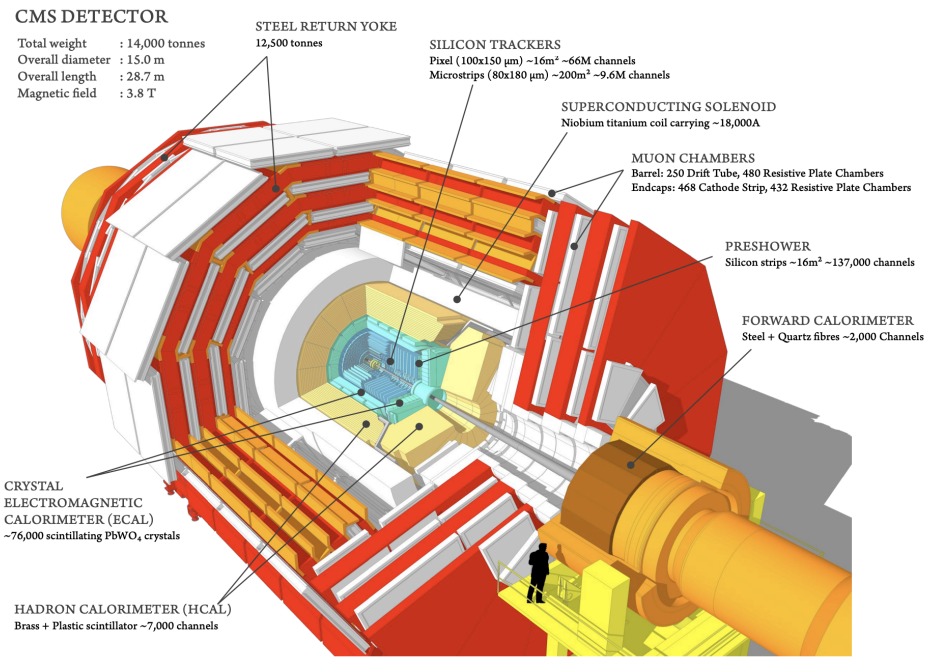

Figure 3.3: *A cutaway view of the CMS detector.*[17]

3.2.1 Coordinate system

The detector for a collider experiment should be able to wrap around and enclose the collision point as much as possible. Therefore, it is natural to have cylindrical geometry as shown in Fig. 3.3. The beam goes along the orange beam pipe which is designated as the z axis. The x axis points toward the center of the accelerator ring and y axis points up towards the sky. While collisions are azimuthally symmetric, cosmic rays, which are utilized in calibration, usually travel in an nearly vertical direction.

The spherical coordinates (r, θ, ϕ) can be defined with r as the radial distance from the beamline. The azimuthal angle ϕ is measured in the x-y plane from the x axis and polar angle θ is measured from the z axis in the plane orthogonal to the x-y plane as well as z axis. Unlike azimuthal angle ϕ, the radial angle θ is occasionally used, but is instead used to define the more widely used pseudorapidity η:

$$\eta = -\ln\left[tan(\frac{\theta}{2})\right] \tag{3.1}$$

The pseudorapidity is a good approximation of the rapidity y

$$y = \frac{1}{2}\ln\left[\frac{E + p_z}{E - p_z}\right] \tag{3.2}$$

in the high energy limit ($E \gg m$). The difference of the rapidities of two particles is invariant under a Lorentz boost in the z direction, which are frequently used in order to work in the center-of-mass frame of a collision. Besides, the number of particles produced in between η and $\eta + \Delta\eta$ is roughly constant as a function of η.

3.2.2 Magnet

The CMS magnet surrounds the silicon trackers and calorimeters. It produces an almost constant magnetic field of 3.8 T, along the beam pipe of the LHC as shown in Fig. 3.4. The magnetic field deflect every particle that has a magnetic moment, and the radius of the trajectory is proportional to the momentum of the particles. A stronger magnetic field also

produce smaller curvature, allowing a more precise momentum measurement. The magnetic field strength was one of the motivations for the CMS detector design choices.

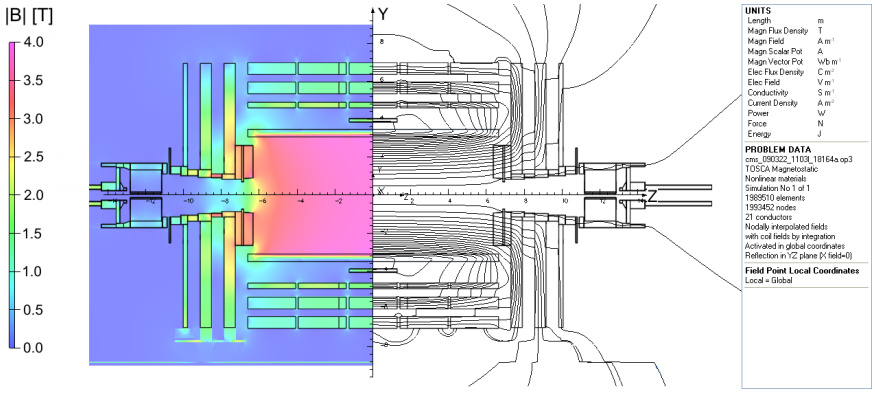

Figure 3.4: *Magnetic field (left) and field lines (lines) predicted on longitudinal section of the superconducting solenoid magnet in the CMS detector.*[18]

The magnetic field is produced by a superconducting solenoid magnet, 12.9 m long and an inner diameter of 5.9 m. It carries a current of 19.5 kA. A 10,000 ton iron return yoke acts as a filter to ensure that only muons and invisible particles goes into the muon chambers.

3.2.3 Parts of the CMS Detector

Silicon tracker

The tracker is the closest layer to the beam pipe out of all sub-detectors. It records the tracks of all kind of charged particles, which can be utilized to estimate their momentum with a great precision, or to reconstruct secondary vertices from long-lived particle decays, like b quarks and τ leptons. It is cylindrical in shape, of length 5.8 m, and has a diameter of 2.6 m. It is essential that the tracker has a high level of granularity because of the large number of particles produced in overlapping proton-proton collisions. While the tracker had to have a high power density of electronics and to be radiation resistant, the amount of material used had to be kept low to limit photon conversion, bremsstrahlung, multiple scattering and nuclear interactions.

The inner-most tracking material composed of four barrel layers[19] of silicon pixel detec-

tors (BPIX), and two forward/backward disks (FPIX) at longitudinal positions. The pixel detectors receive the most hits and require the highest precision. Directly outside the pixel detector is composed of the silicon strip detector with 10 layers of silicon strip in the barrel region, 3 layers in the inner endcap region, and 9 layers in the outer endcap region. The silicon strips provide coarser spatial resolution than the silicon pixels, but are essential to provide full coverage without gaps.

The combination of pixels and strips provides measurements along trajectory of particles. This information is used to measure the momentum of the particle, reconstruction of tracks and vertices from the primary and secondary vertices with very high accuracy.

Electromagnetic calorimeter

ECAL is the inner layer of the CMS calorimeter system. It measures the energy of photons and electrons. ECAL provides the only measurement of energy and momentum for photons while the momentum measurement from tracker complements ECAL measurement for electron. Therefore, the presence or absence of a track in the tracker enable CMS to distinguish between photons and electrons. ECAL is constructed by a dense material called lead tungstate ($PbWO_4$) crystal scintillators that has a radiation length of about 8.99 mm. The barrel region of ECAL is constructed from 61200 lead tungstate crystal and 7323 more in the endcap. As an electron or positron pass through these crystals, it radiates a photon in the electrostatic fields of a nucleus. This process is called *bremsstrahlung* which is the primary energy loss mechanism of electrons in the multi-GeV range. Similarly, photons with high energy can convert into an electron-positron pair in the field of the heavy nucleus resulting in a cascade that is known as an electromagnetic shower. Consequently photons with low energy can be produced and travel to the end of ECAL, where they are captured and measured. Furthermore, photomultipliers are used in the form of avalanche photodiodes and vacuum phototriodes in the barrel and endcap region respectively to amplify the signal. The average distance over which an electron's energy decrease by a factor of 1/e is called the radiation length. In case of photons, it is roughly 7/9 of the mean free path of the electron-

positron pair production. ECAL's width designed to be around 25 times the radiation length, so that electrons and photons deposit all of their energy in ECAL and do not affect the measurements in the rest of the detector. Other particles, such as muons or even the lightest hadrons, are significantly heavier than electrons. Thus, they do not lose a significant amount of energy via bremsstrahlung in ECAL and can continue to the outer parts of the detector.

Hadronic calorimeter

The succeeding subdetector, the hadronic calorimeter or HCAL, measures the energy of hadrons. The measurements from charged hadrons can be matched to tracks in the tracker, while neutral hadrons, such as neutrons, only leave energy deposits in HCAL.

HCAL is designed as a sampling calorimeter in actuality, therefore, not all of the energy is measured. It is formed from the alternating brass absorber and plastic scintillator layers. The nuclear interaction length of the brass absorber is 16 cm. Charged hadrons lose their energy by the ionisation process as they pass through HCAL. In addition, both charged and neutral hadrons interact with the nuclei and proceeds through the strong force. The scintillator layers measure the energies of the hadronic particles that traverse through them. The energy of the original particle can be determined by this captured energy.

HCAL is divided into barrel and endcap regions as the other sub-detectors. Additionally, HCAL has an outer region outside the solenoid that designed to capture the particles that missed by the barrel region. Also, there is a forward region to capture radiation that travels almost parallel to the beamline. The width of the barrel region is approximately 10 times the nuclear interaction length, which allow to captures almost all of the energy of hadronic particles. At the end of HCAL, only muons and neutrinos left.

Muon system

Muons are the only remaining charged particles that are not absorbed by two calorimeters. The muon system is designed to provide precise measurement of muons, which is the only

particles that leave tracks in it. To detect muons drift tubes (DTs) are used in the barrel region, while cathode strip chambers (CSCs) are used in the endcap region. Within the barrel and endcap regions, there are also resistive plate chambers (RPCs). All three types of chambers are filled with gas and muons ionize the gas atoms when they pass through. Electrons drift to wires in the DTs, and both the electrons and ions are collected by arrays of cathode and anode wires in the CSCs. In the RPCs, the electrons are collected by strips, therefore, RPCs are especially precise at determining the time of arrival of muons. On the other hand, DTs and CSCs have a better spatial resolution.

3.2.4 Trigger

Over a billion collisions per second at the LHC, and most of those collisions produce nothing interesting. It is unfeasible to store all of these collisions data. As a consequence, a solution is to discard the useless event and record the interesting events. This has to be fast enough so that the events can be saved or dropped.

A two-level triggering system is used to achieve this goal. First one is the fast level 1 (L1) trigger, using loose and raw measurements to decide whether to keep or discard the event. All is done using custom hardware which reduces the rate to about 100,000 events per second. Secondly, events that pass the L1 trigger are then transferred to a higher-level trigger (HLT), which has access to more information about the event. About 100 events per second pass HLT trigger. Only events that pass both triggers can be saved and used in analyses.

3.2.5 Particle identification and reconstruction

The particle flow algorithm is used to construct a full picture of a collision event using all of the information from each subdetector. During LHC Run2 period (2016 2018), each event involves an average of 37-38 collisions which may produce particles in overlapping regions of the detector. As a consequence, a method to sort the particles and match the information between different parts of the detector is necessary. Figure 3.5 showcases a summary of how

each type of particle interacts with the various parts of the CMS detector which are detailed in Section 5.1.

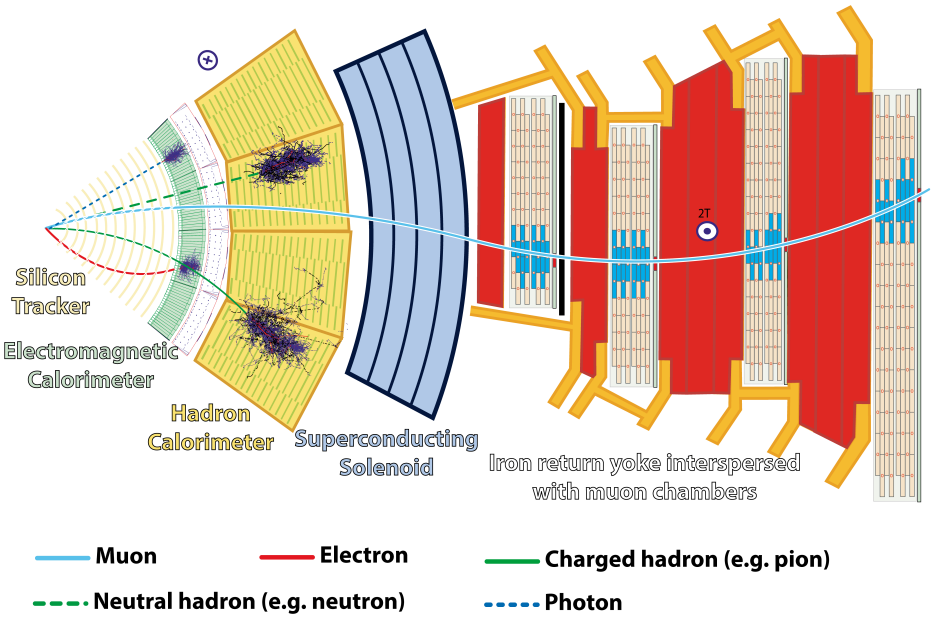

Figure 3.5: *A cross sectional view of the CMS detector.*[2]

Chapter 4

Statistics

The physics analyses presented in the following chapters heavily rely on statistical tools and interpretation. In Section 4.1, the notion of likelihood is introduced, and the way to include systematic uncertainties is described. In order to find the values of the parameters that give the best match between prediction and data maximum likelihood fits are performed, which are detailed in Section 4.2. The value of a parameter of interest, such as the signal strength can be extracted, or the validity of a model from the pulls of the nuisance parameters can be checked by using the maximum likelihood fits. To set upper limits on a signal process, the CL_s method, described in Section 4.3, uses ratios of likelihoods. Goodness-of-fit tests are a kind of generalized chi-square test that measure the agreement between data and predictions. They are used to check the background modeling, described in Section 4.4.

4.1 Likelihood

4.1.1 Basics

A counting experiment follows a Poisson law, which is a discrete probability law describing the repartition of the number of events occurring in a fixed interval of time if their average rate, λ, is known and if they occur independently of the time since the last event. The Poisson probability function is given by Equation 4.1.

$$f(n) = \frac{e^{-\lambda}\lambda^n}{n!} \tag{4.1}$$

It can be shown that the best estimation of the average rate λ is the expected number of events in the given time interval.

For a simple counting experiment, the likelihood is simply given as Equation 4.2 when n events are observed while b events are expected. The likelihood, therefore, quantifies the agreement between expectation and the observation.

$$\mathcal{L}(n|b) = \frac{e^{-b}b^n}{n!} \tag{4.2}$$

In the analyses in this thesis, data are binned in histograms. In such cases, each bins can be considered as independent of each other and the likelihood in simply given by the product of the likelihoods of every bins of the histograms.

$$\mathcal{L}(\vec{n}|\vec{b}) = \prod_{i=1}^{N} \frac{e^{-b_i}b_i^{n_i}}{n_i!} \tag{4.3}$$

Where N is the total number of bins, $\vec{b}$ is the vector of the expected processes in the different bins, and $\vec{n}$ is equivalent for the observed data. If, in contrast, we have unbinned distributions which are described by a probability distribution function (pdf) $f_b(x)$ of observable x, the likelihood is[20]

$$\mathcal{L}(\vec{x}|b, f_b(x)) = k^{-1} \prod_{i=1}^{k} b f_b(x_i) e^{-b} \tag{4.4}$$

if b events are expected and k events are observed in the full range of x.

4.1.2 Introducing systematic uncertainties

Systematic uncertainties are handled by nuisance parameters which are not of immediate interests but which influence the model. In that, they must be accounted for in the analyses. Systematic uncertainties come from the following three sources in the analyses presented

later.

- Theoretical uncertainties: uncertainties on the cross section, branching ratio, etc.
- Statistical uncertainties: uncertainties from the limited number of events in the MC simulations.
- Experimental uncertainties: uncertainties from luminosity or trigger efficiency measurements, etc.

They can be considered in the likelihood[20]. The number of expected events can be written as $b(\vec{\theta})$ because it is impacted by the nuisance parameters $\vec{\theta}$. Introducing the probability density function $p(\theta|\tilde{\theta})$ where $\tilde{\theta}$ is the best estimate of the nuisance, the likelihood becomes

$$\mathcal{L}(\vec{n}|\vec{b}) = \prod_{i=1}^{N} \frac{e^{-b_i} b_i^{n_i}}{n_i!} \prod_{j=1}^{M} p(\theta_j|\tilde{\theta}_j) \tag{4.5}$$

in the case of a binned histograms with M nuisance parameters. The Bayesian probability $p(\theta|\tilde{\theta})$ can be approximated as a function of the frequentist probability $\rho(\theta|\tilde{\theta})$, as given by the Bayes' theorem.

Most systematic uncertainties could be described by a Gaussian pdf of the type

$$\rho(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(\theta-\tilde{\theta})^2}{2\sigma^2}\right) \tag{4.6}$$

which well describes uncertainties on parameters that can be both positive and negative. But this has the inconvenient side-effect for positively defined observables. Alternatively, log-normal pdfs of parameter κ are preferred, which have longer tails than Gaussian distributions for comparable uncertainties and go to zero at $\theta = 0$ in order to avoid difficulties of the truncated Gaussian.

$$\rho(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi} ln(\kappa)} exp\left(-\frac{(ln(\theta|\tilde{\theta}))^2}{2(ln\kappa)^2}\right)\frac{1}{\theta} \tag{4.7}$$

While the Gaussian with a relative uncertainties ϵ and the log-normal with $\kappa = 1+\epsilon$ are asymptotically identical for small uncertainties, the log-normal pdf clearly behaves more

appropriately for large uncertainties as illustrated in Figure 4.1 (left).

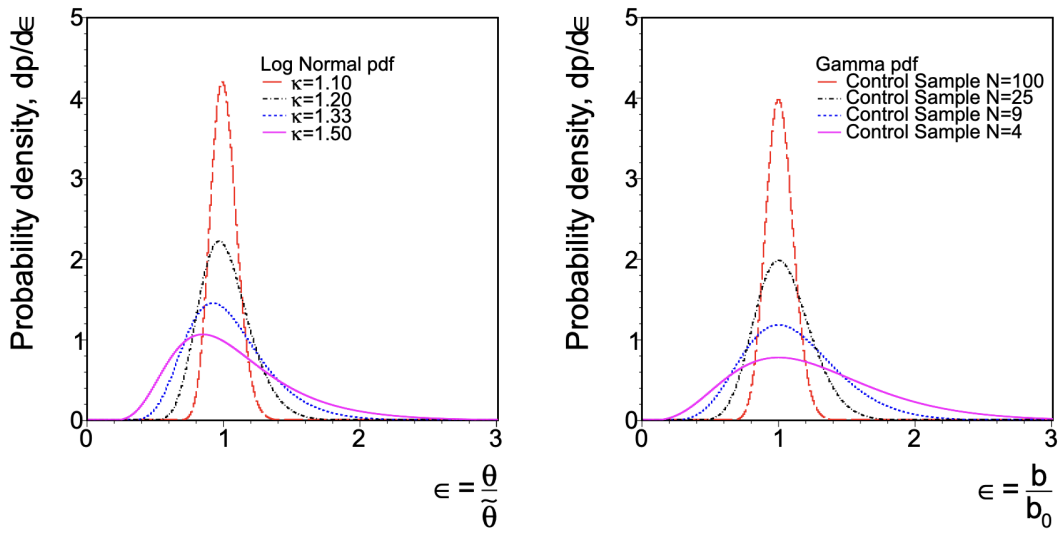

Figure 4.1: *Log-normal (left) and Gamma (right) distributions for various parameter values.*

In the case of uncertainties coming from statistically limited numbers of events in simulation or a data control sample, gamma distributions are used. If the event rate n in the signal region is proportional to the small number N in MC or data via a relationship $n = \alpha \cdot N$, the gamma distribution reads

$$\rho(n) = \frac{1}{\alpha}\frac{(n/\alpha)^N}{N!}exp(-n/\alpha). \tag{4.8}$$

The gamma distributions are shown in Figure 4.1 (right) for a given set of parameter N.

In addition, shape uncertainties have an impact on the distribution of the parameter of interest[21] and are modeled with a linear extrapolation method[22]. In practice, shape uncertainties are implemented by providing two different shapes corresponding to the variation by ± 1 standard deviation of the nuisance parameter. A parameter θ is added to interpolate smoothly between the two alternative shapes in the likelihood. If there are N shape nuisance parameters modeled with the parameters $\vec{\theta} = (\theta_1, \theta_1, ..., \theta_N)$, the histogram distribution is

given by

$$h(\vec{\theta}) = h_0 + \sum_{j=1}^{N} (a(\theta_j)h_j^+ + b(\theta_j)h_0 + c(\theta_j)h_j^-) \tag{4.9}$$

where h_0, h_j^+ and h_j^- correspond to the nominal histogram, the histogram for a variation by $+1$ standard deviation of the jth nuisance parameter and the histogram for a variation by -1 standard deviation of the jth nuisance parameter respectively with

$$a(\theta) = \begin{cases} \theta(\theta+1)/2 & \text{if } |\theta| \leq 1 \\ 0 & \text{if } \theta < -1 \\ \theta & \text{if } \theta > +1 \end{cases}, \tag{4.10}$$

$$b(\theta) = \begin{cases} -\theta^2 & \text{if } |\theta| \leq 1 \\ -(|\theta|-1) & \text{if } |\theta| < 1 \end{cases}, \tag{4.11}$$

and

$$c(\theta) = \begin{cases} \theta(\theta-1)/2 & \text{if } |\theta| \leq 1 \\ 0 & \text{if } \theta > +1 \\ |\theta| & \text{if } \theta < -1 \end{cases}. \tag{4.12}$$

It is worth noting that the effect of various shape uncertainties is additive.

A special kind of shape uncertainty, called bin-by-bin (bbb) uncertainties, is associated to the uncertainty on the number of MC events in every bin of the distribution[23]. A nuisance parameter is introduced for every bin of every process separately, then multiplies the number of expected events in this particular bin for a given process. Bin-by-bin uncertainties affect the shape of the distribution, but the behavior of each bin is independent from the behavior of the others for a single process.

4.2 Maximum likelihood fit

In the analyses presented later, maximum likelihood fits are performed to find the parameters of interest that give the best agreement between expectation and observation. Commonly, the background-only and signal-plus-background hypotheses are used in the fits. In case of background-only fit, the nuisance parameters are varied to the value $\hat{\vec{\theta}}$ that maximize the likelihood $\mathcal{L}(\vec{n}|\vec{b},\vec{\theta})$. On the other hand, the nuisance parameters and the freely floating signal strength μ of the expected signal distributions $\vec{s}$ are varied together to their optimal values $\hat{\vec{\theta}}$ and $\hat{\mu}$ to maximize the likelihood $\mathcal{L}(\vec{n}|\mu\vec{s}+\vec{b},\vec{\theta})$ in case of signal-plus-background fit. The difference between the nuisance parameters before and after a maximum likelihood fit are called pulls. Abnormally large pulls may indicate an incoherence in the background or signal modeling.

4.3 Exclusion limits

When no significant excess of data is observed over the expected backgrounds, it is natural to set an upper limit on the cross section of a hypothetical signal. Typically, the CL_s method[20],[24] is used in CMS physics analyses. The test statistic is defined using a likelihood ratio

$$q_\mu = -2ln\frac{\mathcal{L}(\vec{n}|\mu\vec{s}+\vec{b},\vec{\theta_\mu})}{\mathcal{L}(\vec{n}|\hat{\mu}\vec{s}+\vec{b},\vec{\theta})}, \text{where } 0 \leq \hat{\mu} \leq \mu. \tag{4.13}$$

In this method, the optimal signal strength $\hat{\mu}$ that maximizes the likelihood is constrained to be in between zero and the signal strength μ, for which the test statistic is computed, to avoid negative limit. While the signal strength and the nuisance parameters can both float in the denominator, the signal strength is fixed in the numerator when the likelihood is maximized. The observed value of the test statistic, q_μ^{obs}, can be determined by assuming a value of the signal strength μ from Equation 4.13. The nuisance parameters $\hat{\vec{\theta}}_\mu^{obs}$ also can be computed where $\mu = 0$ in the background-only hypotheses.

The probability density functions of the test statistics are estimated from toy MC pseudo-data, using the optimal value of the nuisance parameters, $\hat{\vec{\theta}}_0^{obs}$ and $\hat{\vec{\theta}}_\mu^{obs}$, computed in the ear-

lier step. Therefore, the probability to obtain the observation under signal-plus-background hypothesis is

$$p_{\mu s+b} = P(q_\mu \geq q_\mu^{obs}|\text{signal-plus-background}) = \int_{q_\mu^{obs}}^{\infty} f(q_\mu|\mu\vec{s}+\vec{b},\hat{\vec{\theta}}_\mu^{obs})dq_\mu \tag{4.14}$$

with the probability density function $f(q_\mu|\mu\vec{s}+\vec{b},\hat{\vec{\theta}}_\mu^{obs})$ and the probability to obtain the observation under background-only hypothesis is

$$1-p_b = P(q_\mu \geq q_\mu^{obs}|\text{background-only}) = \int_{q_\mu^{obs}}^{\infty} f(q_\mu|\vec{b},\hat{\vec{\theta}}_0^{obs})dq_\mu \tag{4.15}$$

with the probability density function $f(q_\mu|\vec{b},\hat{\vec{\theta}}_0^{obs})$. Then the CL_s value is given by the ratio of two probabilities

$$CL_s(\mu) = \frac{p_{\mu s+b}}{1-p_b} \tag{4.16}$$

for a given signal strength μ. This allows to treat cases where the signal is small enough to make both hypotheses are compatible with the observation, or where a deficiency in data would result in a negative signal strength with large significance if only $p_{\mu s+b}$ was considered.

4.4 Goodness-of-fit test

Goodnees-of-fit (GOF) tests are performed to check the agreement between the observation and expected processes. They are tests of the null hypothesis when the alternative hypothesis is not specified. In *ssaturated model*[25], an alternative hypothesis for which each bin of the distribution is exactly match to the observed data, and corresponds to a likelihood $\mathcal{L}_{sat}(\vec{n}|\vec{n})$ is taken. The test statistic in a saturated GOF test is then

$$q_\mu = -2ln\frac{\mathcal{L}(\vec{n}|\mu\vec{s}+\vec{b},\vec{\theta})}{\mathcal{L}_{sat}(\vec{n}|\vec{n})} \tag{4.17}$$

Using Poisson pdfs

$$q_\mu = -2ln\left[\prod_{i=1}^{N}\left(\frac{\mu s_i(\vec{\theta}) + b_i(\vec{\theta})}{n_i}\right)^{n_i} exp(-(\mu s_i(\vec{\theta}) + b_i(\vec{\theta})) + n_i)\right]. \tag{4.18}$$

The observed value of the test statistic q_{obs} can be obtained by minimizing q_μ.

The observed test statistic is tested whether it is compatible with what could predicted given the background (or signal-plus-background) expectations using toy MC pseudo-datasets generated according to the likelihood for given hypothesis generated. A minimal test statistic q_μ is obtained from each pseudo-dataset to generate distribution. q_{obs} is then compared to the q_μ distribution. If q_{obs} sit in the middle of the distribution, it means that data is compatible with expectation. On the other hand, if it lies in the tails of the distributions, one need to check if the expected processes are mismodeled. An example of GOF test which shows good agreement between data and expectation is shown in Figure 4.2.

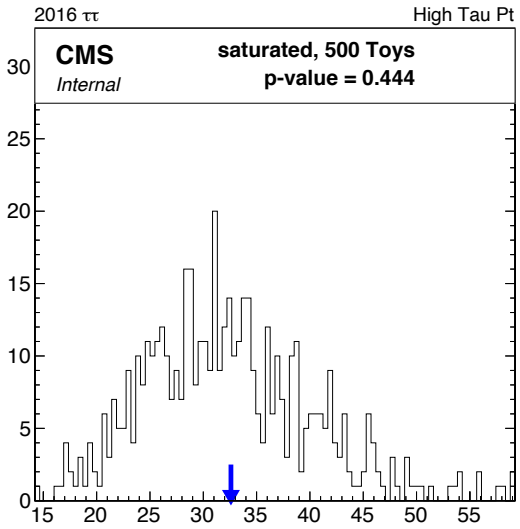

Figure 4.2: *Example of goodness-of-fit test. The observed value of the test statistic is represented by the blue arrow and black distribution is obtained from toy datasets.*

4.5 Chapter summary

Statistical methods are used to extract results from data analyses in the CMS experiment. A specific test statistics defined for LHC experiments using likelihood functions associated to data and predicted background and signal distributions. Parameter of interest, such as the signal strength, can be extracted by performing maximum likelihood fits. The likelihood also permit to check the agreement between data and predicted processes with goodness-of-fit tests. These tools are exploited in the analyses described in Chapters 7-9.

Chapter 5

Object reconstruction and analysis tools

This chapter reviews the tools that are common to the three analyses described in this thesis. Section 5.1 begins with a description of the various algorithms used to reconstruct physics objects then Section 5.2 outlines the datasets used in the analyses and the triggers used to record the events. Section 5.3 and Section 5.4 describe the simulation of proton-proton collision events and how to correct the simulated events for the better description of the data. Finally, the methods used to model backgrounds are detailed in Section 5.5 and Section 5.6 describes algorithm that used to construct di-tau invariant mass.

5.1 Object reconstruction

The reconstruction of τ lepton is particularly challenging because it is an unstable particle and decays quickly to lighter particles always accompanied by τ lepton's partner, the tau neutrino. Its decay products include also charged and neutral hadrons, electrons, and muons, which are visible. Neutral hadrons from τ lepton are usually decaying to a pair of photons, while charged hadrons from τ lepton decay are producing a shower of particles, or jet, in the calorimeter system. This section describes the algorithms used to reconstruct physics

objects in the CMS detector. Especially the objects used for the analyses described in this thesis are highlighted.

5.1.1 Tracks and vertices

Particle tracks are reconstructed using the Combinatorial Track Finder (CTF) algorithm in CMS[26]. This algorithm performs the following four steps.

1. Track seeds are generated using hits in the first few layers of the tracker (only 2 or 3 hits). A seed defines an initial estimate of the trajectory parameters and corresponding uncertainties.

2. The seed trajectories are extrapolated using a Kalman filter (KF)[27]. This extrapolates outwards along the expected flight path of a charged particle, and searches for additional hits in successive detector layers. At each layer, associated hits are added to the track candidate and the trajectory parameters are updated. This process continues until it reaches to the final detector layer.

3. The final trajectory is fitted by means of a KF and smoother iteratively. Spurious hits are looked for and discarded after each iteration. This continues until no more spurious hits are found. This track-fitting module provides the best possible estimates of the trajectory parameters.

4. Tracks are finally required to pass a set of quality flags and any failing tracks are removed.

A series of six iterations of the track reconstruction algorithm are performed. The initial iteration start with easy to identify tracks (e.g., of relatively high p_T, and/or near the interaction region). The subsequent iterations are then used more difficult classes of tracks (e.g., low p_T, or far from the interaction region). Hits associated with tracks are removed after each iteration so that the combinatorial complexity can be reduced making the latter

iterations become simpler. Measured tracking efficiencies for muons and pions with $p_T > 500$ MeV are greater than 99.3% and 98.5% respectively using $\sqrt{s} = 7$ TeV collision data[28].

The position of all interaction vertices, including the primary vertex as well as additional pile-up vertices can be determined after the reconstruction of tracks. It starts by selecting promptly produced tracks that are close to the center of the luminous region, which is known as the beamspot. Tracks that appear to originate from the same vertex then clustered together using a deterministic annealing algorithm[29], which identifies vertex candidates and assigns tracks to them. The vertex is kept if it has at least two of the associated tracks, which are incompatible with other vertices. Vertex candidates are then fitted using an adaptive vertex fitter[30] to find the best estimate of the 3D vertex positions. The vertex with the largest value of transverse momentum of the summed physics objects is defined as the primary interaction vertex. The physics objects are defined as jets clustered from tracks associated with the candidate vertex using anti-k_T algorithm[31] and the associated missing transverse energy given as the negative vectorial sum of the p_T of those jets[32].

5.1.2 Particle flow

The reconstruction of all stable particles in an event, such as electrons, photons, muons, and hadrons, is performed using the particle flow (PF) algorithm[33]. The PF algorithm performs a correlation of the basic elements (i.e. traks and calorimeter clustes) by combining the information from all of the CMS sub-detectors in order to achieve superior measurements of particle energies, directions, and types. The output of the algorithm is a list of particles called PF candidates that may then be used to build higher-level objects, such as jets, taus that decay hadronically, and missing transverse energy E_T^{miss}.

5.1.3 Muon

Muons leave hits in the inner tracker and in the muon system. Threfore, muons are reconstructed by matching tracks in the inner tracker to tracks in the muon system[34]. The tracks are reconstructed independently first. Track reconstruction in the muon system, or stan-

dalone muon tracks, starts from seeds consisting of CSC or DT segments, then extrapolated using a KF to gather hits from DT, CSC, and RPC subdetectors.

There are two approaches to match tracker and standalone muon tracks: tracker muon reconstruction, and global muon reconstruction. The tracker muon reconstruction is an inside-out approach. First tracks in the inner tracker with $p_T > 0.5$ GeV and total momentum $p > 5$ GeV are selected and then extrapolated to the muon system. If the tracker track is matched to at least one muon segment, a muon candidate can qualify as a tracker muon. In contrast, the global muon reconstruction is an outside-in approach, begins with standalone muon tracks then matches them to tracks in the inner tracker. Two tracks are matched by comparing parameters propagated to a common surface. If the tracks are matched, a combined fit with the KF is performed using the hits from both the tracker and standalone muon tracks.

While global muons typically require matching to at least two muon segments in different muon stations, tracker muons require only one muon segment. Thus, tracker muon reconstruction is more efficient than the global muon reconstruction for soft muon because they will penetrate through to the first muon station only. As shown in Fig. 5.1, muons with $p_T > 200$ GeV benefits from the global muon reconstruction, while below 200 GeV the momentum resolution is driven by the inner tracker.

Fake muons is given rise to hadron shower remnants that reach the muon system (hadron punch-through). Fake rate is small for global muons but can be larger for tracker muons because they may include hits from the innermost muon station only. This background is largely from real muons produced by, for instance, in-flight hadron decays, and heavy-flavor decays and reducible by requiring additional identification and isolation criteria. The cut based muon identification discriminant is used for the analyses covered in this thesis which is defined as *medium* ID[32].

Isolation discriminants further reduce the background. The muon isolation is defined as scalar sum of the energy deposited within a cone with size $\Delta R = 0.4$ centered on the muon direction[36].

$$I = \sum P_T^{h^\pm} + max(0, P_T^{h^0} + P_T^{\gamma} - \Delta\beta P_T^{h^{\pm,PU}}) \tag{5.1}$$

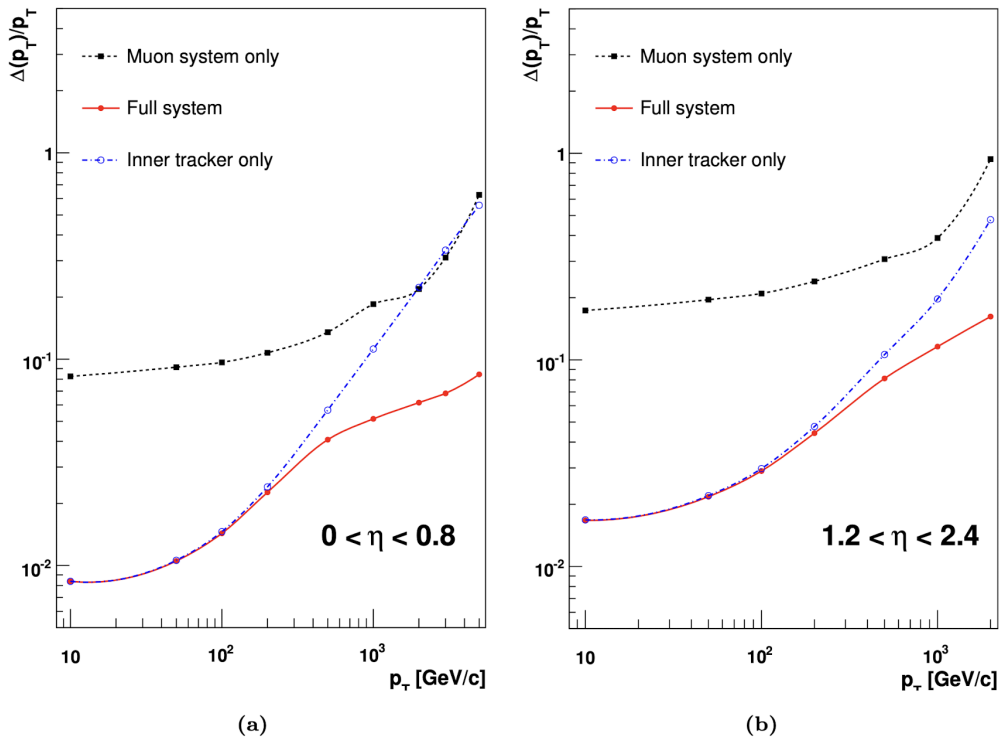

Figure 5.1: *The muon p_T resolution as a function of p_T is shown for the region $|\eta| < 0.8$ (left) and $1.2 < |\eta| < 2.4$ (right), using only the muon system (black), only the inner tracker (blue) and both subdetectors (red)*[35]

where the sums runs over all charged hadrons from the primary vertex ($h^{\pm}$) and other vertices ($h^{\pm,PU}$), neutral hadrons (h^0), and photons (γ). The contribution of pile-up to the neutral energy sums is estimated using $h^{\pm,PU}$ multiplied by the factor $\Delta\beta = 0.5$ to account for the fraction of neutral to charged energy in pile-up interaction. The relative muon isolation is defined as $I_{rel} = I/P_T^{\mu}$.

5.1.4 Electron

Electrons are reconstructed by matching tracks in the tracker with energy deposits in the ECAL. A standalone electron reconstruction algorithm[37] is used together with the PF algorithm described in Section 5.1.2. Electrons travel a significant amount of material that compose tracker before they reach the ECAL. A significant fraction of their energy is deposited through the emission of bremsstrahlung photons in ≈ 0.4 to 2 radiation lengths of

material depending on η. At η region where the material is thickest (e.g. at $|\eta| = 1.4$) electrons lose an average of 86% of their initial energy while on average 33% of the electron energy is lost at $|\eta| = 0$.

It is essential to collect the energy of all the radiated photons to measure the accurate initial electron energy. This is performed by different supercluster (SC) algorithms in the barrel and in the endcaps as the subdetector geometries are different. In the barrel, SC algorithm known as the"hybrid" starts from the seed crystal that contains the largest energy deposit above a minimum threshold of 1 GeV. Stripts of 5×1 crystals in the $\eta \times \phi$ plane are delimited around the seed crystal, and are merged to contiguous strips if their energy is greater than 0.1 GeV. A SC is then formed from the strip clusters with their energy exceeding 0.35 GeV. In the endcaps, the "multi 5×5" method begins with seed crystals with threshold of 0.18 GeV and the energy is collected around the seeds in secondary clusters of 5×5 crystals. The SC is built from the primary cluster and all secondary clusters with energy deposit grater than 1 GeV, within a η range of ± 0.07 and a ϕ range of ± 0.3 rad. The energy weighted positions of all secondary clusters are extrapolated to the Preshower, and the most energetic cluster is used as a reference point. The Preshower cluster ϕ range is defined by the maximum ϕ distance between the reference point and the secondary clusters extended by ± 0.15 rad with the η range set to ± 0.15 rad. The energy collected in the Preshower is added to the SC.

In order to reconstruct electron tracks, the standard CTF algorithm is adopted and compromised due to the large radiative losses in the tracker material. These losses motivate the use of a dedicated tracking procedure for electrons because they lead to a reduced hit-collection efficiency as well as a poor estimate of the trajectory parameters. The procedure starts by finding seeds of two or three hits in the tracker using two complementary methods. The first method is known as ECAL-based seeding which uses the SC energy and position to build the trajectory in the first layer of the tracker. Only seeds that contain hits within windows around the extrapolated trajectory can be selected. Another method is called a tracker-based seeding, which begins with tracks reconstructed by the CTF algorithm. In this method, seeds from CTF track are selected if a direction of the CTF track is compatible

with the position of the closest PF cluster, and the ratio of the cluster energy to the track momentum is within a certain range. The selected seeds are then extrapolated and smoothed following the same procedure described in Section 5.1.1 with the Gaussian sum filter (GSF) instead of the KF.

In order to reject backgrounds, such as misidentified jets, electron from photon conversions, and electrons produced in heavy-flavor decays, MVA-based and cut-based electron identification discriminants are used in the analyses. Shower shape variables, track quality variables, the fraction of energy deposited in the HCAL, and many other variables are used to build the discriminants. A few working points on the MVA discriminant are decided using cuts on the MVA score[37] and two working points, with efficiencies of 90% and 80% for genuine electrons are used in the analyses described in this thesis.

Furthermore, isolation discriminants are utilized to reduce the backgrounds. The isolation variable is defined as

$$I = \sum P_T^{h^\pm} + max(0, P_T^{h^0} + P_T^{\gamma} - \rho A_{eff}) \tag{5.2}$$

within a cone size of $\Delta R = 0.3$, where the sums run over all charged hadrons from the primary vertex ($h^\pm$), neutral hadrons (h^0), and photons (γ). The pile-up is estimated by the last term, ρA_{eff}, where ρ is the average p_T density for particles defined within the area of jets with $p_T > 3$ GeV and $|\eta| < 2.5$, and A_{eff} is the effective area[37].

5.1.5 Jets

Once quarks and gluons are produced, they almost instantly fragment and hadronize resulting in a collimated sprays of energetic hadrons known as a jet[38]. It is essential to detect and combine all constituents of a jet in order to accurately measure the kinematics of the initial quark or gluon. This procedure is performed by jet clustering algorithms[38]. The start of the clustering is defining two distance parameters: d_{iB}, which is the distance

between object i and the beam, and d_{ij}, which is the distance between object i and j.

$$
\begin{aligned}
d_{ij} &= min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \\
d_{iB} &= p_{Ti}^{2p},
\end{aligned}
\tag{5.3}
$$

where the parameter $p = -1, 0, 1$ defines the behavior of the algorithm, ΔR_{ij} is the ΔR separation between object i and j. R is the radious parameter that determines the typical size of the jets. The clustering is formulated as follows:

1. For all objects, d_{ij} and d_{iB} are computed according to Equation 5.3 and the minimum distance is found.

2. If it is a d_{ij}, combine the objects i and j into a single new object and return to step 1.

3. If it is a d_{iB}, then declare object i to be a jet and remove it from the list of objects. Return to step 1.

4. The procedure is repeated until there is no remaining objects.

CMS uses the anti-k_T jet clustering algorithm[31] which corresponds to setting $p = -1$ in Equation 5.3. The anti-k_T algorithm tends to cluster jets around the hardest particles. The analyses described in this thesis use jets clustered with a radius of $R = 0.4$. Simulated jets are also clustered by using all stable generator-level particles but for neutrinos.

The effect of pile-up on jet energy and substructure is partially mitigated using the charged hadron subtraction (CHS) procedure[33,39], whereby jets are reconstructed from all reconstructed particles except charged hadrons associated with pile-up vertices. Noise jets can occur from calorimeter and/or electronic noise and a loose identification discriminant is used to suppress those misidentified jets. In particular, for data collected in 2017, a number of noise jets are arose due to large amount of ECAL noise in the $2.65 < |\eta| < 3.14$ region. The noise jets have a similar geometrical structure to pile-up jets, with both being more dispersed compared to genuine jets. To reject the noise jets in this η region, they are required to pass a pile-up jet identification discriminant.

Jet Energy Corrections

The purpose of the jet energy corrections is to relate, on average, the measured energies for the detector jets to the energies of the corresponding true particle jets at hadron level. The correction is applied to each component of the uncorrected jet four-momentum, p_μ^{raw} as described by the equation below[40]:

$$\mathcal{C} = C_{\text{offset}}(p_T^{raw}) \cdot C_{\text{MC}}(p_T^{'}, \eta) \cdot C_{\text{rel}}(\eta) \cdot C_{\text{abs}}(p_T^{''}) \cdot p_\mu^{raw}, \tag{5.4}$$

where C_i are the correction factors, p_T^{raw} is the p_T of the uncorrected jet, $p_T^{'}$ is the p_T of the jet after applying the C_{offset} correction, and $p_T^{''}$ is the p_T of the jet after applying all the other corrections. C_{offset} is offset correction applied to remove the contribution of pile-up and electronics noise. The MC calibration factor, C_{MC}, corrects the reconstructed jet energy to equal, on average, the energy of the generator-level jets using simulation. The relative residual correction factor, C_{rel}, and the absolute jet energy response correction, C_{abs}, correct the response to be flat as a function of η and p_T, respectively.

5.1.6 b-jets

Jets originating from the hadronisation of b quark is called b-jet. Lifetime of the b hadron is relatively long ($\approx$ 1.5 ps), which results in flight distances of between 1-10 mm from its production points[32]. This means that the tracks of b-jet decay products are likely to be displaced from the primary vertex and secondary vertices are reconstructed often. b-jet identification algorithms usually rely on the presence of these displaced tracks and/or secondary vertices. The algorithms also may exploit variables which can be sensitive to other properties of b hadrons, such as hard fragmentations and their heavy mass.

An upgraded version of the combined secondary vertex (CSV) algorithm[41] is used for b-tagging in the analyses described in this thesis. The algorithm is known as CSVv2[32] and takes jets with $p_T > 20$ GeV and $|\eta| < 2.4$ as inputs. Secondary vertices are reconstructed using the inclusive vertex finding (IVF) algorithm, which contains all tracks with $p_T > 0.8$

GeV, impact parameters > 50 μm, and impact parameter significances > 1.2 in the vertex reconstruction. Information related to the displaced tracks and associated secondary vertices are combined together using a multivariate technique by the CVSv2 algorithm. To be used as discriminants, several working points are defined and *medium* working point is used in the analyses described here. The selected working point has an efficiency of $\approx 60\%$ for misidentification probability of $\approx 1\%$ for light-flavor jets[32] .

5.1.7 Missing transverse energy

Stable or long-lived weakly interacting particles travel through the CMS detector without any interactions, which include neutrinos and hypothetical BSM particles. Their presence cannot be directly observed, but instead is inferred by a momentum imbalance in the transverse plane defined as

$$\vec{E}_T^{miss} = -\sum \vec{p}_T^{\,i} \tag{5.5}$$

where the sum runs over all the reconstructed PF particles in a given event. The magnitude of $\vec{E}_T^{miss}$ is called the missing transverse energy, E_T^{miss}. The resolution of the E_T^{miss} gets influenced by tracker inefficiencies, minimum energy threshold in the calorimeter, and nonlinearity of the calorimeter response for hadronic particle[42]. The impact of these factors on the E_T^{miss} measurement is alleviated by propagating the jet energy corrections to $\vec{E}_T^{miss}$ as follow:

$$\vec{E}_T^{miss,corr} = \vec{E}_T^{miss,raw} - \sum_{jets} (\vec{p}_{T,jet}^{\,corr} - \vec{p}_{T,jet}^{\,raw}) \quad , \tag{5.6}$$

where *corr* refers to the corrected values and *raw* refer to uncorrected values.

Recoil corrections are applied to some of the simulated processes to enhance the agreement with data. The hadronic recoil $\vec{U}_T$ is

$$\vec{U}_T = \vec{E}_T^{miss,raw} - \sum \vec{p}_T^{\,\nu} \quad , \tag{5.7}$$

where the sum runs over all neutrinos, ν, in the event. $\vec{U}_T$ is measured in simulated $Z \to \mu\mu$

events using the following relationship:

$$\vec{U}_T = -\vec{H}_T - \vec{p}_T^{\,Z} \quad , \tag{5.8}$$

where $\vec{H}_T$ is the total transverse momentum of the recoiling jets. In data and simulation, the projection of $\vec{U}_T$ onto the perpendicular ($\vec{U}_\perp$) and parallel ($\vec{U}_\parallel$) planes to the Z boson $\vec{p}_T$ are fitted using a superposition of Gaussians. The following corrections are then applied to the simulation for each component.

$$\begin{aligned} U_\perp^{'} &= U_\perp \frac{\sigma_{data}(U_\perp)}{\sigma_{MC}(U_\perp)}, \\ U_\parallel^{'} &= < U_\parallel >_{data} + (U_\parallel - < U_\parallel >_{MC}) \frac{\sigma_{data}(U_\parallel)}{\sigma_{MC}(U_\parallel)} \end{aligned} \tag{5.9}$$

σ_i are resolutions of the fitted functions and the $\vec{E}_T^{miss}$ is recomputed using Equation 5.7.

5.1.8 Hadronic taus

τ leptons are unstable with a lifetime of 0.3 ps[43] and they decay close to the primary vertex before reaching the detector, thus, they cannot be directly detected. Taus are massive enough with a mass of 1.8 GeV and can decay either leptonically or hadronically. 64.8% of taus decay hadronically[43], where the tau decays to a composite of neutral and charged mesons together with a singe neutrino. The hadronic decay modes are summarised in Table 5.1. In this thesis, hadronically decaying taus will be denoted as τ_h. In the leptonic case, the tau decays to electron or muon accompanied by their partner, the electron neutrino and muon neutrino. The branching fractions for these leptonic decays are 17.8% and 17.4%, respectively[43].

Hadronic tau leptons are reconstructed using the dedicated algorithm called the hadrons plus strips (HPS) algorithm[44]. The HPS algorithm starts from seed jets with $p_T > 14$ GeV and $|\eta| < 2.5$. For a given jet, the algorithm generate all possible combinations of hadrons for five decay modes ($h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$, $h^\pm h^\mp h^\pm$ and $h^\pm h^\mp h^\pm \pi^0$) covering 96.8% of hadronic

Decay mode	Meson resonance	Branching fraction (%)
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	ρ (770)	26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	a_1 (1260)	9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	a_1 (1260)	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other hadronic decay		3.2
total		64.8

Table 5.1: *Summary of the most common hadronic tau decay modes and the corresponding branching fractions. The generic symbol $h^\pm$ represents either a $\pi^\pm$ or a $K^\pm$. Intermediate resonances are listed where relevant*[43].

tau decay as shown in Table 5.1. The jets are required to have at least one charged hadron originated from the primary vertex with $p_T > 0.5$ GeV. In the CMS detector, π^0 measons decay almost promptly to a pair of photons with a branching ratio of $\approx 99\%$ then convert to e^+e^- pairs in the tracker. The electron pairs are bent due to the magnetic field dispersing them largely in the ϕ direction. The e/γ constituents of the jets are, therefore, clustered into rectangular $\Delta\eta \times \Delta\phi$ strips. The strips are reconstructed as follows:

1. The e/γ with the highest p_T is selected as a seed for a new strip and set the initial η and ϕ values of the strip.

2. The e/γ with the second highest p_T within a $\eta \times \phi$ window centered on the strip is absorbed into the strip. The size of the strip window is defined

$$\begin{aligned} &\Delta\eta = f(p_T^{e/\gamma}) + f(p_T^{strip}) \text{ and} \\ &\Delta\phi = g(p_T^{e/\gamma}) + g(p_T^{strip}) \\ &\text{with } f(p_T) = 0.20 \cdot p_T^{-0.66} \text{ and } g(p_T) = 0.35 \cdot p_T^{-0.71}, \end{aligned} \tag{5.10}$$

   where $p_T^{e/\gamma}$ and p_T^{strip} are p_T of the candidate e/γ being merged and current strip respectively. The new η and ϕ values are computed using all e/γ constituents in the strip.

3. The above steps are repeated until there are no remaining e/γ candidates within the

$\Delta\eta \times \Delta\phi$ window. The process continues by choosing the next seed and the entire procedure is repeated until every e/γ constituent has been merged into strips.

Any strips which have $p_T > 2.5$ GeV are considered as π^0 candidates. Furthermore, the visible mass is required to be compatible with the intermediate meson resonances. The mass constraints are given as

- $h^{\pm}\pi^0$ mode: $0.3 - \Delta m_{\tau_h} < m_{\tau_h} < 1.3\sqrt{p_T/100\text{GeV}} + m_{\tau_h}$ GeV, the upper limit constrained to lie between 1.3 and 4.2 GeV,
- $h^{\pm}\pi^0\pi^0$ mode: $0.4 - \Delta m_{\tau_h} < m_{\tau_h} < 1.2\sqrt{p_T/100\text{GeV}} + m_{\tau_h}$ GeV, the upper limit constrained to lie between 1.2 and 4.0 GeV, and
- $h^{\pm}h^{\mp}h^{\pm}$ mode: $0.8 < m_{\tau_h} < 1.5$ GeV,

where Δm_{τ_h} is the change in the mass of the τ_h candidate because of the merging process of the e/γ to strips[44].

All τ_h constituent candidates are required to be bounded within the signal cone $0.05 < R_{sig} < 0.10$ where $R_{sig} = (3\text{ GeV})/p_T$. If there are additional charged particles, the candidates are discarded. τ_h candidate with the largest p_T is retained in case there are multiple τ_h candidates are found in the same jet. The $h^{\pm}\pi^0$ and $h^{\pm}\pi^0\pi^0$ decay modes are consolidated into the $h^{\pm}\pi^0$ decay mode to be analysed together. In this thesis, the four decay modes will be referred to as the 1-prong ($h^{\pm}$), 1-prong + 1-π^0 ($h^{\pm}\pi^0$), 3-prong ($h^{\pm}h^{\mp}h^{\pm}$) and 3-prong + 1-π^0($h^{\pm}h^{\mp}h^{\pm}\pi^0$) decay modes.

The jet $\rightarrow \tau_h$ fake background is significant even after the HPS reconstruction, and isolation requirement helps to reduce this contamination. The isolation of the τ_h candidates, I_{τ_h}, is given as

$$I_{\tau_h} = \sum p_T^{charged}(d_Z < 0.2cm) - max\left(0, \sum p_T^{\gamma} - \Delta\beta \sum p_T^{charged}(d_Z > 0.2cm)\right) . \quad (5.11)$$

I_{τ_h} is computed by summing the scalar p_T of the charged particles and photons within a cone with size $\Delta R = 0.5$ centered on the direction of the τ_h candidate. In this computation,

the constituents of the τ_h are excluded from the sum.

The DeepTau algorithm is developed to discriminate reconstructed hadronic decays of τ leptons that originated from genuine τ leptons against that originataed from quarks or gluon jets, electrons, or muons. The algorithm employs a deep neural network trained using a large number of features including all the inputs to the HPS algorithm, such as hits, tracks, as well as tau decay mode. It returns the probability, y_α, for the τ_h candidate to come from one of the four target classes (τ_h, jet, μ and e). The final discriminants against jets, muons, and electrons are then given by

$$\mathcal{D}_\alpha(y) = \frac{y_\tau}{y_\tau + y_\alpha} \qquad \text{with } \alpha \in \{jet, \mu, e\} \ . \tag{5.12}$$

The efficiencies for genuine generated τ_h with $p_T > 20$ GeV to be reconstructed with $p_T > 20$ GeV and pass the $\mathcal{D}_{jet}$ discriminant is shown in Figure 5.2 as a function of generated τ_h p_T. The reconstruction efficiency is over 80 % for $p_T > 30$ GeV and exceeds 90 % for $p_T > 100$ GeV. The efficiency is degraded by around 10 % when decay modes with missing charged hadrons (two-prong decay modes) are excluded. Eight working points for the $\mathcal{D}_{jet}$ discriminant are defined and *Medium* working point is used in the analyses described in this thesis, which targets efficiency of $\approx 70\%$.

5.2 Datasets and triggers

The datasets collected during 2016, 2017 and 2018 are used for the analyses described in this thesis. These datasets correspond to integrated luminosity of 138 fb^{-1}. Interesting events were collected by a set of triggers that designed to select events with tau pairs decaying to the $\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$ and $\tau_e\tau_\mu$ final states, which will be referred to as "channels" hereafter.

Double-tau triggers were used to select a pair of hadronically decaying taus ($\tau_h\tau_h$) events. These triggers require two isolated hadronically decaying taus with $p_T^{\tau_h} > 35$ or 40 GeV.

The $\tau_\mu\tau_h$ channel was selected using single-muon trigger and muon+tau cross trigger. The single-muon trigger requires an isolated muon with $p_T^\mu > 22$ for 2016 data taking period,

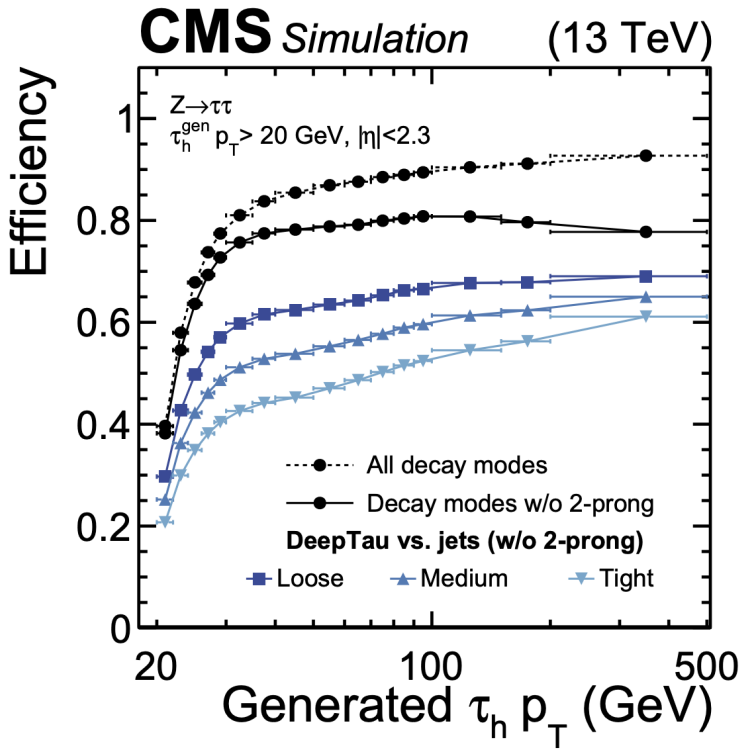

Figure 5.2: *Efficiencies for simulated τ_h decays with $p_T > 20$ GeV and $|\eta| < 2.3$: 1) to be reconstructed with $p_T > 20$ GeV and $|\eta| < 2.3$ (black dashed line) 2) to be reconstructed with $p_T > 20$ GeV and $|\eta| < 2.3$ without missing charged hadrons (black solid line), and 3) to be reconstructed and pass the Loose, Medium, or Tight working points of the $\mathcal{D}_{jet}$ discriminant (blue lines) without missing charged hadrons, obtained with a $Z \to \tau\tau$ event sample.*

whereas for 2017 and 2018 the threshold was either 24 or 27 GeV depending on the run period. In 2016, Muon+tau cross triggers require both an isolated muon and an isolated hadronic tau with $p_T^{\mu} > 19$ GeV and $p_T^{\tau_h} > 20$ GeV, while the p_T thresholds on the muon and tau candidates were increased to 20 GeV and 27 GeV respectively in 2017.

Similarly, single-electron and electron+tau triggers were used to target the $\tau_e\tau_h$ channel. The single-electron trigger requires an isolated electron candidate with a $p_T^e > 25$ GeV for 2016, $p_T^e > 27$, 32, and 35 GeV for 2017, and $p_T^e > 32$, and 35 GeV for 2018. Electron+tau cross triggers were not utilized for 2016 data taking period. For 2017 and 2018 data taking period, the electron+tau cross triggers require one isolated hadronic tau with $p_T^{\tau_h} > 30$ GeV and one isolated electron with $p_T^e > 24$ GeV.

Finally, electron+muon triggers target the $\tau_e\tau_\mu$ channel. Two such triggers were used

requiring the presence of one isolated muon and one isolated electron. Each places different p_T thresholds on the leptons; one requires $p_T^{\mu} > 8$ GeV and $p_T^{e} > 23$ GeV, and another one requires $p_T^{\mu} > 23$ GeV and $p_T^{e} > 12$ GeV.

All the triggers described above is commonly used to select di-tau events in the three analyses covered in this thesis.

5.3 Event simulation

This section gives a concise summary of tools used to simulate proton-proton collision events based on the more detailed review given in Ref.[45]. The generation of a simulated physics event normally beging with the hard subprocess resulting from a highly energetic collision of a pair of partons, such as gluons or a quark-antiquark pair from the incoming partons producing a few outgoing fundamental particles. The portion of momenta of the colliding protons carried by an incoming parton which are selected by sampling the parton distribution functions (PDFs) of the proton at the energy scale of the subprocess. For example, function $f_i(x, \mu_F)$ is PDFs which describe the probability of a parton species i to carry a momentum fraction x probed at a chosen factorisation scale μ_F.The hard subprocess, by definition, involves either large momentum transfer or heavy particle production, resulting in violently accelerated partons that are asymptotically free. Thus, these processes can be described by matrix-elements using perturbation theory. The cancellation of ultraviolet and infrared divergences are involved in the matrix-element computations, which results in a dependence on μ_F and the renormalisation scale μ_R. Depending on the level of precision required, the hard subprocess may be treated at leading order (LO) or next-to-leading order (NLO). Furthermore, progress on next-to-next-to-leading order (NNLO) generation also has been made[46].

As if accelerated electric charges emit QED radiation (photons), the accelerated colored partons will emit QCD radiation (gluons) which lead parton showers. The most dominant contributions are associated with the emission of many additional soft or collinear partons. At the lowest scale hadronization occurs, where the partons are bound into colorless hadrons

interacting non-perturbatively. Due to the non-perturbative nauture, the models requires tunable parameters to describe the data. These tools also provide a description of the UE which is secondary interaction of proton constituents not involved in the hard subprocess. The input parameters of these models are tuned using collision data. Unstable particles, such as taus, heavy-flavor measons have to be decayed into a set of stable final-state particles. Finally, the detector response to these final-state particles is also simulated.

5.3.1 Monte Carlo samples

Backgrounds

The Monte Carlo (MC) simulation samples used to model background processes are identical across all analyses described in this thesis. MC samples were produced for each year separately although all datasets were collected at the same center of mass energy. This means that there shouldn't be much change to the true event topologies, however, the sample vary due to different detector conditions, such as different detector responses or different numbers of pile-up interaction. In addition, the PDFs and UE tunes were updated in 2017 and 2018 which was reflected in the MC production. Therefore, the generator-level event topologies also can mildly vary between eras.

Depending on the background process, different MC generators are used. The QCD-induced W+jets and Z+jets processes are modelled using MadGraph5[47]. They are simulated at leading order (LO) with MLM jet matching and merging[48] to allow for large statistics compare to using the NLO sample with negative weights. A correction of the Z p_T spectrum in the LO simulation is performed to describe the data correctly. The single-top quark production are simulated at NLO accuracy using powheg 2.0 and 1.0 respectively, whereas the MadGraph5_aMC@NLO generator is used for diboson production including ZZ, WZ, and WW at NLO accuracy with FxFx jet merging scheme[49]. The generators are interfaced with Pythia8.212[50] to model the parton showering, hadronization, as well as the decay of the τ leptons. The Pythia parameters affecting the description of the underlying events are set to the CUETP8M1 tune[51] in 2016, and CP5 tune[52] in 2017 and 2018.

Signal

POWHEG V2 generator is used to obtain descriptions of the SM Higgs boson production for the five main production modes: 1) ggH including quark mass effects, VBF, and associated production (WH, ZH and ttH). In case of WH and ZH production modes, the MINLO HVJ extension of POWHEG is utilized.

For the analysis described in Chapter 9, the analysis-specific signal samples are also used. The JHUGEN 7.0.2[53],[14] MC program is used to simulate anomalous couplings in H boson production and in $H \to \tau\tau$ decay. The kinematics of ggH events with two associated jets are modified by anomalous Hgg couplings, which are simulated for $H + 0, 1, 2$ jets separately using MADGRAPH5_AMC@NLO. Associated production is simulated using JHUGEN at LO QCD. In the VBF and VH production modes, the JHUGEN and POWHEG simulations are compared after parton showering in the SM case, and no significant differences are found in kinematic observables. Therefore, the JHUGEN simulation is adopted to describe kinematics in the VBF and VH production modes with anomalous couplings, with expected yields taken from the POWHEG simulation.

The NNPDF 3.0 (3.1)[54] Parton Distribution Functions (PDF) are used in 2016 (2017 and 2018), and all simulated samples are interfaced with PYTHIA with the CUETP8M1 (CP5) tune to describe parton showering and hadronization in 2016 (2017 and 2018). Pile-up interactions are generated by PYTHIA and overlaid on all simulated events, according to the luminosity profile of the analyzed data.

All the generated signal and background MC samples are processed with the simulation of the CMS detector based on GEANT4[55] to model detector effects observed in data.

5.3.2 Embedded samples

The description of background with genuine di-τ pairs can be improved using an embedding technique[32]. This method involves selecting $Z \to \mu\mu$ events in data and the energy deposits of the recorded muons are replaced by the energy deposits of simulated τ leptons decays with the same kinematic properties. As illustrated in Figure 5.3, the procedure of

the production of the embedded samples can be summarized in four steps:

1. Selection: Select di-muon events in data. The selection criteria are intentionally maintained as loose as possible to minimize any biases.

2. Cleaning: Calorimeter deposits and tracks associated with the selected di-muon pair are removed from the event.

3. Simulation: A di-tau system is simulated using the kinematics of muon Lorentz vector. The decay of the τ leptons are modeled using PYTHIA in the same way as a typical $Z \to \tau\tau$ decay.

4. Merging: All hits and energy deposits from the simulated tau pair are extrapolated to the cleaned event.

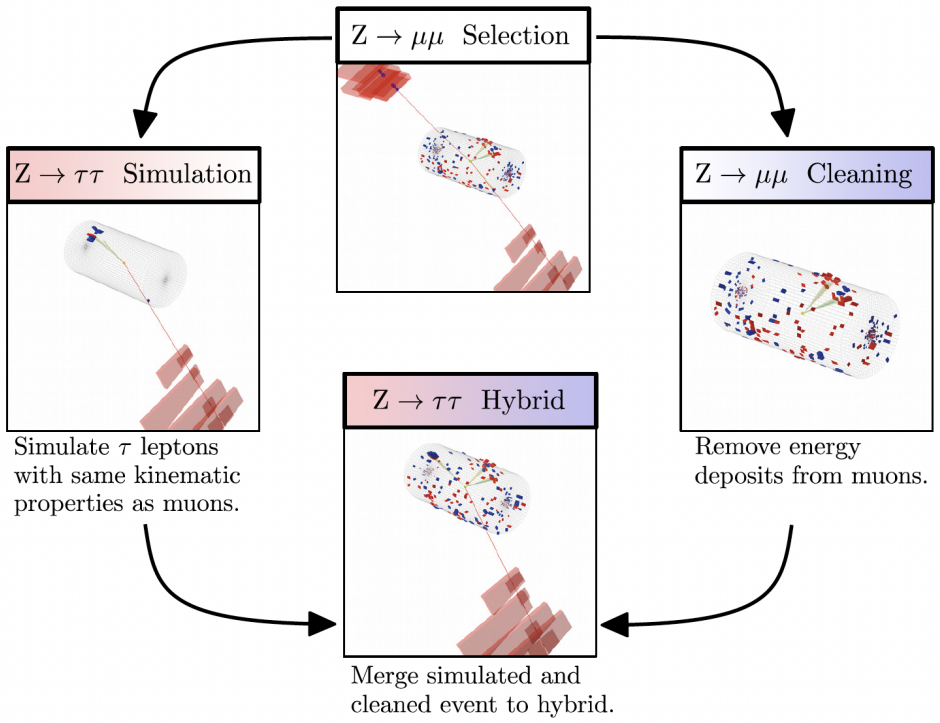

Figure 5.3: *Illustration of the four main steps of the τ-embedding technique: selection, cleaning, simulation, and merging*[32].

Z boson decays is predominant in the selected di-muons ($\approx 99\%$), and remaining 1 % is mainly from $t\bar{t}$ and di-boson events. Therefore, the embedding technique models not only $Z \to \tau\tau$ decay, but also provides an estimate for entire backgrounds producing a genuine di-τ pair. The primary benefit of utilizing the embedded sample is that the majority of the event information is derived directly from data, with the exception of the well described decay of τ lepton. This is often challenging to model jet topology, pile-up and underlying event using MC simulation. By construction, the embedded samples model these properties with extremely high accuracy. Additionally, description of the Z boson kinematics can also be challenging with MC generators, which is remedied by the embedded method, as $Z \to \tau\tau$ and $Z \to \mu\mu$ kinematics are near identical. Moreover, the statistics of the embedded samples increase linearly with the amount of collected collision data, which allows to avoid to generate the computationally intensive simulation of large MC samples.

5.4 Simulation corrections

5.4.1 E_T^{miss} recoil corrections

The E_T^{miss} recoil corrections are applied to correct for the mismodeling in the simulated samples of Drell-Yan Z/γ^*, W+jets and Higgs production. Figure 5.4 shows the improvement in the data/MC agreement for 2018 $Z \to \mu\mu$ events.

5.4.2 Pile-up reweighting

The MC events are reweighted so that the number of pile-up interactions match the pile-up conditions found in data. The number of interactions is estimated from the measured instantaneous luminosity per bunch crossing.

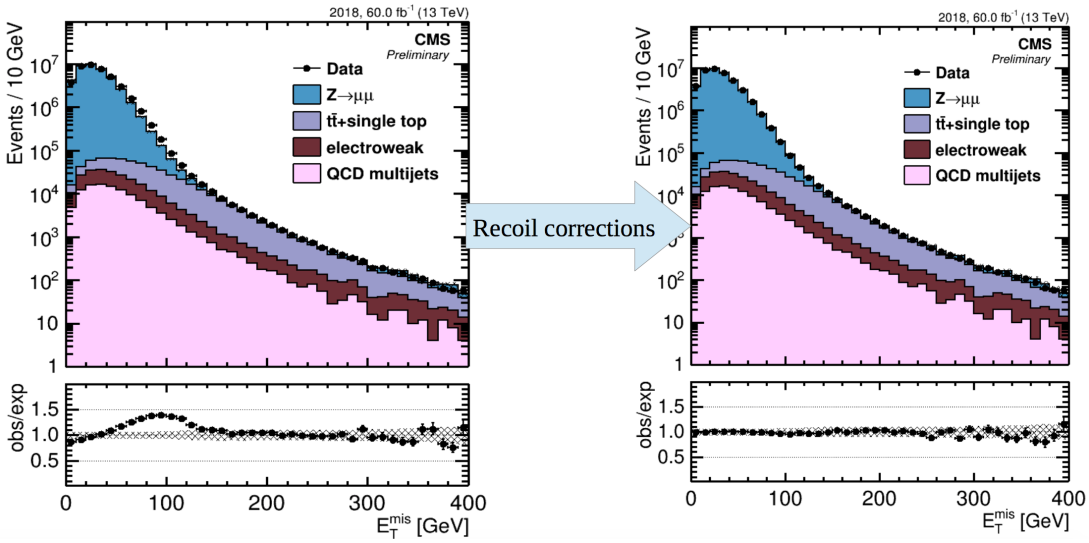

Figure 5.4: *p_T^{miss} distribution before (left) and after (right) recoil correction in 2018 $Z \to \mu\mu$ events.*

5.4.3 Top p_T reweighting

The agreement of the $t\bar{t}$ MC with the data is improved from the NLO generated MC to NNLO by reweighting the top p_T distribution. This reweighting is computed as a function of the (anti)top quark p_T for each quark independently, and then the overall reweighting is the square root of the weights for each top quark in the $t\bar{t}$ pair multiplied together. $t\bar{t}$ simulations use different tunes in 2016 (CUTETP8PM2T4) from 2017+2018 (CP5), therefore, corrections depends on their tune. The weights are first derived for CP5 simulation as Equation. 5.13 and directly applied to 2017 and 2018 samples.

$$w = exp(0.088 - 0.00087 \times p_T + 0.00000092 \times p_T^2) \tag{5.13}$$

In 2016, the same weights derived from CP5 simulation applied after reweighting top p_T distribution in CUTETP8PM2T4 to that in CP5.

5.4.4 Z/$\gamma^* p_T$-mass reweighting

The Drell-Yan MC samples are generated by MadGraph5 at LO accuracy. A reweighting is applied to the samples in order to correct the generator level p_T and m_{ll} distributions to match what is observed in data. The weights are derived in a pure $Z \to \mu\bar{\mu}$ control region. The weights change the shape of the p_T and m_{ll} distributions, but then the distributions are normalized to ensure that they do not change the yield. The weights are measured separately for 2016 and other two years because Drell-Yan samples generated in 2016 use a different tune than those generated in 2017 and 2018.

5.4.5 Electron and muon efficiency corrections

Measurement of the tracking, identification, isolation, and trigger efficiencies may differ between data and MC. To account for the differences, scale factors are derived with a tag-and-probe method using samples of $Z \to \mu\mu$ and $Z \to ee$ events. These efficiencies are measured by the CMS HTT group, and are applied to MC samples. Dedicated corrections are measured and applied to the embedded samples.

5.4.6 Hadronic tau efficiency corrections

Genuine τ_h identification efficiency can differ in data and MC. Corrections are measured in an inclusive $\tau_\mu\tau_h$ channel, using genuine Drell-Yan sample. Dedicated corrections are measured in the same way for the embedded samples.

5.4.7 e$\to \tau_h$ and $\mu \to \tau_h$ fake rate corrections

The efficiency of the discriminators against electrons or muons misidentified as τ_h candidates can be mismodeled in simulation. To correct the mismodeling, the rate at which electrons and muons pass the anti-electron and anti-muon discriminants are measured with a tag-and-probe method using sample of Z/$\gamma^* \to$ ee and Z/$\gamma^* \to \mu\mu$ events. Scale factors are applied to correct for the observed differences in the fake rates in data and simulation.

5.4.8 Tau energy scale correction

Differences in τ_h energy scale are commonly observed between MC and data, which can occur due to, for example, mismodeling of the τ_h and detector effects. Corrections must be applied to the τ_h four-vector to account for this. The τ_h energy scale is measured using $Z/\gamma^* \rightarrow \tau_\mu\tau_h$ events for each tau decay mode separately. It is measured in two regions of phase space based on τ_h p_T and the corrections from the low p_T region are used because the measured energy scales are consistent within one standard deviation. The energy scale of each τ_h that can be matched to a generator-level τ is varied in steps of 0.2 % to construct templates. A maximum likelihood fit is used to extract the most likely τ_h energy scale.

5.4.9 e$\rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ energy scale corrections

The energy scale corrections are also measured for taus originated from lepton fakes. Corrections are derived for e$\rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ candidates independently. To extract the correction, the $\tau_{e,\mu}\tau_h$ invariant mass distribution is fitted because the energy scale moves the position of the $Z \rightarrow ee$ or $Z \rightarrow \mu\mu$ mass peaks. Corrections are measured for 1-prong and 1-prong+π^0 separately. The misidentification probability of the 3-prong decay mode is negligible, therefore, it is not considered.

5.5 Background methods

5.5.1 Modelling backgrounds using MC

The generated MC events are corrected using the procedures described in Section 5.4 in order to model background processes. The events are then scaled to their expected theoretical cross sections using the weight given as

$$w = \frac{L_{int} \cdot \sigma \cdot BR \cdot w_{gen}}{N_{effective}} \quad , \tag{5.14}$$

where σ is cross section and BR is branching ratio for the process, L_{int} is the integrated luminosity, w_{gen} is the generator weight, and $N_{effective}$ is the effective event number ($N_{effective} = \sum_i w^i_{gen}$ where i runs over all events).

5.5.2 Modelling $Z/\gamma^* \to \tau\tau$ using embedding

It is crucial to have a good description of $Z/\gamma^* \to \tau\tau$ process because it is the most dominant background for $H \to \tau\tau$ analyses. The embedded samples described in Section 5.3.2 are utilized to enhance the modelling of background with genuine di-τ pairs, including $Z/\gamma^* \to \tau\tau$ process. The corrections described in Section 5.4 are applied to the embedded events, together with the weights described in Section 5.5.1. Embedded events are further corrected for the inefficiencies in the initial di-muon reconstruction by scale factor given as

$$w = \frac{1}{\epsilon_{ID} \cdot \epsilon_{trig}} \quad , \tag{5.15}$$

where ϵ_{ID} is the efficiency for the muons to pass the ID requirements, and ϵ_{trig} is the double-muon trigger efficiency. The agreement of the embedded $Z/\gamma^* \to \tau\tau$ with the data is checked in every final state considered. Relatively high purity of $Z/\gamma^* \to \tau\tau$ can be obtained using invariant mass of two tau candidates in spite of a non-negligible contribution from other background processes which is shown in Figure 5.5 with decent agreement with the data in both the yields and the shapes is observed.

5.5.3 The fake-factor method

It is difficult to model backgrounds in which a jet fakes a τ_h using MC because of the poor description of the jet $\to \tau_h$ fake rate in simulation. In addition, a significant computational expense is required to MC samples as the probability of a jet being misidentified as a τ_h is not huge. These motivate the use of data-driven methods for these processes, and the fake-factor (FF) method is one of such procedure. The general idea behind the FF method is that a control region enriched in jet $\to \tau_h$ fake backgrounds can be defined using side-band

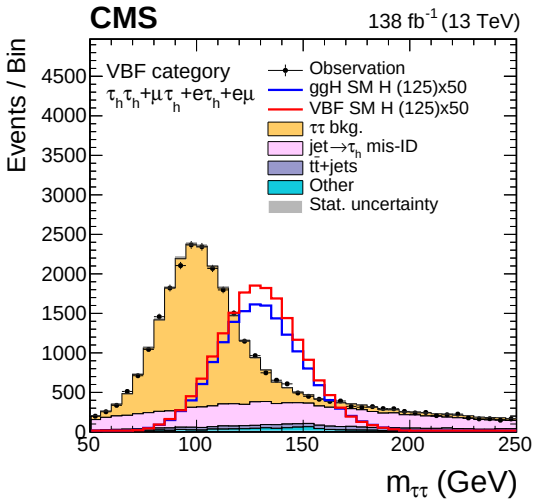

Figure 5.5: *The invariant mass of Higgs boson candidate, $m_{\tau\tau}$, in VBF topology. All events selected in the $\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$ and $\tau_e\tau_\mu$ channels are included. Only statistical uncertainties are shown.*

regions in data, in which a τ_h fails the nominal *Medium* DeepTau ID still passing the the loosest working point (veryveryvery-loose or VVVLoose).

$$FF_i = \frac{N(Medium)}{N(VVVLoose \;\&\&\; !Medium)} \tag{5.16}$$

where $N(x)$ is the number of events meeting the identification requirement x. The distributions in this side-band regions can then be scaled by the ratio of isolated to non-isolated events (called the fake factor) to give a distributions in the signal region. The FF method is based on the assumption that the side-band region and the signal region have similar kinematics. The FF, therefore, provide the correct normalization for those shapes in the signal region with the assumption. Additional corrections are also measured and applied in order to account for small, understood differences in shape between two regions.

The FF method describes any background, which has at least one jet faking a τ_h candidate in the $\tau_h\tau_h$, $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels. Dominant processes estimated by this method are QCD multi-jet, W+jets, and $t\bar{t}$ events with fully-hadronic or semi-leptonic decays. The fake factors are separately measured for each channel, data taking period and processes. For

each process, a dedicated side-band region is determined to be enriched with events from the target process. The FF will not agree between channels due to the differences in trigger and event selection. The trigger thresholds are slightly changed in each year leading to separated FF measurements. It is expected that the FF to depend on the physics process and jet multiplicity. The fake rates typically decreases by 10 % when the jet multiplicity is increased by one.

The FF method is composed of the following steps:

1. The FF is measued as a function of τ_h p_T because the τ_h identification working points are derived to have a flat efficiency with respect to p_T (Figure 5.7). Requiring the loosest working point for every event deduces the p_T dependence drastically, and the remaining dependence can be parameterized.

2. Measured FF are then corrected as a function of p_T of the other leg (Second energetic τ_h for the $\tau_h\tau_h$ channel, and leptonic decay tau for the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels). The FF is further corrected to account for slightly different kinematics in each side-band regions and signal region (Figure 5.7).

3. Determine the fraction of QCD multi-jet, W+jets, and $t\bar{t}$ events in the signal region as a function of the visible mass of Higgs boson candidate, m_{vis}.

4. Scale corrected FF by the corresponding fractions and average of the three FF in order to obtain jet $\rightarrow \tau_h$ fake backgrounds in the signal region as follow

$$FF_{tot} = \sum_{i=1}^{3} n_i \cdot FF_i \tag{5.17}$$

   where i runs over three background processes, FF_i is the FF estimate for a given background side-band region and n_i is the fraction of background i.

5. Estimated jet $\rightarrow$ τ_h fake backgrounds can contain contamination from other background processes with real τ_h that are not isolated. The same procedure is applied

to all background processes with real τ_h to account for this contamination, then subtracted from data.

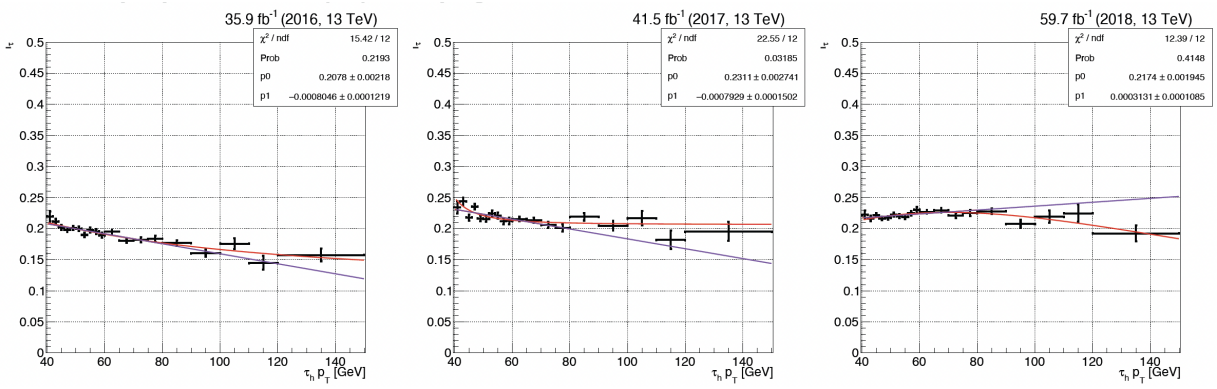

Figure 5.6: *Fake factors determined in the QCD multijet sideband region with 0jet in the $\tau_h\tau_h$ final state in 2016 (left), 2017 (middle) and 2018 (right). They are fitted with the sum of a Landau function and polynomial as a function of the leading τ_h p_T.*

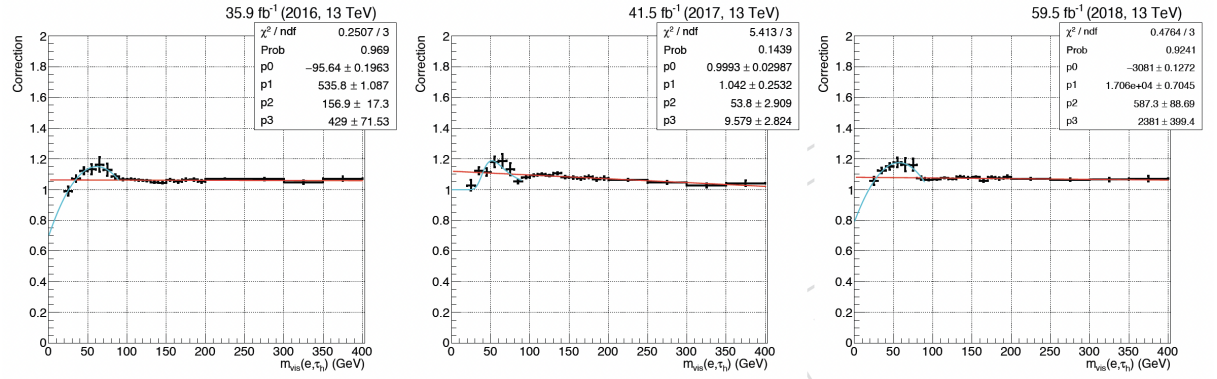

Figure 5.7: *Correction for the inversion of the sign requirement in the $\tau_h\tau_h$ final state of the QCD multijet sideband region in 2016 (left), 2017 (middle) and 2018 (right).*

It is observed that jet $\rightarrow$ τ_h fake backgrounds estimated by data-driven method have better agreement with the data than what could be estimated using MC simulation. This method had been developed by the HTT group in the CMS collaboration and already used in several analyses.

5.6 SVFit for Di-tau mass

Invariant mass of the tau pair mass, $m_{\tau\tau}$, is an unique property in decays of neutral particles into pairs of tau leptons. There are several SM processes as well as scenarios for physics beyond the SM that containing tau pairs. The sensitivity of those analyses can be enhanced significantly by distinguishing the source of the tau pair using $m_{\tau\tau}$. However, the presence of neutrinos in the τ lepton decay prevent us from reconstructing full $m_{\tau\tau}$ as they are invisible. The CMS detector can only provide the energy imbalance in the plane transverse to the colliding proton beams, while at least one neutrino will decay from each tau lepton. The imbalance, therefore, will also include any other particles escaping detection for any reason, detection effects such as noise, and unaccounted physics processes. Dedicated algorithm, *SVfit*[56], for reconstruction of $m_{\tau\tau}$ has been developed and analyses covered by this thesis uses a simplified version of the *SVfit* called *FastMTT*.

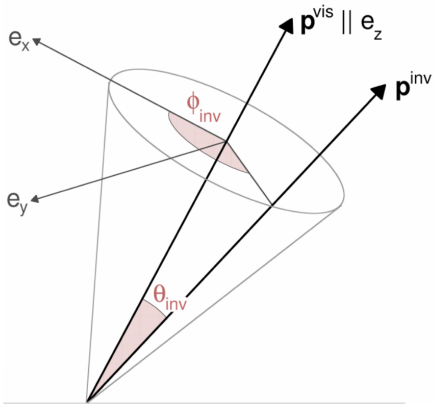

Figure 5.8: *Parametrization of tau lepton decay in tau rest frame.*

For a single tau decay, the kinematic of hadronic decay is parameterized by two decay angles θ_{inv} and ϕ_{inv} as shown in Fig. 5.8. In addition to these angles, leptonic tau decay has one more parameter which is the invariant mass of the neutrino system, $m_{\nu\nu}$, in order to fully describe the neutrino system. Returning to the case of the tau pair, it is clear to see there are four to six unknown parameters depending on how the taus decay. For example, if taus decay hadronically there are four unknowns, $\vec{a} = (\theta_1, \phi_1, \theta_2, \phi_2)$, on the other hand if

one tau decays to electron or muon there are five unknowns, $\vec{a} = (\theta_1, \phi_1, m_{\nu\nu,1}, \theta_2, \phi_2)$. These unknown variables are constrained by two observables which constrain the momenta of the neutrinos produced in the tau decays.

$$\begin{aligned} \Sigma p_x^{\nu} &= E_x^{miss} \\ \Sigma p_y^{\nu} &= E_y^{miss} \end{aligned} \tag{5.18}$$

The problem of reconstructing $m_{\tau\tau}$ is underconstrained by measured observables and handled via a likelihood approach in the *SVfit* algorithm. In that, the algorithm determine the most likely $m_{\tau\tau}$ by maximizing the likelihood given in Equation. 5.18 where $\vec{y}$ is the four-momenta of the visible decay products. The model makes a prediction for the probability density $p(\vec{x}|\vec{y},\vec{a})$ to observe the values $\vec{x} = (E_x^{miss}, E_y^{miss})$ measured in an event, given $\vec{a}$ and $\vec{y}$.

$$\mathcal{L}(m_{\tau\tau}(\vec{y},\vec{a})|\text{event data}) = \int p(\vec{x}|\vec{y},\vec{a})\delta(m_{\tau\tau} - m_{\tau\tau}(\vec{y},\vec{a}))d\vec{a} \tag{5.19}$$

Practically, the probability density $p(\vec{x}|\vec{y},\vec{a})$ can be decomposed to a product of three independent terms: 1) the matrix element for tau decays (ME), 2) the transfer function (TF), and 3) a regularization term (REG).

$$p(\vec{x}|\vec{y},\vec{a}) = ME \cdot TF \cdot REG \tag{5.20}$$

where the transfer function is

$$\begin{aligned} TF_{MET} &= \mathcal{L}(\vec{E}_T^{miss,reco}|\vec{E}_T^{miss,hypo}) \\ &= \frac{1}{2\pi\sqrt{|V|}} exp[-\frac{1}{2}((\vec{E}_T^{miss,reco} - \vec{E}_T^{miss,hypo})V^{-v}(\vec{E}_T^{miss,reco} - \vec{E}_T^{miss,hypo}))] \end{aligned} \tag{5.21}$$

However, a number of calculation should be performed for each event in order to evaluate systematic uncertainties which can be very costly in terms of time and CPU. For this reason,

the *FastMTT* algorithm was developed based on *SVfit* to attempt to simplify the calculation of Equation. 5.18 while maintaining similar performance.

The major difference between two algorithms is the ME term which is simply set to 1 in *FastMTT*. It is noted that the ME term has relatively small impact compared to the TF and the REG terms. In addition, for sufficiently boosted taus, the angle θ becomes small enough to be assumed the tau direction and neutrino direction are identical, called the collinear approximation. In the collinear approximation, angles θ and ϕ are no longer well defined. As a results, we can simplify Equation. 5.19 to Equation. 5.22.

$$\mathcal{L}(m_{\tau\tau}(\vec{y},\vec{a})|\text{event data}) = TF \cdot \int \delta(m_{\tau\tau} - m_{\tau\tau}(\vec{y},\vec{a}))d\vec{a} \tag{5.22}$$

FastMTT lead to an estimate of $m_{\tau\tau}$ which is similar in performance to *SVfit* while running almost 100 times faster with these simplifications. The $m_{\tau\tau}$ distribution reconstructed using the *FastMTT* algorithm is shown in Figure 5.5.

Chapter 6

Systematic uncertainties

The following systematic uncertainties affecting the shapes and normalizations of the signal and background templates are commonly included in every analysis described in this thesis. In this section, all the uncertainties that apply to MC samples are described first, while uncertainties in embedded samples are described after. Some analyses specific uncertainties will be detailed in relevant chapters.

6.1 Uncertainties in the object reconstruction

6.1.1 τ_h identification

For τ_h candidates matched to τ_h at generated level, the ID scale factors are computed in 4 decay mode bins for the $\tau_h\tau_h$ channel where the decay mode distribution is affected by the trigger selection, and in 3 p_T bins for the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels. A shape uncertainty is considered for each of the decay mode or p_T bins. The size of these uncertainties range between 2 and 3% and they are uncorrelated between bins and years. The uncertainties are fully correlated between the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels because two channels use the same τ_h identification working point.

In addition, a normalization uncertainty of 3% is considered for real τ_h related to the discriminators against electrons and muons per each hadronic tau leg. This uncertainty is

uncorrelated between channels because different electron and muon discriminators working points are used in each channel.

In summary, there are a total of 4 τ_h ID uncertainties (3 shape + 1 normalizations) for the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels, and 5 ID uncertainties (4 shape + 1 normalizations) for the $\tau_h\tau_h$ channel.

6.1.2 τ_h energy scale

The energy scale of real τ_h is corrected by decay mode. Corresponding uncertainty depends on the decay mode and on the year and are uncorrelated between them. The uncertainties ranges between 0.7 and 1.2% and are treated as shape uncertainties as they affect the mass and p_T distributions.

6.1.3 $e \to \tau_h$ misidentification

In the $\tau_e\tau_h$ channel, the shape uncertainties applied depending on the CMS ECAL geometry (barrel and endcap regions) for τ_h candidates matched to electrons at generated level. The uncertainty is dominated by statistical sources and uncorrelated across years and the measured bins.

On top of that, 10% uncertainty per p_T range (30-40, 40-50, 50+ GeV) is taken into account because there is no evidence that there is no dependence with the τ_h p_T although the correction is measured inclusively.

6.1.4 $e \to \tau_h$ energy scale

For τ_h candidates matched to electrons at generated level is corrected in the 1 prong and 1 prong+π^0 decay modes in the $\tau_e\tau_h$ and $\tau_h\tau_h$ channels. The shape uncertainties is considered depending on τ_h decay modes and the pseudorapidity of the τ_h. They are uncorrelated across decay modes, years, and the pseudorapidity bins.

6.1.5 $\mu \to \tau_h$ misidentification

In the $\tau_\mu\tau_h$ channel, the shape uncertainties applied depending on the CMS ECAL geometry (barrel and endcap regions) for τ_h candidates matched to muon at generated level. The uncertainty is dominated by statistical sources and uncorrelated across years and the measured bins.

On top of that, 10% uncertainty per p_T range (30-40, 40-50, 50+ GeV) is taken into account because there is no evidence that there is no dependence with the τ_h p_T although the correction is measured inclusively.

6.1.6 $\mu \to \tau_h$ energy scale

For τ_h candidates matched to muons at generated level in the $\tau_\mu\tau_h$ channel, a normalization uncertainty of 1% is applied and they are uncorrelated in τ_h decay modes.

6.1.7 Muon identification

The normalization uncertainty in the muon reconstruction, efficiency, and tracking is 2% for all years. It is uncorrelated between years but fully correlated between channels which have muon.

6.1.8 Muon energy scale

The uncertainty in the muon energy scale depends on the muon pseudorapidity, and the exact values are given in Table 6.1. The uncertainties are uncorrelated between pseudorapidity bin, while the uncertainties are fully correlated by year. These uncertainties are treated as shape uncertainties because they modify the muon p_T, $m_{\tau\tau}$, D_ζ, and m_T distributions.

$\lvert\eta\rvert$ bin	[0.,1.2]	[1.2,2.1]	[2.1,2.4]
Uncertainty	0.4%	0.9%	2.7%

Table 6.1: *Uncertainty in the muon energy scale*

6.1.9 Electron identification

The normalization uncertainty in the electron reconstruction, efficiency, and tracking is 2 % for all years. It is uncorrelated between years but fully correlated between channels which have electron.

6.1.10 Electron energy scale

The energy scale of electrons is corrected event-by-event, therefore, scale uncertainties are also derived event-by-event. They can affect and are treated as shape uncertainties because they affect the electron p_T, $m_{\tau\tau}$, D_ζ, and m_T distributions. The uncertainties are fully correlated between years as well as between channels.

6.1.11 Trigger

The uncertainty in the trigger efficiency is p_T and decay mode-dependent for the τ_h leg and ranges between 3 and 15 % per leg. For the electron and muon leg, a uncertainty of 2 % per leg is applied.

6.1.12 Jet energy scale

The jet energy scale is corrected event-by-event with the jet energy corrections. There are 27 sources of uncertainties, which are merged into different groups based on the correlations between them which is closely related to the detector geometry. The uncertainties are propagated to the p_T^{miss} too on an event-by-event basis and affect distributions of several variables, such as m_{jj} or $m_{\tau\tau}$ through the p_T^{miss} variation.

6.1.13 Jet energy resolution

In simulation, jets are smeared in order to match the jet energy resolution in data. Uncertainties in the smearing are considered as an additional shape uncertainty. These uncertainties are uncorrelated between years.

6.1.14 p_T^{miss} reconstruction

The uncertainty in the recoil corrections should be applied for events where the p_T^{miss} distribution is corrected with recoil corrections. The corrections and uncertainties are split into resolution and scale components. These are treated as shape uncertainties because they affect the $m_{\tau\tau}$, m_T, Higgs p_T, and D_ζ distributions. The uncertainties are statistically dominated and ,therefore, uncorrelated between years.

For events where no recoil corrections are applied (single-top, di-boson, $t\bar{t}$ events), the uncertainty in the unclustered p_T^{miss} is considered as a shape uncertainty. This uncertainty is treated as uncorrelated between years.

6.2 Uncertainties in the the background estimation methods

6.2.1 $jet \to \tau_h$ fake background

Jet $\to \tau_h$ fake backgrounds are estimated by data-driven method, called the fake factor method, as described in Section 5.5.3. There are multiple sources of uncertainty related to this method.

In the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels, the FF are parameterized with a linear function as a function of the τ_h p_T and two uncertainties per fit function are taken into account. The magnitude of the correction is considered as the uncertainty, which means that the up and down shapes correspond to applying the correction twice or not at all. Lastly, the m_T correction for the W fake factor is parameterized with a linear function. The two uncorrelated uncertainties in the linear function are taken into account as systematic uncertainties.

The FF are parameterized with the sum of a Landau function and a linear polynomial in the $\tau_h\tau_h$ channel as shown in Figure 5.6. The uncertainty is taken as using a linear function up to 80 GeV and then a constant fake rate above. On top of that, a 5 % normalization uncertainty for jet multiplicity bin is applied to cover for normalization differences seen in

the closure tests.

6.3 Uncertainties in the background cross sections

The uncertainty in the $t\bar{t}$, di-boson, single-top, and Drell-Yan cross sections are 4.2 %, 5 %, 5 %, 2 %, respectively.

6.4 Theoretical uncertainties for the signal

The uncertainties in the signal production cross section and $H \to \tau\tau$ branching fraction were provided by LHC Higgs Cross Section Working[9]. The uncertainties in the cross section include a 0.5 % (0.9 %, 0.8 %) uncertainty on VBF (ZH, WH) production due to variations of the QCD scales and 3.2 % (2.1 %, 1.3 %, 1.8 %) on ggH (VBF, ZH, WH) due to PDF+α_S variations. The uncertainty on the $H \to \tau\tau$ branching ratio includes a 1.7 % uncertainty because of missing higher order corrections, a 0.62 % parametric uncertainty on α_S, and a 0.99 % parametric uncertainty on the quark masses. The 3.9 % on ggH production that would come from variations of the QCD scales are split into nine shape uncertainties provided by LHC Higgs Working Group, two of which correspond to scale change and seven others to category migrations. For the ggH production, uncertainties due to the description of the underlying event description and the parton shower (UEPS) is taken into account.

6.4.1 Uncertainties related to the embedded samples

The uncertainties in the τ_h energy scale and ID efficiency have the same magnitude as in the MC, and are 50 % correlated with the MC uncertainties. In addition, a normalization uncertainty of 5 % per tau leg is considered for real τ_h related to the discriminators against electrons and muons. This uncertainty is uncorrelated between channels because different working points are used in each channel.

Embedded samples estimate all events with two τ candidates, mainly Drell-Yan events

but also some di-boson and $t\bar{t}$ events. An uncertainty related to the contamination from these non Drell-Yan events is treated as a shape uncertainty. The contamination from $t\bar{t}$ and di-boson events is estimated from MC samples, and 10 % is added or subtracted from the embedded shape. The uncertainty is correlated between channels but uncorrelated between years.

An uncertainty in the muon trigger used to select data events before replacing the muons by taus is treated as a normalization uncertainty of 4%.

6.4.2 Other uncertainties

The uncertainty in the luminosity affects all processes estimated from simulation, which is decomposed in multiple components. This uncertainty is in the range 1.2–2.5%, with partial correlations between data-taking years[57 58 59].

Bin-by-bin uncertainties are used to take into account the insufficient statistical precision in every bin of the final distributions.

An uncertainty in the simulated effect of the prefiring of the L1 trigger is considered for all simulations. It is correlated between years (2016 and 2017 only), and between channels.

Chapter 7

Measurement of the Higgs boson cross section in STXS framework

This chapter describe measurement of production cross section of the Higgs boson decaying to the $\tau\tau$ final state. The analysis targets the gluon-gluon fusion and the vector boson fusion productions. The signal strengths are measured following the simplified template cross section (STXS) binning scheme (stage 0, and stage 1.2)[9 60]. The measurement is performed using the dataset collected by the CMS during full Run-2 (2016, 2017, and 2018), which corresponds to a total integrated luminosity of 138 fb^{-1}.

7.1 Introduction to the STXS framework

Lately, Higgs boson production rates have been studied in the framework of the STXS scheme, developed by the LHC Higgs Working Group. STXS scheme defines a set of kinematic and topological phase space regions to account for the evolving experimental sensitivity (henceforth referred to as STXS bins). The measurements are unfolded to the STXS bins which are defined identically for all analyses. Thus, it allows the global combination of measurements across different Higgs boson decay channels as well as across experiments. The STXS bins have been chosen in order to reduce the dependence on any underlying theoreti-

cal models folded into the measurements. The number of STXS bins has evolved with time, such that more and more fine-grained measurement is possible as the anticipated statistical power of the data required to perform the measurements increases. For this purpose, they have been defined in stages. At STXS stage 0, events are split according to the Higgs boson production mechanisms, of which ggH, and qqH are of relevance for this thesis. ggH includes the gluon-initiated VH associated production with hadronic vector boson decays as well as ggF production. Likewise, qqH includes both VBF production and the quark-initiated VH associated production with hadronic vector boson decays. At stage 1.2 these production modes have division of the basic categories into finer bins according to the jet multiplicity (N_{jets}), the invariant mass of the two leading jets (m_{jj}), and the Higgs boson transverse momentum (p_T^H).

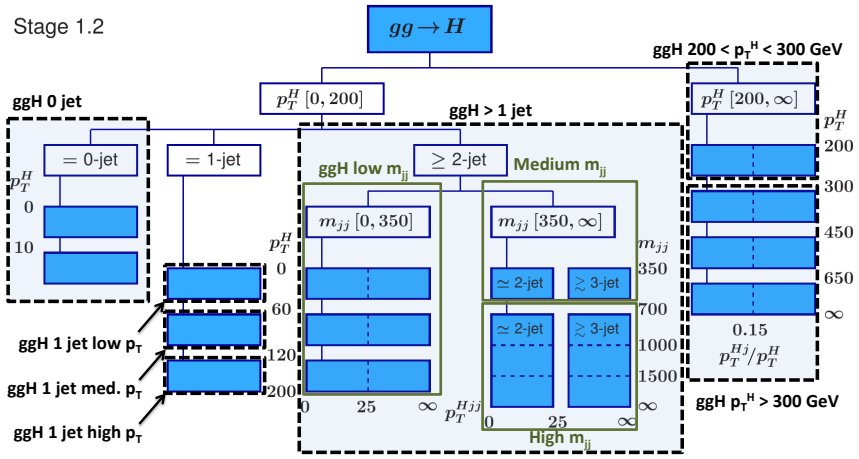

Figure 7.1: *Stage 1.2 bins for gluon-gluon fusion Higgs production gg→H. The dashed black boxes show the process-based merging explained in Section 7.6.2. The green boxes indicate the differences in merging for the topology-based merging detailed in the same section.*[3]

Figs. 7.1 and 7.2 show stage 1.2 STXS bins. The analyses under consideration do not have the statistical power to measure all stage 1.2 bins, therefore, they have been combined resulting in 7 measurements for ggH, and 4 measurements for qqH.

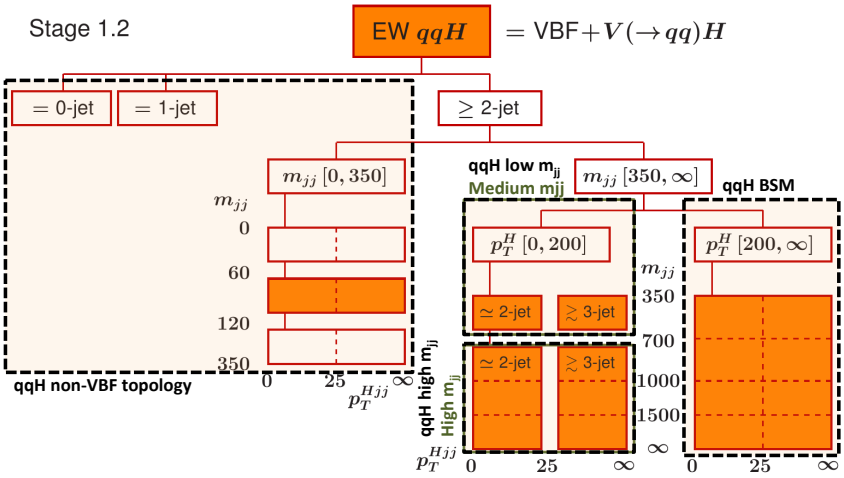

Figure 7.2: *Stage 1.2 bins for electroweak qqH production, VBF+V(→qq)H. The dashed black boxes show the process-based merging explained in Section 7.6.2. The green boxes indicate the differences in merging for the topology-based merging detailed in the same section.*[3]

7.2 Event selection

Selected events are classified into the various final states according to the numer of selected electrons, muons, and τ_h cndidates. The analysis covers four different final states. The di-tau from the Higgs boson can decay to $\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$, $\tau_e\tau_\mu$, $\tau_e\tau_e$ and $\tau_\mu\tau_\mu$. Two di-tau final states, $\tau_e\tau_e$ and $\tau_\mu\tau_\mu$ are not studied because they overlap with the $H \to ZZ$ analysis[61] and have anyway an extremely low branching fraction compare to the rest four final states. Events with additional loosely identified and isolated muons or electrons are discarded to make the resulting event samples in different final states mutually exclusive. The lepton pair assigned to the Higgs boson decay are required to have opposite-sign electric charge.

In the $\tau_h\tau_h$ channel, events are selected with triggers that require two isolated τ_h objects. In the $\mu\tau_h(e\tau_h)$ channel, the trigger system requires at least one isolated muon(electron) candidate, or at least one isolated muon(electron) and one τ_h candidate, depending on the offline muon p_T. Whereas in the $e\mu$ channel, the triggers rely on the presence of both an electron and a muon, allowing lower online p_T thresholds. The selection criteria are summarized in Table 7.1.

Channel	Trigger requirement	Year	Selection criteria p_T (GeV)	η	Isolation
$\tau_h\tau_h$	$\tau_h(35)$ & $\tau_h(35)$	2016	$p_T^{\tau_h} > 40$	$\lvert\eta^{\tau_h}\rvert < 2.1$	DNN τ_h ID
	$\tau_h(40)$ & $\tau_h(40)$	2017, 2018			
$\mu\tau_h$	$\mu(22)$	2016	$p_T^{\mu} > p_T^{\text{trigger}} + 1$ GeV	$\lvert\eta^{\mu}\rvert < 2.1$	$I^{\mu} < 0.15$
	$\mu(19)$ & $\tau_h(21)$	2016	$p_T^{\tau_h} > 30$	$\lvert\eta^{\tau_h}\rvert < 2.3$	DNN τ_h ID
	$\mu(24)$	2017, 2018			
	$\mu(20)$ & $\tau_h(27)$	2017, 2018			
$e\tau_h$	$e(25)$	2016	$p_T^{e} > p_T^{trigger} + 1$ GeV	$\lvert\eta^{e}\rvert < 2.1$	$I^{e} < 0.15$
	$e(27)$	2017	$p_T^{\tau_h} > 30$	$\lvert\eta^{\tau_h}\rvert < 2.3$	DNN τ_h ID
	$e(32)$	2018			
	$e(24)$ & $\tau_h(30)$	2017, 2018			
$e\mu$	$e(12)$ & $\mu(23)$	all years	$p_T^{e} > 15$, $p_T^{\mu} > 24$	$\lvert\eta^{e}\rvert < 2.4$	$I^{e} < 0.15$
	$e(23)$ & $\mu(8)$	all years	$p_T^{\mu} > 15$, $p_T^{e} > 24$	$\lvert\eta^{\mu}\rvert < 2.4$	$I^{\mu} < 0.15$

Table 7.1: *Kinematic selection requirements for the four di-tau decay channels. The trigger requirement is defined by a combination of trigger candidates with p_T over a given threshold, indicated inside parentheses in GeV. The pseudorapidity thresholds come from trigger and object reconstruction constraints.*[3]

In the $\mu\tau_h$ and $e\tau_h$ channels, the large W + jets background is suppressed by requiring the transverse mass m_T smaller than 50 GeV. The m_T is defined as follow:

$$m_T \equiv \sqrt{2 p_T^l p_T^{miss} [1 - cos(\Delta\phi)]}\,, \tag{7.1}$$

where l is either μ or e in the $\mu\tau_h$ and $e\tau_h$ channels respectively, and $\Delta\phi$ is the azimuthal angle between its direction and the $\vec{p}_T^{miss}$.

In the $e\mu$ channel, m_T is also required to be smaller than 60 GeV, which defined by the transverse momentum of $e\mu$ system, $p_T^l = p_T^{e\mu}$, to reduce overlap with $H \to WW$ analysis. In addition, the $t\bar{t}$ background is reduced by requiring $p_\zeta > -30$ GeV, where p_ζ is the component of the $\vec{p}_t^{miss}$ along the bisector of the transverse momentum of the two leptons.

Events with a b-tagged jet are discarded to further suppress the $t\bar{t}$ background in the $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels.

7.3 Event categorization

Events are split into categories designed to maximize the sensitivity to the signal by isolating regions with large signal-to-background ratios, and to provide sensitivity to the stage 0 as well as stage 1.2 ggH and qqH STXS bins. The categories are chosen to be exclusive to avoid double counting and the combination of all categories covers entire phase space that passed event selection. Event categories aim to separate the different Higgs boson production modes, corresponding to the stage 0 processes in the STXS scheme. A 0-jet category contains all events with no reconstructed jet. This category has very large statistics predominantly containing background. This category is mostly used to constrain background normalization in the fit to data. Some signal events, which are mainly originated from ggH production, are also contained in the 0-jet category. Events with at least two jets are categorized to target VBF production with additional selection. In the $\tau_h\tau_h$ final state, the pseudorapidity between two jets ($\Delta\eta_{jj}$) is required to be greater than 2.5 and p_T^H is required to be greater than 100 GeV, while other channels m_{jj} must be greater than 350 GeV. It is split into two categories by requiring an estimate of p_T^H ($\hat{p}_T^H$) is larger or smaller than 200 GeV, where $\hat{p}_T^H$ is obtained from the sum of the $\vec{p}_T$ of the two τ candidates and $\vec{p}_T^{miss}$. All other events that do not fit within the 0-jet or two VBF categories, enter to so-called *boosted* categories. These categories distinguished by the presence of exactly one or at least two jets in an event. The boosted categories can contain contributions from VBF events that did not pass the VBF category selection, but they mainly contain ggH events with the Higgs boson recoiling against one or several jets. Also, VH events where V decays hadronically are categorized as boosted events. This lead to five categories for each $\tau\tau$ final state, and summarized in Table 7.2.

7.4 Systematic uncertainties

All systematic uncertainties discussed in Chapter 6 are considered in this analysis.

Final state	Category	Selection	Observables
$\tau_h\tau_h$	0-jet	0 jet	$m_{\tau\tau}$
	VBF low p_T^H	≥ 2 jets, $\Delta\eta_{jj} > 2.5$(2.0 for 2016), $100 < \hat{p}_T^H < 200$ GeV	$m_{\tau\tau}$, m_{jj}
	VBF high p_T^H	≥ 2 jets, $\Delta\eta_{jj} > 2.5$(2.0 for 2016), $\hat{p}_T^H > 200$ GeV	$m_{\tau\tau}$, m_{jj}
	Boosted 1 jet	1 jet	$m_{\tau\tau}$, $\hat{p}_T^H$
	Boosted ≥ 2 jets	Not in VBF, ≥ 2 jets	$m_{\tau\tau}$, $\hat{p}_T^H$
$l\tau_h$, $e\mu$	0-jet	0 jet	$m_{\tau\tau}$, $p_T^{\tau_h}$ ($l\tau_h$) $m_{\tau\tau}$ ($e\mu$)
	VBF low p_T^H	≥ 2 jets, $m_{jj} > 350$ GeV, $\hat{p}_T^H < 200$ GeV	$m_{\tau\tau}$, m_{jj}
	VBF high p_T^H	≥ 2 jets, $m_{jj} > 350$ GeV, $\hat{p}_T^H > 200$ GeV	$m_{\tau\tau}$, m_{jj}
	Boosted 1 jet	1 jet	$m_{\tau\tau}$, $\hat{p}_T^H$
	Boosted ≥ 2 jets	Not in VBF, ≥ 2 jets	$m_{\tau\tau}$, $\hat{p}_T^H$

Table 7.2: *Event categorization in STXS measurements. The stage 0 and 1.2 measurements are extracted by performing a maximum likelihood fit of one and two-dimensional distributions in these categories using the observables listed in the last column.*[3]

7.5 Signal extraction and statistical interpretation

The extraction of the results used a global maximum likelihood fit based on 2D distributions in all channels. The form of binned likelihood is

$$\mathcal{L}\left(\{k_i\},\,\{\mu_s\},\,\{\theta_j\}\right) = \prod_i \mathcal{P}\big(k_i|\sum_s \mu_s\, S_s(\{\theta_j\}) + \sum_b B_b(\{\theta_j\})\big)\prod_j \mathcal{C}(\hat{\theta}_j|\theta_j) \tag{7.2}$$

where i labels all bins of the input distributions for each signal process, with index s, and background process, with index b. Signal and background templates are obtained from the data model as discussed in Section 5.5. The Poisson density to observe k_i events in bin i for a prediction of $\sum \mu_s\, S_s$ signal and $\sum B_b$ background events in that given bin i is written as $\mathcal{P}(k_i|\sum \mu_s\, S_s(\{\theta_j\}) + \sum B_b(\{\theta_j\}))$ in Equation (7.2). μ_s are the parameters of interest (POIs) which is the scaling parameters of the signal contributions S_s with respect to the SM expectation.

Four $\tau\tau$ final states have five event categories as explained in Section 7.3. Therefore, 60 input distributions provided for the signal extraction in total. The invariant mass of two

taus candidates, $m_{\tau\tau}$, is chosen to be the primary discriminating observable in every channel and category, and another variable is chosen depending on category with the exception of the 0-jet categories in the $\tau_h\tau_h$ and $e\mu$ channels. An overview of the discriminationg observables is summarized in Table 7.2. Example distributions in the $\tau_h\tau_h$ final state are shown in Figure 7.3, and Figure 7.4.

Practically, systematic uncertainties are integrated in the form of penalty terms in the likelihood for additional nuisance parameters $\{\theta_i\}$. In that, predefined probability density function $\mathcal{C}(\hat{\theta}_j|\theta_j)$ is multiplied to the likelihood. The predefined uncertainties in the $\hat{\theta}_j$ may be constrained by the fit to the data.

7.6 Results

Maximum likelihood estimates for the POIs are obtained from all 60 distributions with an inclusive floating signal strength for all $H \to \tau\tau$ events. The best-fit signal strength is $\mu = 0.85^{+0.12}_{-0.11}$. All distributions are combined into a single 1D $m_{\tau\tau}$ distribution, shown in Figure 7.5. It is obtained by summing all $m_{\tau\tau}$ distributions of each category/slice of the secondary variable with a weight proportional to the signal purity. The weight is the ratio between the signal and background yields in bins with $90 < m_{\tau\tau} < 150$ GeV in each $m_{\tau\tau}$ slice. The distributions are renormalized to keep the same total signal yield before and after reweighting.

7.6.1 Inclusive and Stage 0 best-fit signal strengths

The best-fit signal strengths are shown in Table 7.3 for the qqH and ggH production modes (stage 0), as well as for the inclusive $H \to \tau\tau$ signal. The results are presented individually per year with all final state combined, and per final state with all years combined. In the last column, the results are presented for the combination of all final states and years. The ggZH process with hadronic Z decay is floated with the ggH parameters. The qqVH processes with hadronic V decays are floated with the qqH parameters. The VH processes with leptonic

V decays have a low acceptance due to the extra lepton vetos and are constrained to the SM expectation. The inclusive and stage 0 results are summarized in Figure 7.6 with their decomposed uncertainties.

	2016	2017	2018	$e\mu$	$e\tau_h$	$\mu\tau_h$	$\tau_h\tau_h$	Combined
μ	$1.13^{+0.23}_{-0.21}$	$0.64^{+0.20}_{-0.19}$	$0.93^{+0.20}_{-0.15}$	$1.29^{+0.51}_{-0.47}$	$0.99^{+0.27}_{-0.24}$	$0.95^{+0.17}_{-0.16}$	$0.76^{+0.19}_{-0.17}$	$0.85^{+0.12}_{-0.11}$
μ_{ggH}	$0.83^{+0.39}_{-0.36}$	$0.72^{+0.35}_{-0.34}$	$1.40^{+0.33}_{-0.29}$	$2.47^{+0.91}_{-0.84}$	$0.42^{+0.53}_{-0.51}$	$0.99^{+0.27}_{-0.25}$	$1.29^{+0.44}_{-0.37}$	$0.98^{+0.20}_{-0.19}$
μ_{qqH}	$1.54^{+0.502}_{-0.47}$	$0.51^{+0.48}_{-0.46}$	$0.36^{+0.30}_{-0.29}$	$-0.17^{+0.98}_{-0.95}$	$1.41^{+0.49}_{-0.46}$	$0.89^{+0.38}_{-0.37}$	$0.09^{+0.39}_{-0.38}$	$0.67^{+0.23}_{-0.22}$

Table 7.3: *Observed inclusive and stage 0 signal strengths per year with all final states combined, per final state with all years combined, and with all final states and years combined.*[3]

7.6.2 Stage 1 best-fit signal strengths

Signal strengths are measured for individual or combinations of stage 1.2 processes. This analysis is not sensitive to all of STXS bins, and therefore stage 1.2 bins are merged together when the per-parameter sensitivity is poor. The bins are also merged when there are large correlations between multiple processes due to reconstruction resolution effects. Two merging schemes are chosen and 11 signal strengths in each scheme fitted simultaneously. The first scheme (process-based merging) treats the ggH and qqH productions separately, and following processes are fitted simultaneously.

1. qqH non-VBF topology: qqH events with less than 2 jets or $m_{jj} < 350$ GeV;
2. qqH medium m_{jj}: qqH events with $p_T^H < 200$ GeV, at least 2 jets and $350 < m_{jj} < 700$ GeV;
3. qqH high m_{jj}: qqH events with $p_T^H < 200$ GeV, at least 2 jets and $m_{jj} > 700$ GeV;
4. qqH BSM: qqH events with $p_T^H > 200$ GeV;
5. ggH > 1 jet: ggH events with at least 2 jets and $p_T^H < 200$ GeV;
6. ggH $200 < p_T^H < 300$ GeV: ggH events with $200 < p_T^H < 300$ GeV;
7. ggH $p_T^H > 300$ GeV: ggH events with $p_T^H > 300$ GeV;

8. ggH 0 jet: ggH events with $p_T^H < 200$ GeV and 0 jet;

9. ggH 1 jet low p_T: ggH events with 1 jet and $p_T^H < 60$ GeV;

10. ggH 1 jet medium p_T: ggH events with 1 jet and $60 < p_T^H < 120$ GeV;

11. ggH 1 jet high p_T: ggH events with 1 jet and $120 < p_T^H < 200$ GeV.

In the second scheme (topology-based merging), the ggH production in VBF-like topology, consisting of two jets and a large invariant mass (> 350 GeV) is combined to the qqH production, resulting in the following 11 processes.

1. qqH non-VBF topology: qqH events with less than 2 jets or $m_{jj} < 350$ GeV;

2. Medium m_{jj}: qqH and ggH events with $p_T^H < 200$ GeV, at least 2 jets and $350 < m_{jj} < 700$ GeV;

3. High m_{jj}: qqH and ggH events with $p_T^H < 200$ GeV, at least 2 jets and $m_{jj} > 700$ GeV;

4. qqH BSM: qqH events with $p_T^H > 200$ GeV;

5. ggH low m_{jj}: ggH events with at least 2 jets with $m_{jj} < 350$ GeV and $p_T^H < 200$ GeV;

6. ggH $200 < p_T^H < 300$ GeV: ggH events with $200 < p_T^H < 300$ GeV;

7. ggH $p_T^H > 300$ GeV: ggH events with $p_T^H > 300$ GeV;

8. ggH 0 jet: ggH events with $p_T^H < 200$ GeV and 0 jet;

9. ggH 1 jet low p_T: ggH events with 1 jet and $p_T^H < 60$ GeV;

10. ggH 1 jet medium p_T: ggH events with 1 jet and $60 < p_T^H < 120$ GeV;

11. ggH 1 jet high p_T: ggH events with 1 jet and $120 < p_T^H < 200$ GeV.

The stage 1.2 STXS subdivisions are illustrated in Figures 7.1, and 7.2. The correlation between signal strengths of the STXS parameters in the process-based stage 1.2 merged scheme are shown in Figure 7.7.

The resulting best-fit signal strengths for these processes, as determined from a simultaneous binned maximum likelihood fit, are shown in Figure 7.8 for the two merging schemes. The results show a particularly good sensitivity in VBF-like topologies, such as Higgs bosons produced with high p_T or produced in association with jets. Stage 1.2 signal strength measurements are dominated by statistical uncertainty.

7.6.3 Cross section measurements

The product of the production cross section and branching ratio to a pair of τ leptons is extracted in a similar way as the signal strengths for the stage 0 and stage 1.2 Higgs boson production. For these results, theoretical uncertainties in the cross section and branching ratio predictions are excluded. The measured cross sections are shown in Table 7.4 for the inclusive and stage 0 processes, and Tables 7.5 and 7.6 for the stage 1.2 processes in two merging schemes. A summary of the production cross section σ and branching fraction $\mathcal{B}(H \to \tau\tau)$ for the inclusive, stage 0, and stage 1.2 STXS processes is given in Figure 7.9.

	$\sigma\mathcal{B}(H \to \tau\tau)$ (fb)	
Process	Measured	SM Prediction
Inclusive	2960^{+394}_{-370}	3422^{+172}_{-172}
ggH	3060^{+592}_{-552}	3051^{+160}_{-160}
qqH	$221^{+75.3}_{-73.3}$	$329^{+9.67}_{-9.67}$

Table 7.4: *Product of the production cross section σ and branching fraction $\mathcal{B}(H \to \tau\tau)$ measured for the inclusive and stage 0 processes.*[3]

A scan of the negative log-likelihood difference is performed for $m_H = 125.09$ GeV in the (κ_V, κ_f) parameter space, where κ_V and κ_f quantify the ratio between the measured and the SM value for the couplings of the Higgs boson to vector bosons and fermions, respectively[62]. A likelihood scan is also performed as a function of the signal strengths for the ggF and VBF productions. In this scan, the VH production is constrained to the SM expectation. Decays

	$\sigma\mathcal{B}(H \to \tau\tau)$ (fb)	
Process	Measured	SM Prediction
qqH non-VBF topology	374^{+552}_{-515}	$209^{+6.11}_{-6.11}$
qqH medium m_{jj}	$-18.9^{+45.8}_{-46.0}$	$34.4^{+0.909}_{-0.909}$
qqH high m_{jj}	$32.9^{+17.7}_{-17.3}$	$47.5^{+1.80}_{-1.80}$
qqH BSM	$6.41^{+4.40}_{-4.17}$	$9.90^{+0.339}_{-0.339}$
ggH >1 jet	19.0^{+251}_{-271}	$306^{+70.3}_{-70.3}$
ggH $200 < p_T^H < 300$ GeV	$24.2^{+24.3}_{-24.5}$	$27.5^{+11.5}_{-11.5}$
ggH $p_T^H > 300$ GeV	$13.0^{+7.82}_{-7.90}$	$7.19^{+3.36}_{-3.36}$
ggH 0 jet	-964^{+864}_{-792}	1753^{+160}_{-160}
ggH 1 jet low p_T	-756^{+519}_{-510}	$451^{+63.2}_{-63.2}$
ggH 1 jet med p_T	1010^{+264}_{-261}	$288^{+40.9}_{-40.9}$
ggH 1 jet high p_T	$99.1^{+51.9}_{-47.1}$	$50.0^{+11.5}_{-11.5}$

Table 7.5: *Product of the production cross section σ and branching fraction $\mathcal{B}(H \to \tau\tau)$ measured for the stage 1.2 processes in the process-based merging scheme.*[3]

	$\sigma\mathcal{B}(H \to \tau\tau)$ (fb)	
Process	Measured	SM Prediction
qqH non-VBF topology	191^{+541}_{-510}	$209^{+6.11}_{-6.11}$
qqH BSM	$5.92^{+4.38}_{-4.13}$	$9.90^{+0.339}_{-0.339}$
ggH >1 jet $m_{jj}[0, 350]$	112^{+209}_{-220}	$254^{+58.3}_{-58.3}$
ggH $200 < p_T^H < 300$ GeV	$30.3^{+24.0}_{-24.8}$	$27.5^{+11.5}_{-11.5}$
ggH $p_T^H > 300$ GeV	$15.1^{+7.66}_{-7.87}$	$7.19^{+3.36}_{-3.36}$
ggH 0 jet	117^{+866}_{-791}	1753^{+160}_{-160}
ggH 1 jet low p_T	-695^{+517}_{-510}	$451^{+63.2}_{-63.2}$
ggH 1 jet med p_T	1040^{+261}_{-262}	$288^{+40.9}_{-40.9}$
ggH 1 jet high p_T	$108^{+52.2}_{-47.2}$	$50.0^{+11.5}_{-11.5}$
combined VBF topology $m_{jj}[350, 700]$	$-3.47^{+40.1}_{-40.1}$	$70.8^{+11.9}_{-11.9}$
combined VBF topology $m_{jj} \geq 700$	$38.4^{+21.6}_{-21.5}$	$63.0^{+7.52}_{-7.52}$

Table 7.6: *Product of the production cross section σ and branching fraction $\mathcal{B}(H \to \tau\tau)$ measured for the stage 1.2 processes in the topology-based merging scheme.*[3]

of the Higgs boson to other particles than a pair of τ leptons are treated as a background in both scans. All nuisance parameters are profiled for each of the scan. The observed best-fit values are consistent within one standard deviation with the prediction of the SM as shown in Figure 7.10.

7.7 Chapter summary

A measurement of the $H \rightarrow \tau\tau$ signal strength, using data recorded in proton-proton collisions by the CMS experiment at center-of-mass energy of 13 TeV and correspond to an integrated luminosity of 137 fb^{-1}, has been presented. The event are categorized to increase the signal sensitivity, to separate the gluon fusion and vector boson fusion productions, and to provide sensitivity to the simplified template cross section framework, particularly at high Higgs boson p_T and in event topologies with jets. The results are extracted via maximum likelihood fits in one- or two-dimensional distributions in the form of signal strengths relative to the standard model expectations and products of cross sections and branching ratio to τ leptons in 11 mutually exclusive kinematic regions in two different schemes, following the simplified template cross section scheme of the LHC Higgs Working Group. All results are compatible with the standard model prediction. The best-fit of Higgs boson production signal strength is found to be $\mu = 0.85^{+0.12}_{-0.11}$ for inclusive Higgs boson production.

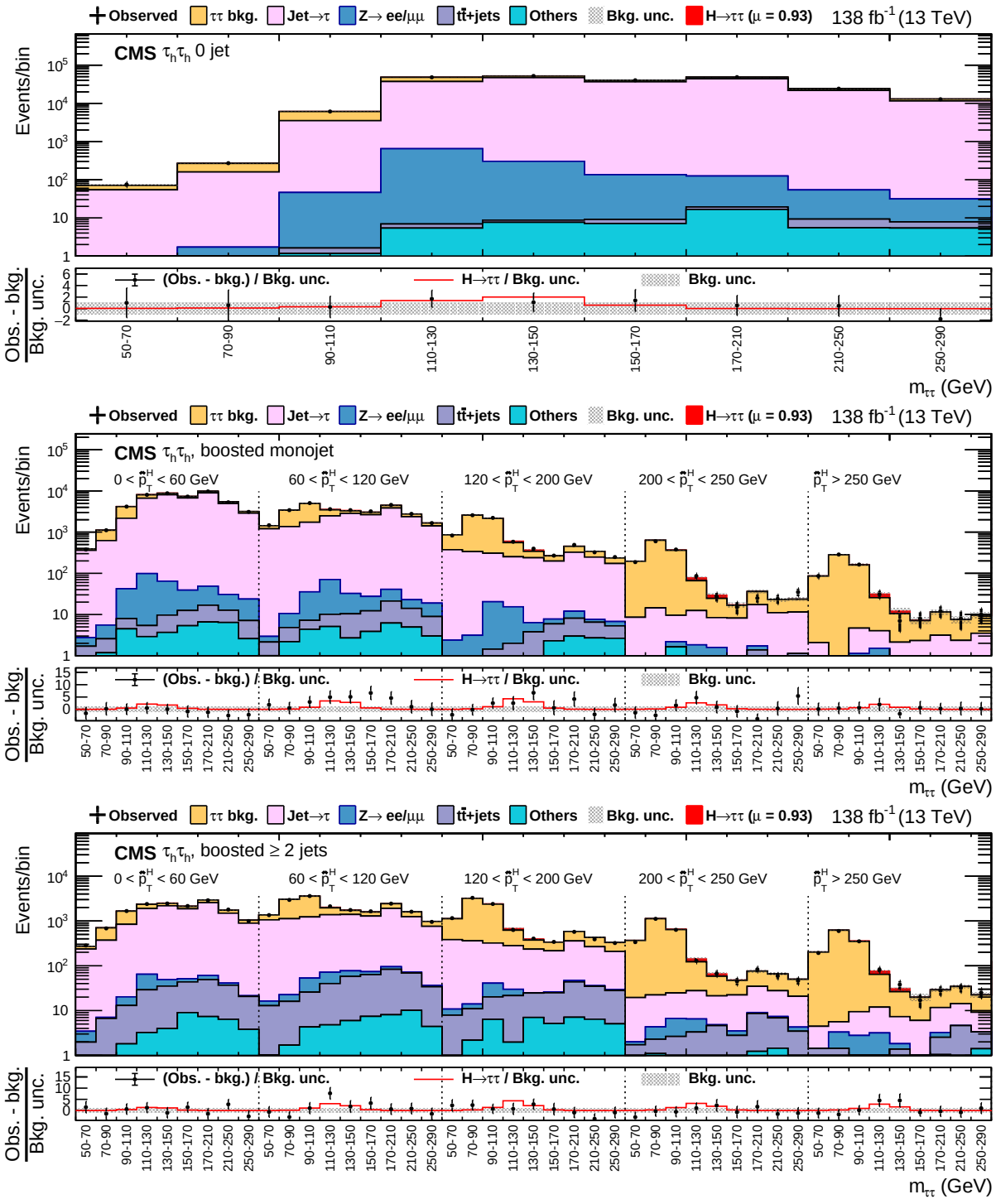

Figure 7.3: *Observed and predicted distributions in the $\tau_h\tau_h$ channel, 0 jet (top), boosted monojet (middle), and boosted multijet (bottom) categories. The signal distribution is normalized to the best fit inclusive signal strength. The uncertainty bands encompass all systematic and statistical uncertainties, after the fit to the data.*[3]

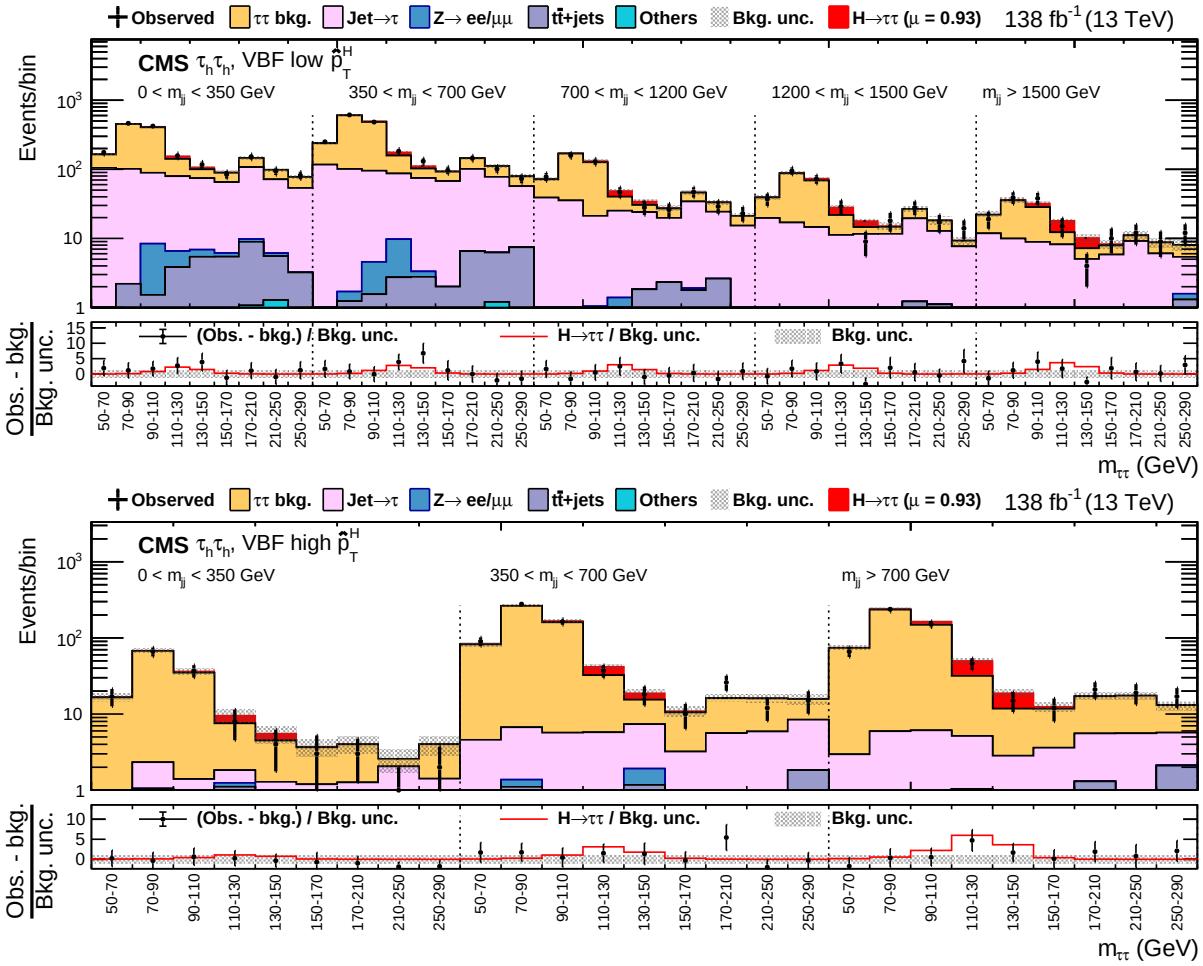

Figure 7.4: *Observed and predicted distributions in the $\tau_h\tau_h$ channel, VBF low Higgs p_T (top), and VBF high Higgs p_T (bottom) categories. The signal distribution is normalized to the best fit inclusive signal strength. The uncertainty bands encompass all systematic and statistical uncertainties, after the fit to the data.*[3]

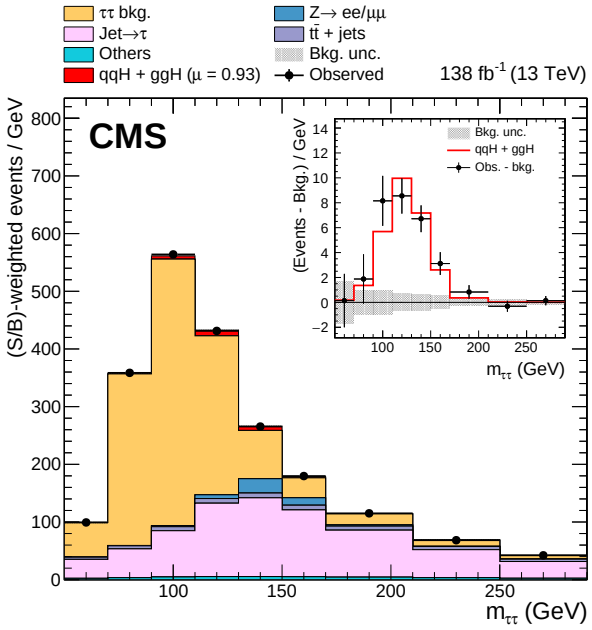

Figure 7.5: *Observed and expected $m_\tau\tau$ distribution obtained by reweighting every $m_\tau\tau$ distribution of each category, year, and final state by the ratio between the signal and background yield in bins with $90 < m_{\tau\tau} < 150$ GeV.*[3]

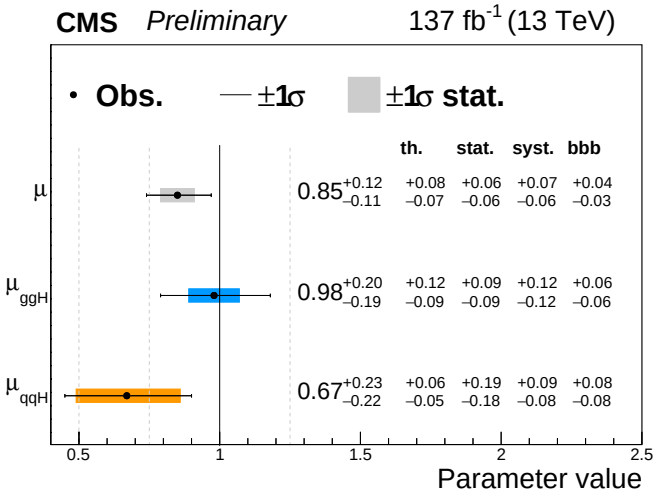

Figure 7.6: *Observed results of the fit to the inclusive signal strength modifier. The contributions to the total uncertainty in each parameter from the statistical systematic, theoretical systematic, bin-by-bin systematic, and other experimental systematic are shown. Also, the results of the fit to production mode signal strength modifiers are shown.*[3]

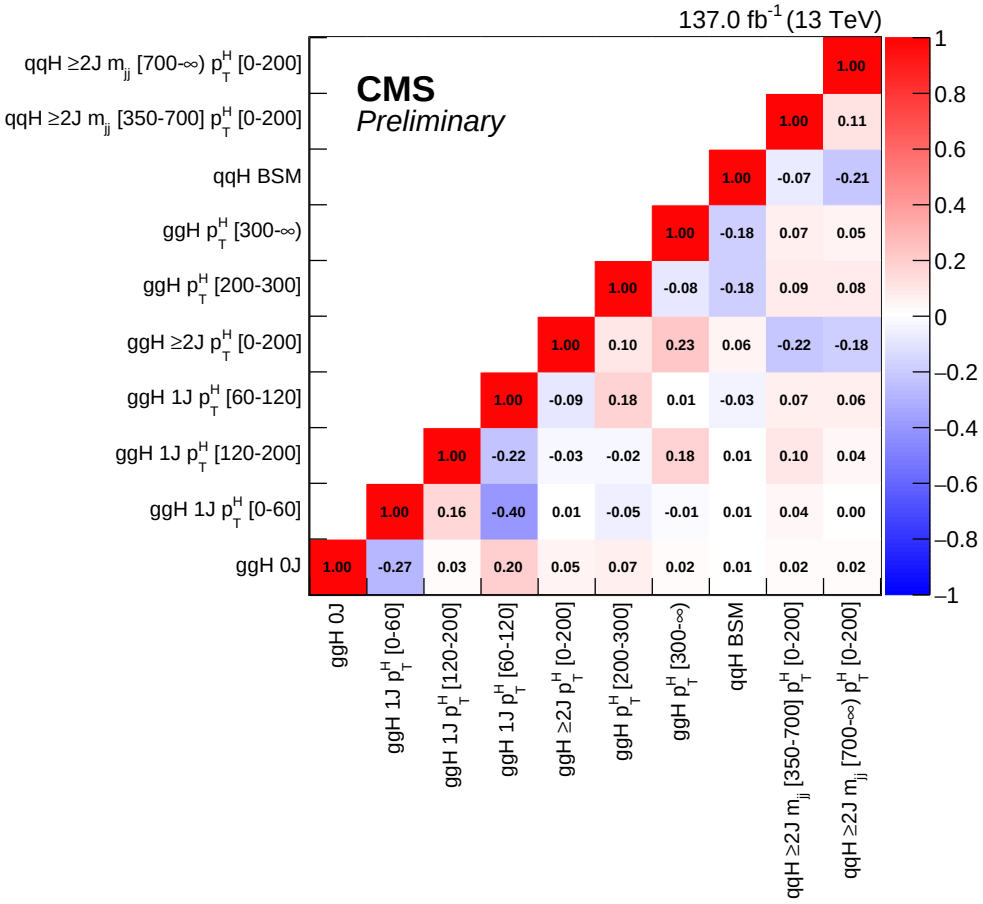

Figure 7.7: *Correlation matrix between merged stage 1.2 STXS parameters.*[3]

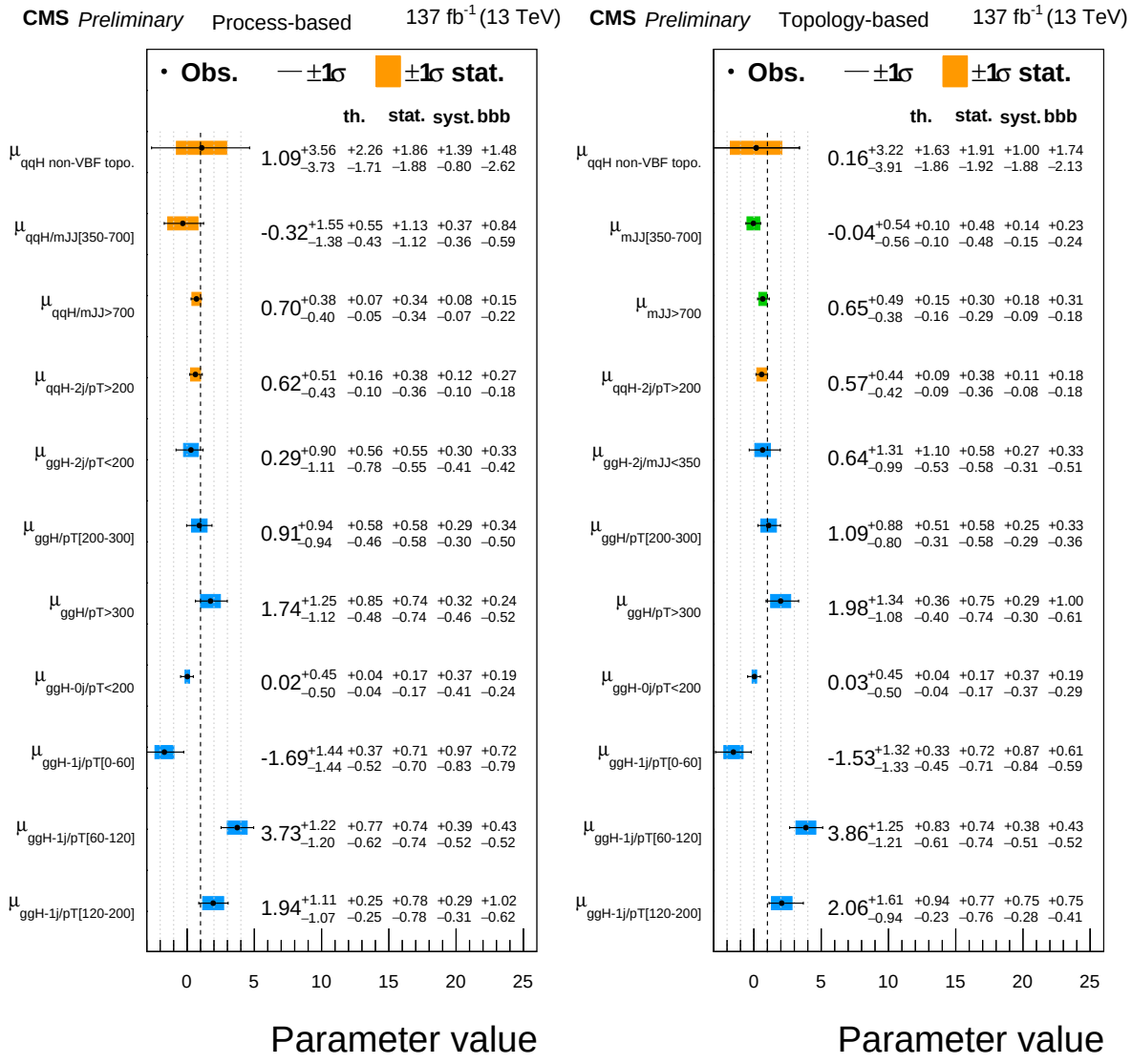

Figure 7.8: *Observed stage 1.2 parameters obtained from the fit of all data taking years, categories, and final states. The left plot corresponds to the process-based merging scheme, and ther right plot to the topology-based merging scheme. The contributions from the statistical uncertainty to the total uncertainty in each parameter are shown as colored squares. The ggH processes are indicated in blue while qqH are indicated in orange. The green squares indicate parameters for the ggH and qqH combined productions in the topology-based merging scheme. All parameters are measured simultaneously.*[3]

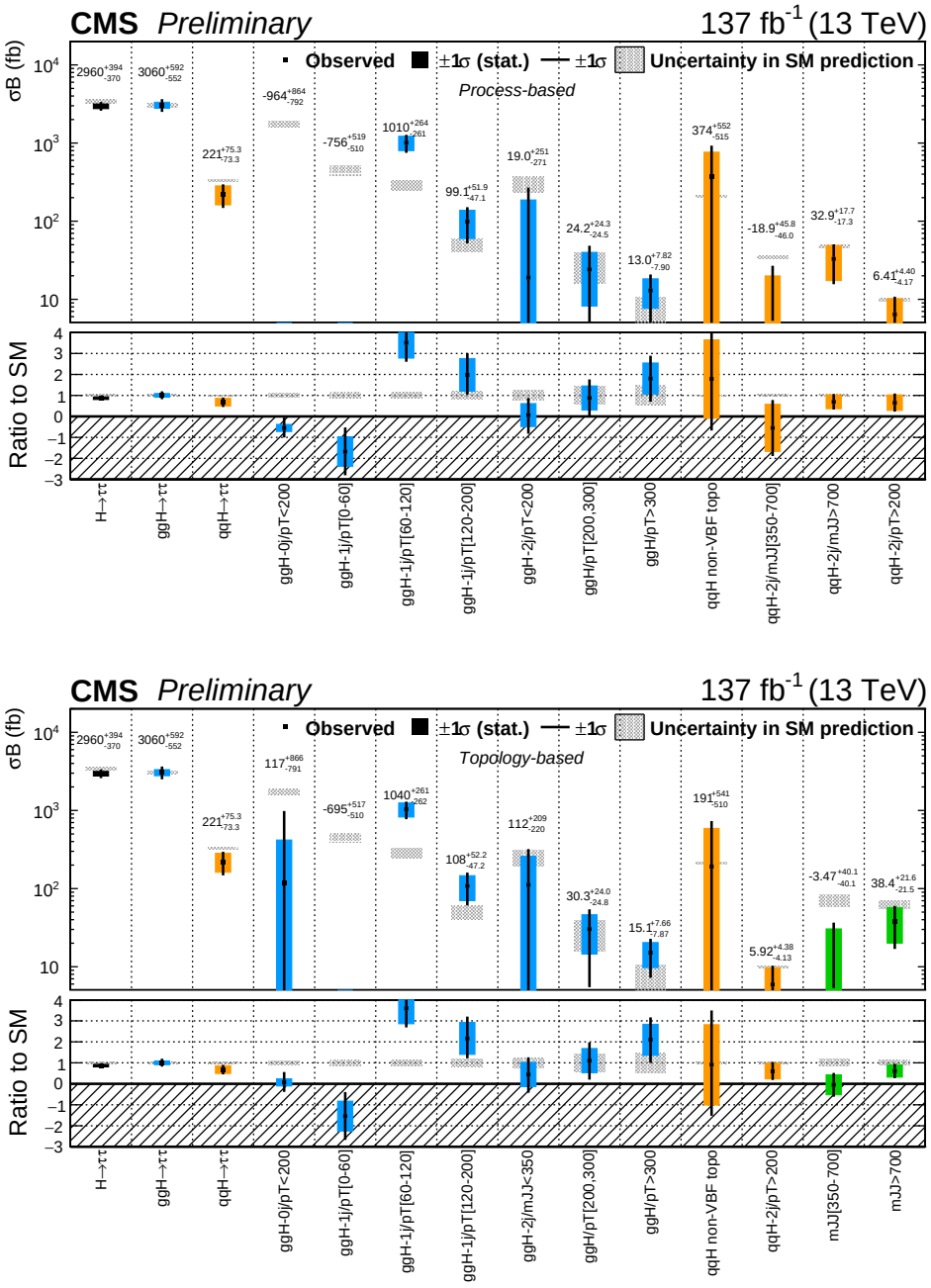

Figure 7.9: *Production of the cross section and branching ratio measured for the inclusive, stage 0 and stage 1.2 parameters. The top plot corresponds to the process-based merging scheme, and the bottom plot to the topology-based merging scheme. The ggH processes are indicated in blue while qqH are indicated in orange. The green squares indicate parameters for the ggH and qqH combined productions in the topology-based merging scheme.*[3]

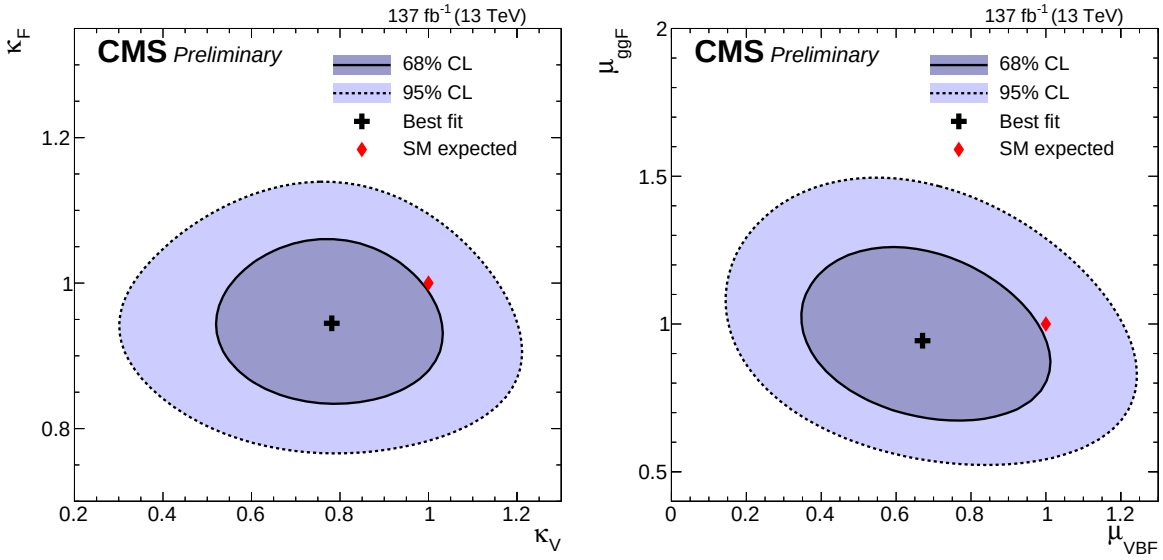

Figure 7.10: *The negative log-likelihood scans as a function of κ_V and κ_f (left) and the coupling modifiers for the ggF and VBF productions (right), for $m_H = 125.09$ GeV.*[3]

Chapter 8

Measurement of the inclusive and differential Higgs boson production cross sections

This chapter describe measurement of inviusive and differential fiducial cross sections of the Higgs boson using the τ lepton decay channel. The differential cross sections are measured as functions of three key observable. kinematic variables: 1) the Higgs boson transverse momentum (p_T^H), 2) the number of jets produced along with the Higgs boson (N_{jets}), and 3) the transverse momentum of the leading jet (p_T^{jet1}). The analysis is performed using proton-proton data collected with the CMS detector at a center-of-mass energy of 13 TeV and corresponding to an integrated luminosity of 138 fb^{-1}. These are the first-ever measurements of the Higgs boson cross section in the Higgs boson production decaying to a pair of tau leptons.

8.1 Event selection

The baseline selection in all cases is that described in Section 7.2.

8.2 Event categorization

Selected events are categorized depending on the $p_T^{\tau_h}$ (or sub-leading $p_T^{\tau_h}$ in the $\tau_h\tau_h$ decay channel) to get a higher signal purity increasing the overall sensitivity of the analysis without introducing a strong model dependence. It is clearly seen that events with sub-leading $p_T^{\tau_h}$ above 70 GeV has high signal purity in Fig. 8.1. Simulations indicate that both backgrounds and Higgs signal are abundant in low $p_T^{\tau_h}$ region. Around 70% of backgrounds and 65% of signal are in sub-leading $p_T^{\tau_h} < 50$ GeV region in the $\tau_h\tau_h$ channel which means low $p_T^{\tau_h}$ category very useful to constraint backgronds.

In the $\tau_h\tau_h$ final state, the categories are defined with the following requirements based on sub-leading $p_T^{\tau_h}$: $40 < p_T^{\tau_h} < 50$, $50 < p_T^{\tau_h} < 70$, and $p_T^{\tau_h} > 70$ GeV. In the $e\tau_h$ and $\mu\tau_h$ channels, lower bound of low $p_T^{\tau_h}$ category is lower because of lower HLT p_T threshold: $30 < p_T^{\tau_h} < 50$, $50 < p_T^{\tau_h} < 70$, and $p_T^{\tau_h} > 70$ GeV. Categorization is not used in the $e\mu$ channel because the signal-to-background ratio does not visibly increase with the lepton p_T.

8.3 Higgs p_T resolution

The reconstructed Higgs p_T is calculated as the vectorial p_T sum of two visible τ's decay products, and the missing transverse momentum p_T^{miss}. Limited resolution of p_T^{miss}, hence, propagated to the resolution of the reconstructed Higgs p_T, which is shown in Figure 8.2 for events with high Higgs p_T in the $\tau_h\tau_h$ final state. In order to have an unbiased estimator of the generated Higgs p_T, a prescalibration factor that is measured in signal simulation and depends on the reconstructed Higgs p_T is multiplied with the reconstructed Higgs p_T. While the correction factor is significantly below 1.0 at low Higgs p_T values, it reaches a plateau between 1.05 and 1.10 at high Higgs p_T values. The corrections are shown in Figure 8.3 for the $\tau_h\tau_h$ final state. The corrections are measured in the Higgs p_T binning that is aligned with the definition of the Higgs p_T thresholds separating different generated-level bins.

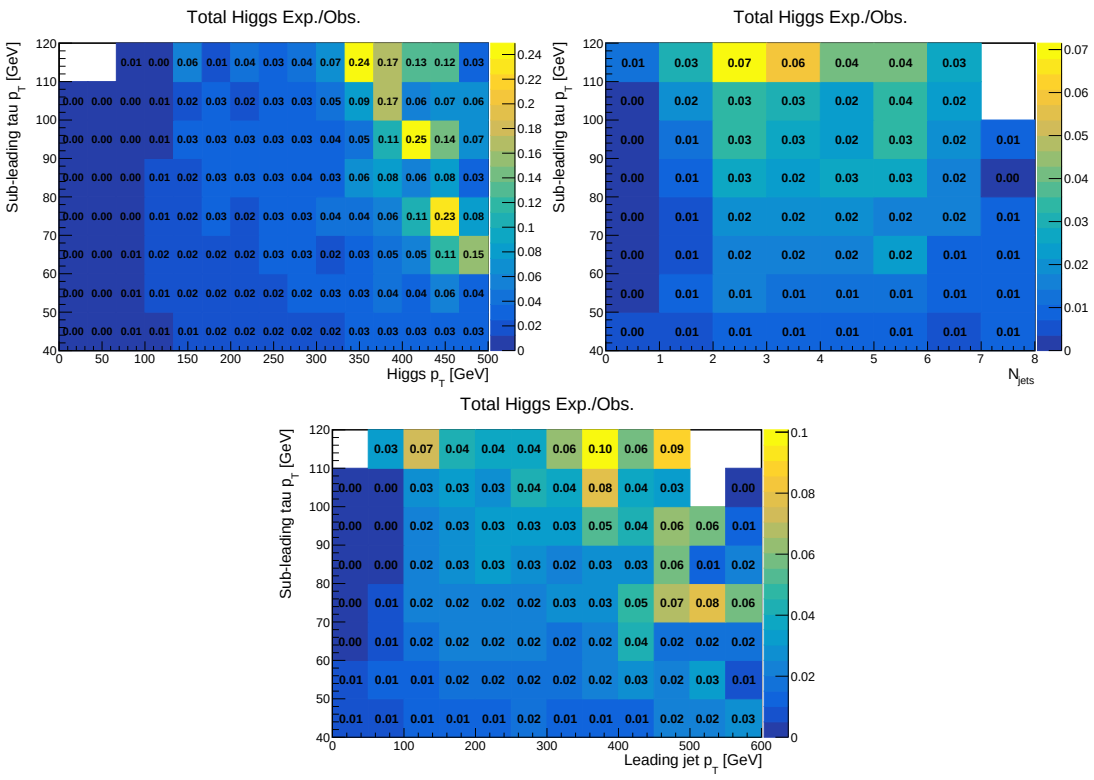

Figure 8.1: *2D distribution ratio of expected total Higgs signal yield over observed yield (Higgs expected/Observed data) in 2018 $\tau_h\tau h$ decay channel.*

8.4 Systematic uncertainties

All systematic uncertainties discussed in Chapter 6 are considered in this analysis.

8.5 Signal extraction and statistical interpretation

8.5.1 Fiducial region

It is difficult to choose a fiducial region that accurately reflects the analysis selection because four different final states use different trigger thresholds and reconstruction conditions. Furthermore, all the analysis selection is based on reconstructed visible taus which means the neutrino(s) from the tau lepton decay is excluded. Hence, the fiducial regions are defined in each final state such that fiducial regions are as close as possible to the reconstructed event

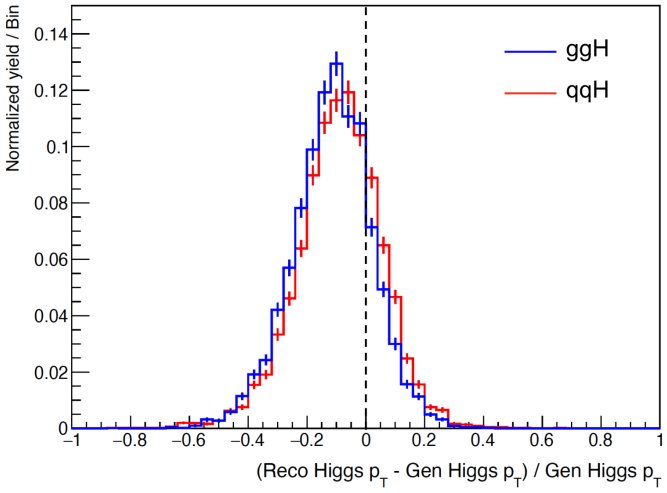

Figure 8.2: *The resolution of Higgs p_T in high p_T (> 200 GeV) region in 2018 $\tau_h\tau_h$ channel.*

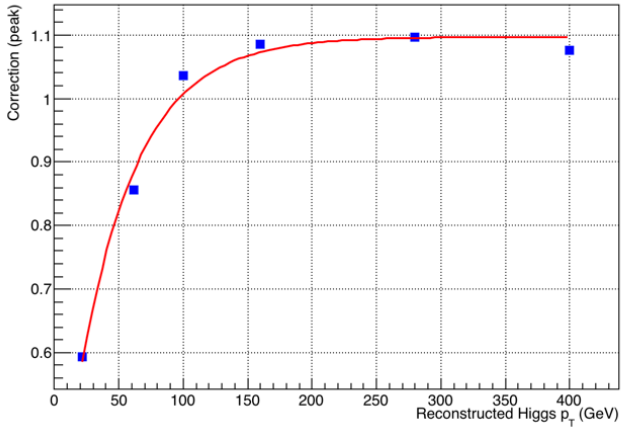

Figure 8.3: *Correction factor applied to the reconstructed Higgs p_T in $\tau_h\tau_h$ channel.*

selection.

Decay of the Higgs boson other than the τ lepton pair are considered to be outside the fiducial region. About 97% of the $\tau_h\tau_h$ final state events passing the reconstructed event selection belong to the fiducial region as estimated from simulation. The selection efficiency for events generated in the fiducial region is 91∼95% in the $e\mu$ final state depending on year. In the $l\tau_h$ final states, 93% of selected events are in fiducial region with the largest loss due to the cut on $m_T(l, \text{MET}) < 50$ GeV at generator level.

H$\rightarrow \tau\tau$ events (e.g. H$\rightarrow$ WW is treated as non fiducial)	
$\tau_h\tau_h$	Generated τ_h $p_T > 40$ GeV and $\|\eta\| < 2.1$ for both legs Generated $\tau_h\tau_h$ decay At least one generated hadronic jet
$\mu\tau_h$	Generated μ $p_T > 20$ GeV and $\|\eta\| < 2.1$ Generated τ_h $p_T > 30$ GeV and $\|\eta\| < 2.3$ Generated $m_T(\mu, \text{MET}) < 50$ GeV Generated $\mu\tau_h$ decay
$e\tau_h$	Generated e $p_T > 25$ GeV and $\|\eta\| < 2.1$ Generated τ_h $p_T > 30$ GeV and $\|\eta\| < 2.3$ Generated $m_T(e, \text{MET}) < 50$ GeV Generated $e\tau_h$ decay
$e\mu$	Generated leading e/μ $p_T > 24$ GeV Generated sub-leading e/μ $p_T > 15$ GeV Generated e and μ $\|\eta\| < 2.4$ Generated $m_T(e+\mu, \text{MET}) < 60$ GeV Generated $e\mu$ decay

Table 8.1: *Fiducial region definition*

The fiducial region is defined in Table 8.1.

8.5.2 Signal splitting at generated level

The signal is split into several templates depending on the generated Higgs p_T for the p_T^H measurement, on the value of the number of generated jet for the N_{jet} measurement, and on the generated leading-jet for the p_T^{jet1} measurement.

At the generator level, p_T^H, N_{jets}, and p_T^{jet1} are evaluated making use of RIVET[63] implementation for the simplified template cross section scheme[9]. In the implementation, generated jets are identified using the anti-k_T algorithm[31] which clustered from final state particles originated from the primary vertex, excluding the decay products of the Higgs boson. Furthermore, the generated jets are required to have $p_T > 30$ GeV, and distance parameter $R = 0.4$.

The binning in the generated variables are chosen to ensure a statistical significance of more than 2σ for the signal process in each bin, or parameter of interest. The binning is also chosen to minimise migrations between bins and aligned with other CMS and ATLAS

measurements in different Higgs decay channels for the combination. The binning is common between the fiducial space and at the reconstruction level, with the exception of p_T^H binning in the $\tau_h\tau_h$ and $e\tau_h$ channels. In those cases, two bins are merged at the reconstruction level because the bins are statistically limited. They are given by the following boundaries.

- p_T^H: [0, 45, 80, 120, 200, 350, 450, ∞) GeV
  p_T^H: [0, 80, 120, 200, 350, 450, ∞) GeV for the $\tau_h\tau_h$ final state, reconstruction level
  p_T^H: [0, 45, 80, 120, 200, 350, ∞) GeV for the $e\tau_h$ final state, reconstruction level
- N_{jets}: [0, 1, 2, 3, 4, ∞)
- p_T^{jet1}: [0 jet, 30, 60, 120, 200, 350, ∞) GeV

Figure 8.4 shows the response matrices of signal processes in bins of p_T^H, N_{jets}, and p_T^{jet1}. All Higgs production modes are taken into account inclusively. Columns of the left plots are normalized to the unity to shows the efficiency, the fraction of reconstructed events that pass the corresponding the generator level selection. On the other hand, the right plots row normalized to effectively show the purity of a given signal bin in each reconstructed category. Bin migration is unavoidable in low p_T^H bins due to p_T^{miss}.

8.5.3 Fit strategy - signal extraction and unfolding procedure

Signal efficiency is imperfect as shown in Figure 8.4 because limited resolution of the detector distorts measured distribution. In that, signal events from one reconstruction level bin contribute to multiple generator level bins. To extract the results, each generator level signal in the fiducial region is modeled as a different freely floating parameter of interest. By performing one simultaneous fit over all reconstruction level bins, the signal strength modifiers of the different generator level observable bins can be determined using all the selected events. There are reconstructed events which are outside the fiducial region, called out of acceptance (OOA). In case of signal, those events are mainly $H \to WW$ decays, as well as a very small fraction of $H \to \tau\tau$ decays. The signal strength of processes outside the fiducial region is constrained to the SM expectation.

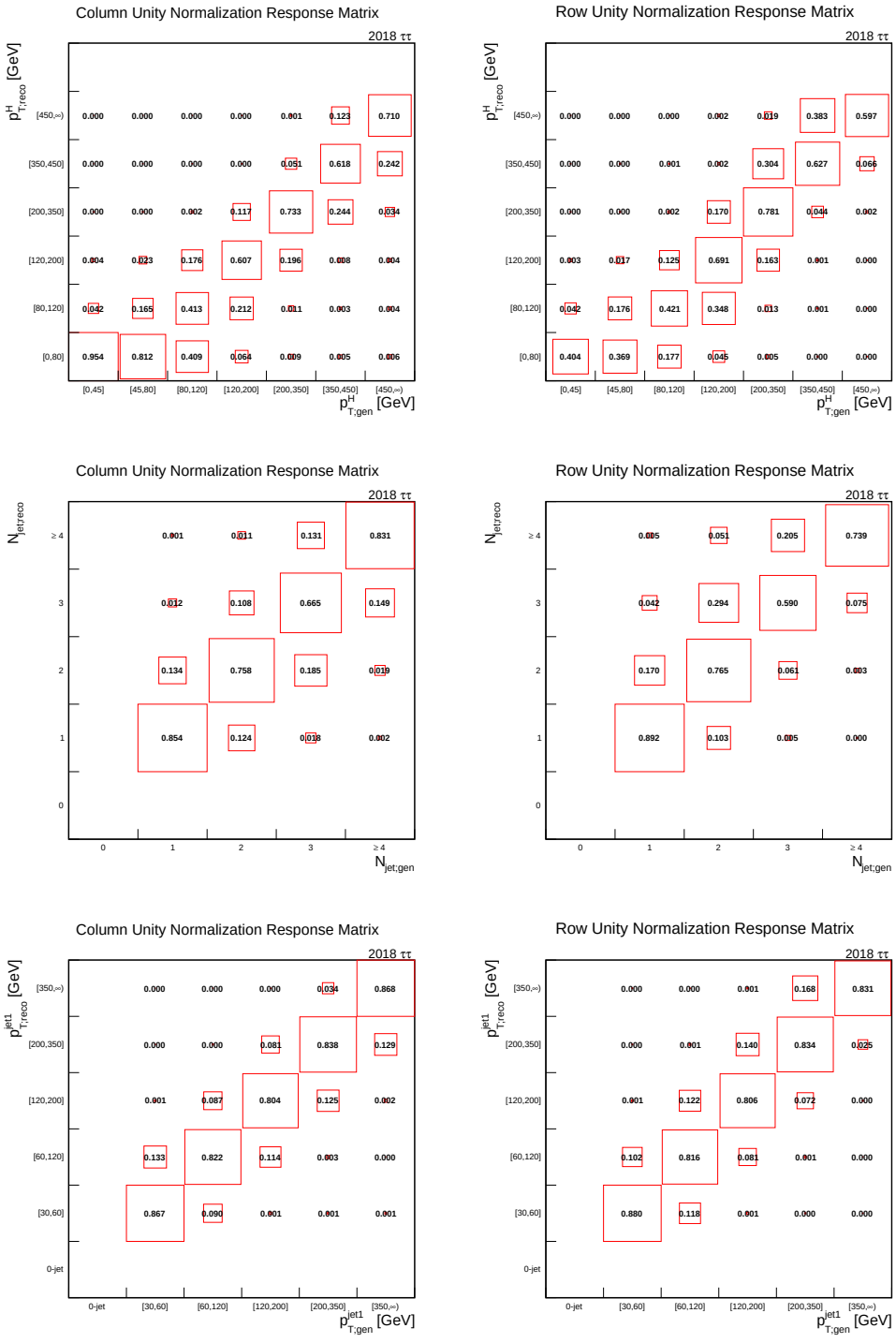

Figure 8.4: *Response matrices for the reconstructed vs. generated p_T^H (top), N_{jets} (middle), and p_T^{jet1} (bottom) in the $\tau_h\tau_h$ channel for 2018. The horizontal axis represents the generator-level bins and the vertical axis represents the reconstruction-level bins. Each column is normalized so that event yield in the column adds up to unity.*

This simultaneous fit can be interpreted as signal extraction in the reconstruction level bins and its unfolding into generator level bins, carried out in a single step. The unfolding problem is essentially solving the linear relationships in $\boldsymbol{n} = R\boldsymbol{\mu} + \boldsymbol{b}$ where $\boldsymbol{n}$, $\boldsymbol{\mu}$, and $\boldsymbol{b}$ are vectors of the observed events, the signal strength modifiers, and the reconstructed background and R is the response matrix. The element of response matrix R_{ij} is not necessarily symmetric with the first index i referring the bin of the observed histogram (reconstruction level) and the second index j denoting to a bin of the true histogram (generator level).

Therefore, the simultaneous fit is performed to maximize the likelihood function written as

$$\mathcal{L} = \prod_i \text{Poisson}\left(n_i; \sum_j \left[R_{ij}(\boldsymbol{\theta})\mu_j\right] + b_i\right) \cdot \mathcal{C}(\theta_i(\mu)) \tag{8.1}$$

In this likelihood function, $\text{Poisson}(n;\lambda)$ represents the Poisson probability for observing n events under the given expectation of λ. θ_i is nuisance parameters in a fit bin with index i, while the second term $\mathcal{C}$ is the constraints for the nuisance parameters θ_i. The parameters μ_j are the same for all final states with full correlation between each others because major difference lies in the branching ratio of τ decays and the selected phase spaces are similar across all final states.

The unfolding procedure is typically sensitive to the statistical fluctuations and observed estimator can be oscillate widely from bin to bin. This can be also shown in correlation matrices as shown in Figure 8.5, 8.7, and 8.10 because of the large variances between the estimators. To tamed this unphysical effect, a regularization of the unfolded distribution is also performed and compared, which reduce the variances at the cost of introducing bias. In this analysis, Tickonov regularization is used by multiplying a penalty term in the likelihood function (Equation 8.1). In Tickonov regularization, measure of smoothness is the mean value of the square of some derivative of the true distribution or finite differences of adjacent bins. In that, the regularization factor $\mathcal{K}(\boldsymbol{\mu})$ is constructed as

$$\mathcal{K}(\boldsymbol{\mu}) = \prod_{j=1}^{M-2} \exp\left(\frac{-\left[(\mu_{j+1} - \mu_j) - (\mu_j - \mu_{j-1})\right]^2}{2\delta^2}\right), \tag{8.2}$$

where M is the number of bins in the true histogram. The parameter δ determines the strength of the regularization by trading-off between the bias and variance of the estimators. δ is optimized by minimizing the mean of the global correlation coefficient[64] obtained from fits to the Asimov dataset. Optimal values for each of the p_T^H, N_{jets}, and p_T^{jet1} fits are 1.85, 1.35, and 2.35 respectively. Figure 8.11 shows the result of the scan of the regularization strength for three variables.

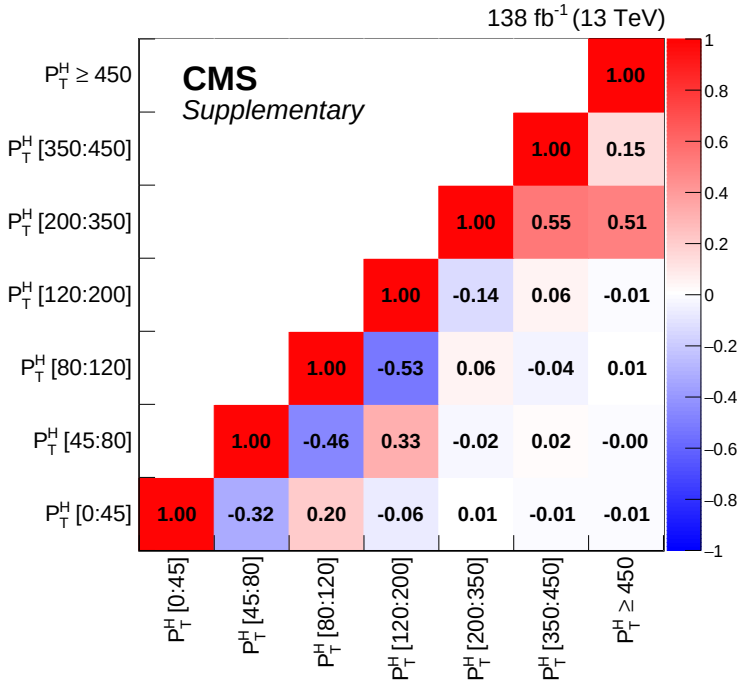

Figure 8.5: *Correlation between signal strength modifiers in fiducial p_T^H bins obtained from fits to data without regularization.*

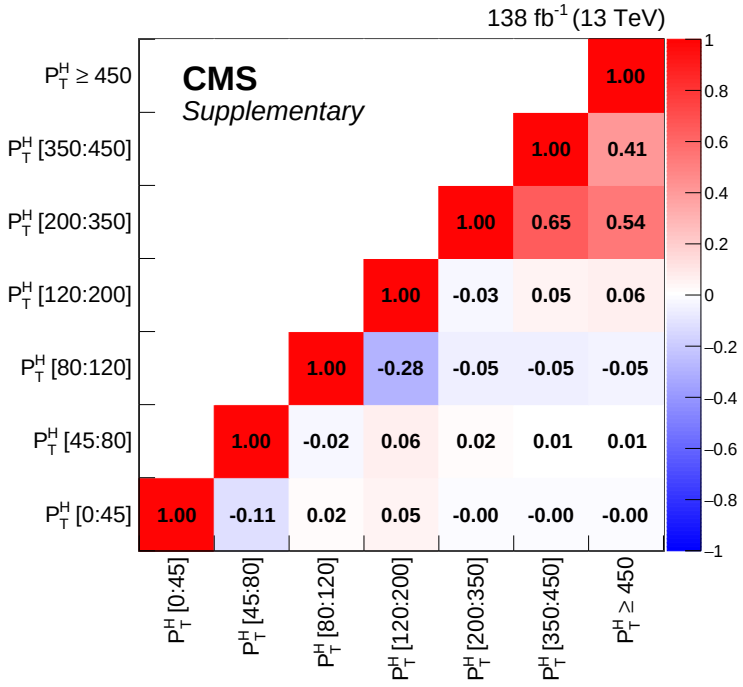

Figure 8.6: *Correlation between signal strength modifiers in fiducial p_T^H bins obtained from fits to data with regularization.*

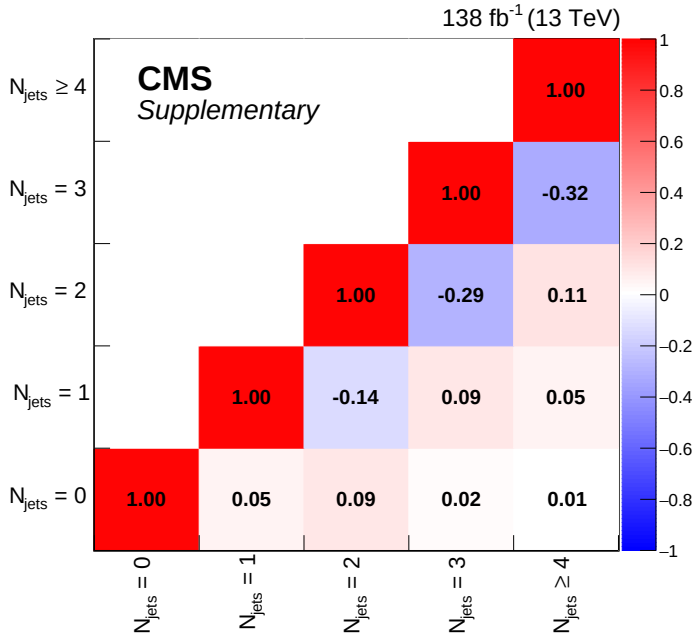

Figure 8.7: *Correlation between signal strength modifiers in fiducial N_{jets} bins obtained from fits to data without regularization.*

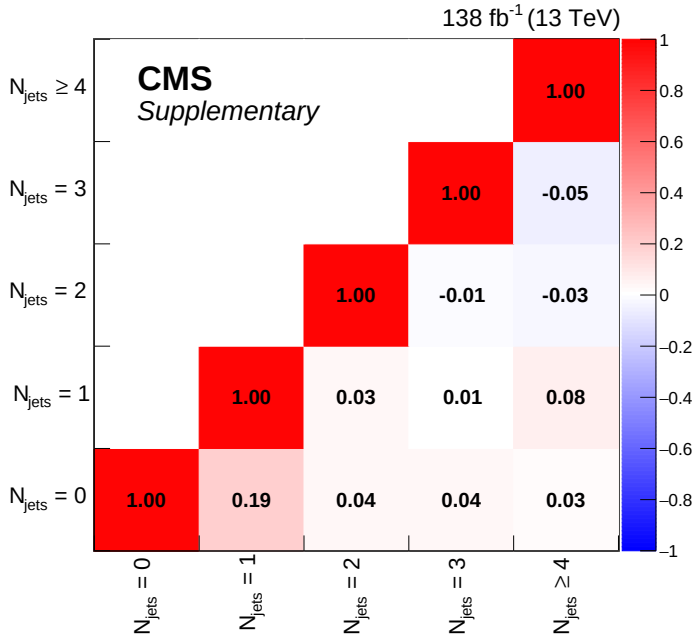

Figure 8.8: *Correlation between signal strength modifiers in fiducial N_{jets} bins obtained from fits to data with regularization.*

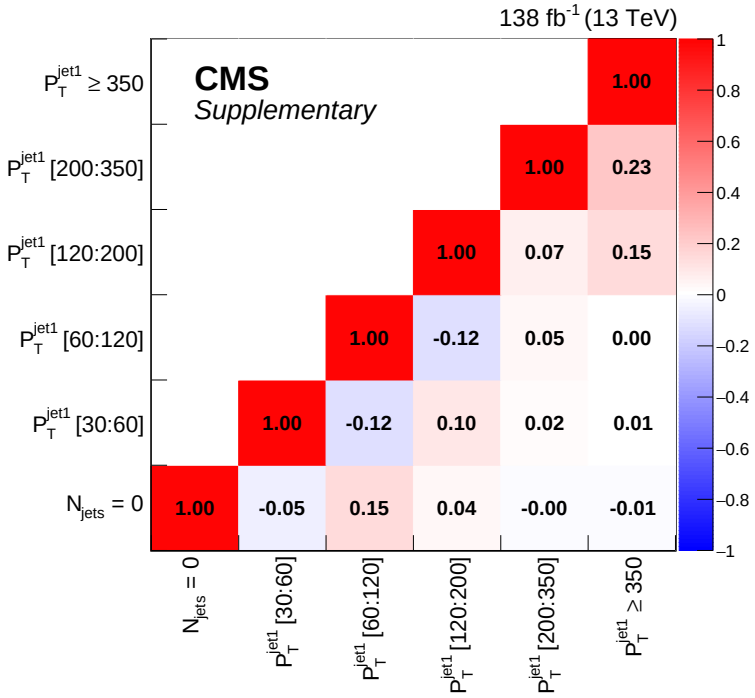

Figure 8.9: *Correlation between signal strength modifiers in fiducial p_T^{jet1} bins obtained from fits to data without regularization.*

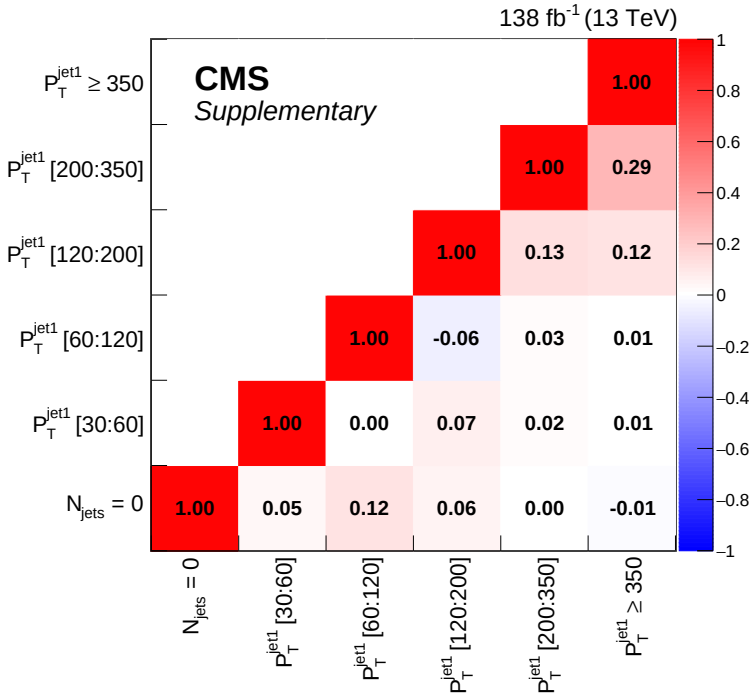

Figure 8.10: *Correlation between signal strength modifiers in fiducial p_T^{jet1} bins obtained from fits to data with regularization.*

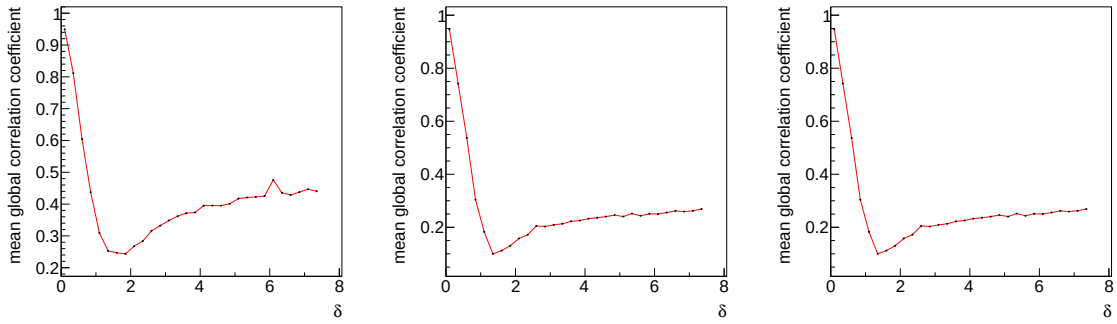

Figure 8.11: *Optimization of the regularization strength δ for p_T^H (left), N_{jets} (middle), and p_T^{jet1} (right) fits, with respect to the mean of the global correlation coefficients.*

Maximum likelihood fits are performed by the CMS Higgs combination tool(`combine`).

8.6 Results

$m_{\tau\tau}$ in bins with $90 < m_{\tau\tau} < 150$ GeV reconstructed with FastMTT algorithm described in 5.6 is chosen as an observable for every parameters of interest. Practically, 2D distributions of $m_{\tau\tau}$ and of the variable considered for the differential measurement (p_T^H, N_{jets}, and p_T^{jet1}) are built in each category and provided as templates. In the summary plots shown in Figure 8.12, all categories and years are summed up and reweighted by the signal-to-background ratio in each of the observable ranges to obtain. The distributions are normalized to conserve signal yield.

The expected and observed fiducial cross sections, regularaized/unregularized signal strength, and regularization biases in the signal strength of the combination of all channels for Run2 are shown in Table 8.2. Figure 8.13 visualized the observed uncertainties with and without regularization. The regularization has the largest influence in the low p_T^H bins, where the correlation between POIs are the strongest. Observed and expected differential fiducial cross section are also shown in Figure 8.14. They are correspond to the regularized fit results.

The total fiducial cross section is measured by using N_{jets} distribution. The fit allows

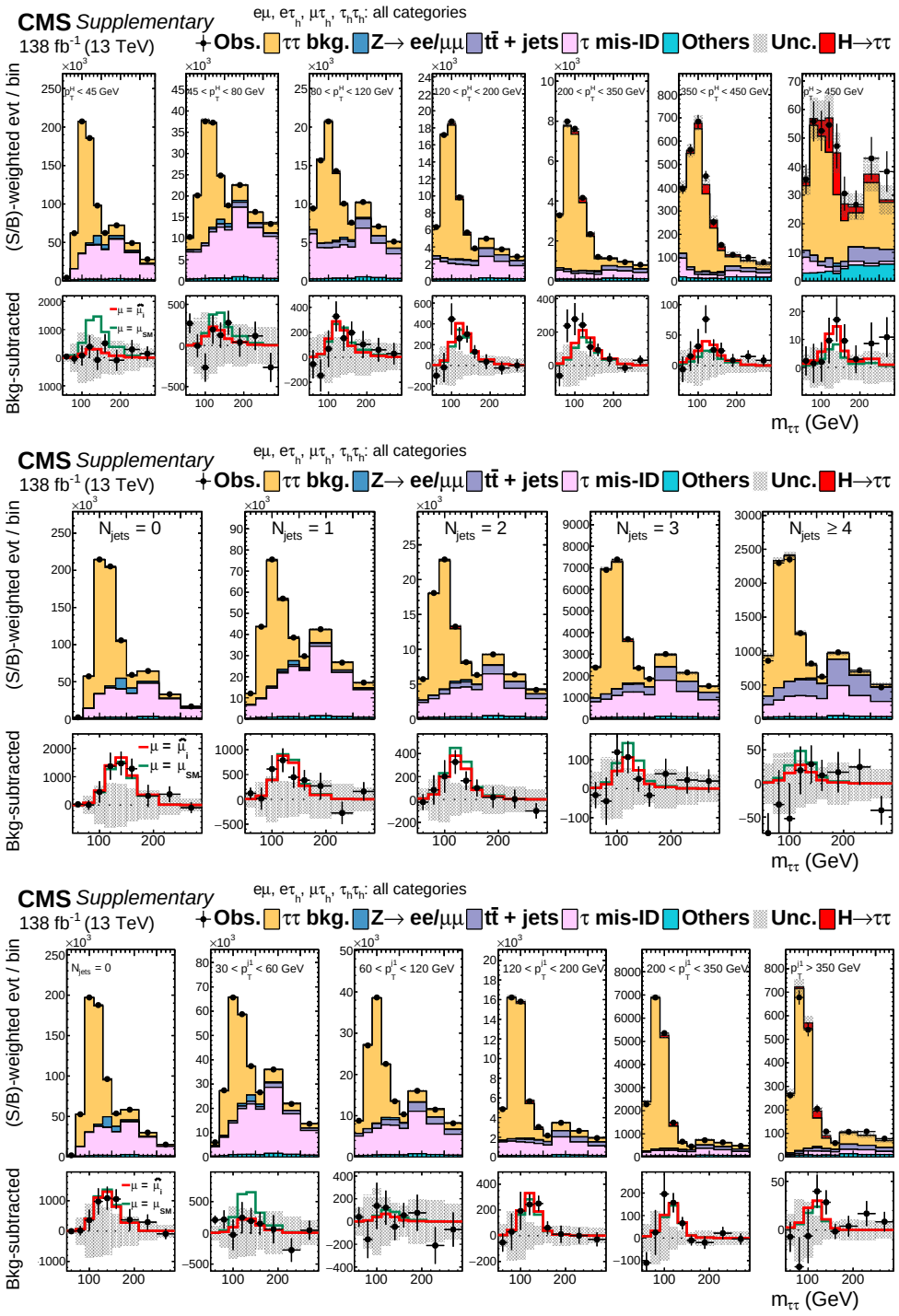

Figure 8.12: *Expected and observed $m_{\tau\tau}$ distributions in the different reconstructed ranges of the differential observable. Every $m_{\tau\tau}$ distribution of each category, year, and final state are reweighted by a factor proportional to the ratio between the signal and background yields. The distributions are normalized to conserve signal yield. The signal and background distributions result from a multidimensional maximum likelihood regularized fit. The green line in the ratio plot shows the expected SM signal, while the red line corresponds to the best-fit signal prediction. The uncertainty band includes systematic and statistical uncertainties.*[4]

Generated bin	$\sigma_{\text{fid}}^{\text{SM}}$ (fb)	Unregularized μ	Regularized μ	$\sigma_{\text{fid}}^{\text{obs}}$ (fb)	Bias
$N_{jets} = 0$	168 ± 15	$1.26^{+0.51}_{-0.50}$	$1.25^{+0.46}_{-0.48}$	210^{+86}_{-84}	-0.13
$N_{jets} = 1$	138 ± 15	$0.99^{+0.30}_{-0.30}$	$1.00^{+0.28}_{-0.27}$	138^{+39}_{-37}	-0.03
$N_{jets} = 2$	71.1 ± 9.4	$0.87^{+0.26}_{-0.25}$	$0.84^{+0.22}_{-0.22}$	$59.7^{+15.6}_{-15.6}$	-0.01
$N_{jets} = 3$	21.1 ± 3.2	$0.48^{+0.50}_{-0.48}$	$0.58^{+0.36}_{-0.35}$	$12.2^{+7.6}_{-7.4}$	$+0.08$
$N_{jets} \geq 4$	10.2 ± 1.7	$0.63^{+0.53}_{-0.50}$	$0.57^{+0.46}_{-0.45}$	$5.81^{+4.69}_{-4.59}$	-0.05
$p_T^H < 45$ GeV	209 ± 16	$0.35^{+0.49}_{-0.50}$	$0.11^{+0.43}_{-0.44}$	$23.0^{+89.9}_{-92.0}$	-0.01
$45 < p_T^H < 80$ GeV	81.7 ± 7.0	$-0.40^{+0.71}_{-0.71}$	$0.34^{+0.46}_{-0.45}$	$27.8^{+37.6}_{-36.8}$	$+0.02$
$80 < p_T^H < 120$ GeV	48.5 ± 4.7	$2.00^{+0.68}_{-0.66}$	$1.28^{+0.43}_{-0.40}$	$62.1^{+20.9}_{-19.4}$	$+0.02$
$120 < p_T^H < 200$ GeV	45.4 ± 5.9	$0.95^{+0.34}_{-0.33}$	$1.18^{+0.26}_{-0.26}$	$53.6^{+11.8}_{-11.8}$	-0.01
$200 < p_T^H < 350$ GeV	20.1 ± 4.9	$1.17^{+0.46}_{-0.37}$	$1.18^{+0.41}_{-0.36}$	$23.7^{+8.2}_{-7.2}$	$+0.01$
$350 < p_T^H < 450$ GeV	2.49 ± 0.62	$1.79^{+0.93}_{-0.72}$	$1.68^{+0.67}_{-0.59}$	$4.18^{+1.67}_{-1.47}$	$+0.03$
$p_T^H > 450$ GeV	0.823 ± 0.209	$1.82^{+1.19}_{-0.94}$	$2.00^{+1.02}_{-0.85}$	$1.65^{+0.84}_{-0.70}$	$+0.06$
$N_{jets} = 0$ (p_T^{jet1}-cat.)	168 ± 15	$1.20^{+0.52}_{-0.52}$	$1.11^{+0.50}_{-0.50}$	186^{+84}_{-84}	N/A
$30 < p_T^{jet1} < 60$ GeV	108 ± 11	$0.13^{+0.46}_{-0.46}$	$0.24^{+0.42}_{-0.41}$	$25.9^{+45.4}_{-44.3}$	N/A
$60 < p_T^{jet1} < 120$ GeV	81.1 ± 7.5	$0.62^{+0.29}_{-0.29}$	$0.60^{+0.27}_{-0.27}$	$48.7^{+21.9}_{-21.9}$	N/A
$120 < p_T^{jet1} < 200$ GeV	34.9 ± 4.2	$1.15^{+0.28}_{-0.27}$	$1.13^{+0.26}_{-0.25}$	$39.4^{+9.1}_{-8.7}$	N/A
$200 < p_T^{jet1} < 350$ GeV	13.5 ± 3.0	$1.02^{+0.39}_{-0.35}$	$1.06^{+0.38}_{-0.34}$	$14.3^{+5.1}_{-4.6}$	N/A
$p_T^{jet1} > 350$ GeV	2.38 ± 0.57	$1.32^{+0.76}_{-0.65}$	$1.30^{+0.72}_{-0.63}$	$3.09^{+1.71}_{-1.50}$	N/A

Table 8.2: *Observed and expected fiducial cross sections in N_{jets}, p_T^H, and p_T^{jet1} bins. The signal strengths, which do not include yield uncertainties in the predictions of the SM Higgs boson production cross sections nor branching fractions, are indicated for the regularized and unregularized fits. The observed cross section corresponds to the result of the regularized fit. Regularization biases in the signal strength[7] are given in the last column and not included in the total uncertainties. They are negligible in the p_T^{jet1} differential measurement. Results for $N_{jets} = 0$ appear twice: the first (second) occurence corresponds to the fit of N_{jets}-based (p_T^{jet1}-based) categories.[4]*

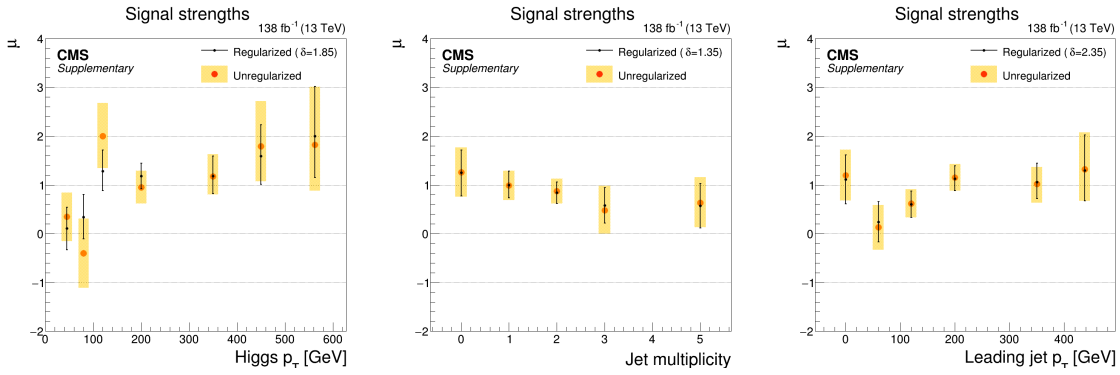

Figure 8.13: *Observed signal strength modifiers in p_T^H (left), N_{jets} (middle), and p_T^{jet1} (right) bins.*

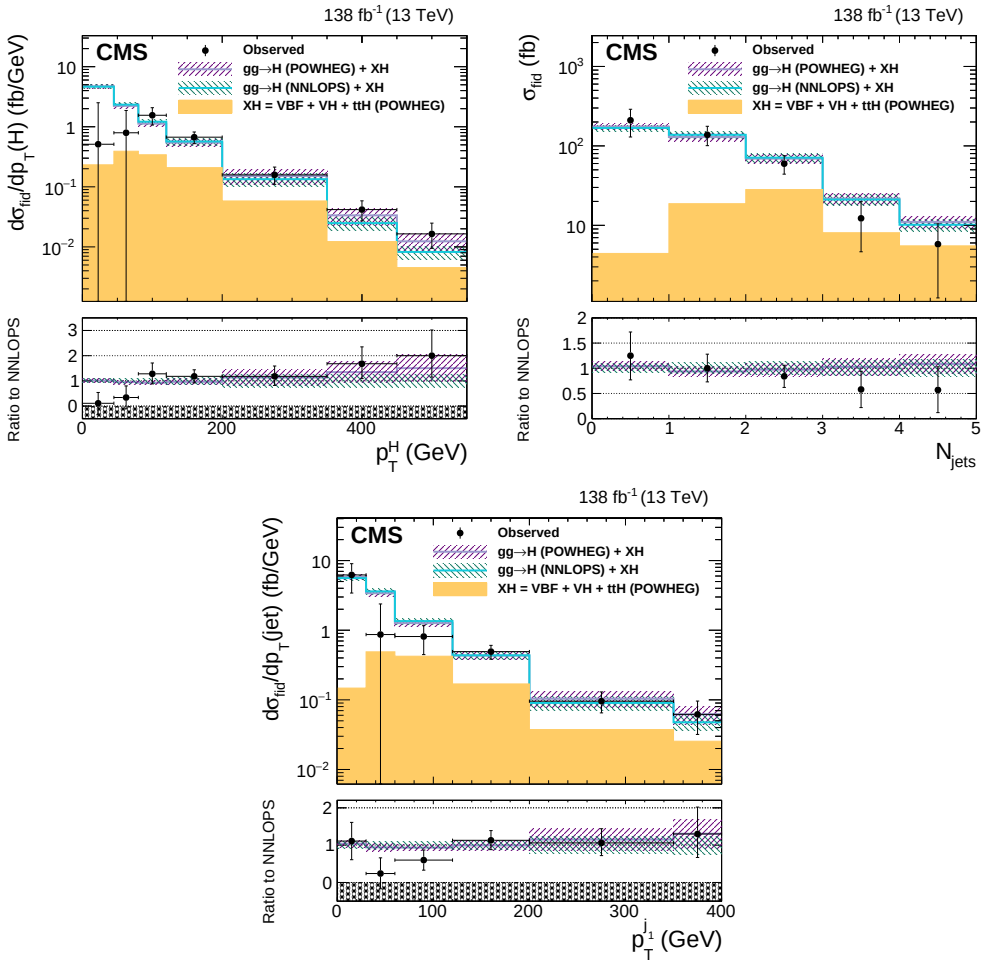

Figure 8.14: *Observed and expected differential fiducial cross section in bins of p_T^H (upper left), N_{jets} (upper right), and p_T^{jet1} (lower). Both regularized (full markers) and unregularized (hollow markers) are shown. The first bin in the p_T^{jet1} plot contains all events without a jet. The shaded bands in the theoretical predictions include uncertainties from the following sources: PDF, renormalization and factorization scale, underlying event and parton showering, and branching fraction of the Higgs boson to τ leptons. Overflow is included in the last bins of each distribution.*[4]

the signal normalization to vary independently in each bin without introducing a model dependence. In this measurement, degrees of freedom is kept to the same but the parameters of interest are reformulated such that one modifies the total inclusive fiducial cross section. Hence, for the $N = 5$ observable bins, each with a fiducial cross section σ_i, each bin is scaled with the product of parameters of interest $\mu_{fid} \times \rho_i$ except for N^{th} bin. On the other hand, bin N is scaled with

$$\mu_{\text{fid}} \times \frac{\sigma_{\text{tot}} - \sum_{i=1}^{N-1} \sigma_i \rho_i}{\sigma_N}. \tag{8.3}$$

The fiducial cross section is measured to be 426 ± 102 fb, compatible with the SM expectation of 408 ± 27 fb.

8.7 Chapter summary

The first differential fiducial cross section of the Higgs boson have been performed at the LHC in the decay channel of two τ leptons. The differential cross sections as functions of the Higgs boson transverse momentum, the jet multiplicity, and transverse momentum of the leading jet, are in agreement with the expectations of the standard model, with competitive precision with respect to measurements in other final states in the phase spaces with a large jet multiplicity, or with a Higgs boson transverse momentum above 120 GeV. In addition, the fiducial inclusisve cross section has been measured to be 426 ± 102 fb, in agreement with the SM expectation of 408 ± 27 fb.

Chapter 9

Search for anomalous couplings of Higgs bosons

This chapter describe a study of the Higgs boson coupling to gauge bosons and fermions. The effects of $\mathcal{CP}$-violation and anomalous couplings in electroweak vector boson or gluon-gluon fusion production mechanisms are analyzed using matrix element and multivariate techniques. The analysis is conducted using data recorded by the CMS experiment at a center-of-mass energy of 13 TeV and correspond to an integrated luminosity of 138 fb^{-1}.

9.1 Event selection

The baseline selection in all cases is that described in Section 7.2.

9.2 Event categorization

Selected events are classified according to four decay channels ($\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$ and $\tau_e\tau_\mu$) to maximize the overall sensitivity of the analysis. The categories are chosen to be mutually exclusive avoiding double counting and all events that passed event selection are covered by the categories. The main purpose of the categorization is to separate the different Higgs

boson production modes, similarly to the categorization used in the STXS measurement descried in Chapter 7. Adopted categories are merged and updated to maximize the sensitivity in this analysis. The categories are defined in Table 9.1.

Final state	Category	Selection
$\tau_h\tau_h$	0-jet	0 jet
	VBF	≥ 2 jets, $\Delta\eta_{jj} > 2.5$
	Boosted	1 jet or (Not in VBF, ≥ 2 jets)
$l\tau_h$, $e\mu$	0-jet	0 jet
	VBF	≥ 2 jets, $m_{jj} > 300$ GeV
	Boosted	1 jet or (Not in VBF, ≥ 2 jets)

Table 9.1: *Event categorization in search for anomalous couplings.*

9.3 Multivariate observables

This analysis used two multivariate techniques to construct powerful discriminants in the VBF category which provides the largest sensitivity in the measurement of anomalous coupling parameters.

9.3.1 Matrix Element Likelihood Approach (MELA)

As discussed in Chapter 2.3.1, kinematics of jets associated with Higgs boson are sensitive to the anomalous couplings. Thus, it is important to have a good description of ggH + 2 jets system. Figure 9.1 shows that the kinematics of the Higgs + 2 jet system can be fully described by five angles and two momenta in the rest frame of the Higgs boson. The distribution of those seven kinematics in the $\tau_h\tau_h$ channel are shown in Figure 9.2 and Figure 9.3. The matrix element likelihood approach, called MELA, is utilized in order to reduce these seven parameters to the minimal number of observables retaining all essential information[5]. The JHUGen relies on matrix element calculations, and it is also possible to extract and use those matrix element probabilities for standalone calculations. MELA package provides an interface to the matrix element calculations contained in JHUGen.

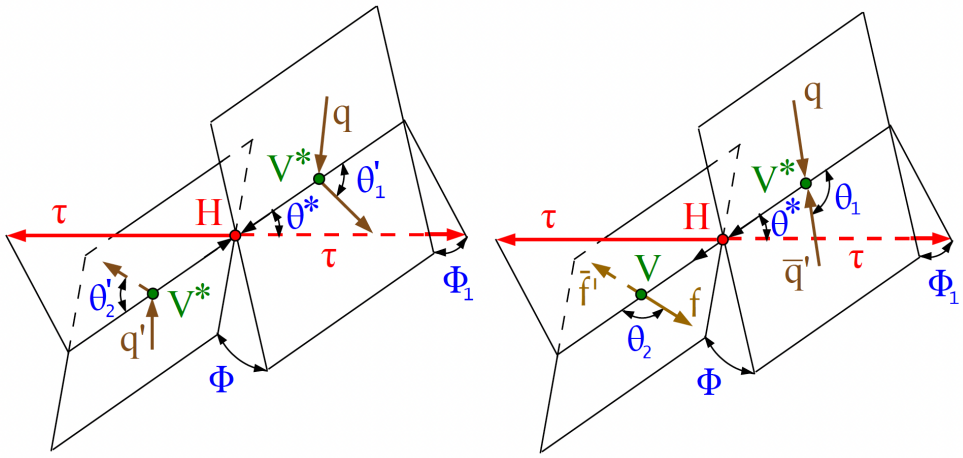

Figure 9.1: *Illustrations of Higgs production in VBF ($qq' \to qq'H$) (left) and VH ($q\bar{q}' \to V^* \to VH \to qq'H$) (right) in the rest frame of the Higgs. Higgs decays to a τ lepton pair is shown without illustrating the further decay chain. Brown shows the incoming partons and fermions in the V decay. The intermediate or final-state particles are shown in red and green. Five angles characterizing kinematic distributions are shown in blue and are defined in the respective rest frames[5]. The illustration for Higgs production via ggH in association with two jets is identical to the VBF diagram, except with $V = g$.*

MELA discriminants

The matrix elements calculated by MELA can be used to build kinematic discriminants that can separate different hypotheses. Three types of MELA discriminants are constructed by taking ratios of probabilities that an event is produced at measured variables $\vec{\Omega}$ under different hypotheses h ($\mathcal{P}_h(\Omega)$). One type of discriminants aims to separate the events with anomalous couplings (denoted as generic BSM hereafter) from the SM events.

$$\mathcal{D}_{BSM} = \frac{\mathcal{P}_{\mathrm{SM}}(\vec{\Omega})}{\mathcal{P}_{\mathrm{SM}}(\vec{\Omega}) + \mathcal{P}_{\mathrm{BSM}}(\vec{\Omega})}, \tag{9.1}$$

The discriminant for each anomalous coupling is listed in Table 9.2.

The second type of discriminant separate the contributions from the interference:

$$\mathcal{D}_{int} = \frac{\mathcal{P}^{\mathrm{int}}_{\mathrm{SM-BSM}}(\vec{\Omega})}{\mathcal{P}_{\mathrm{SM}}(\vec{\Omega}) + \mathcal{P}_{\mathrm{BSM}}(\vec{\Omega})}, \tag{9.2}$$

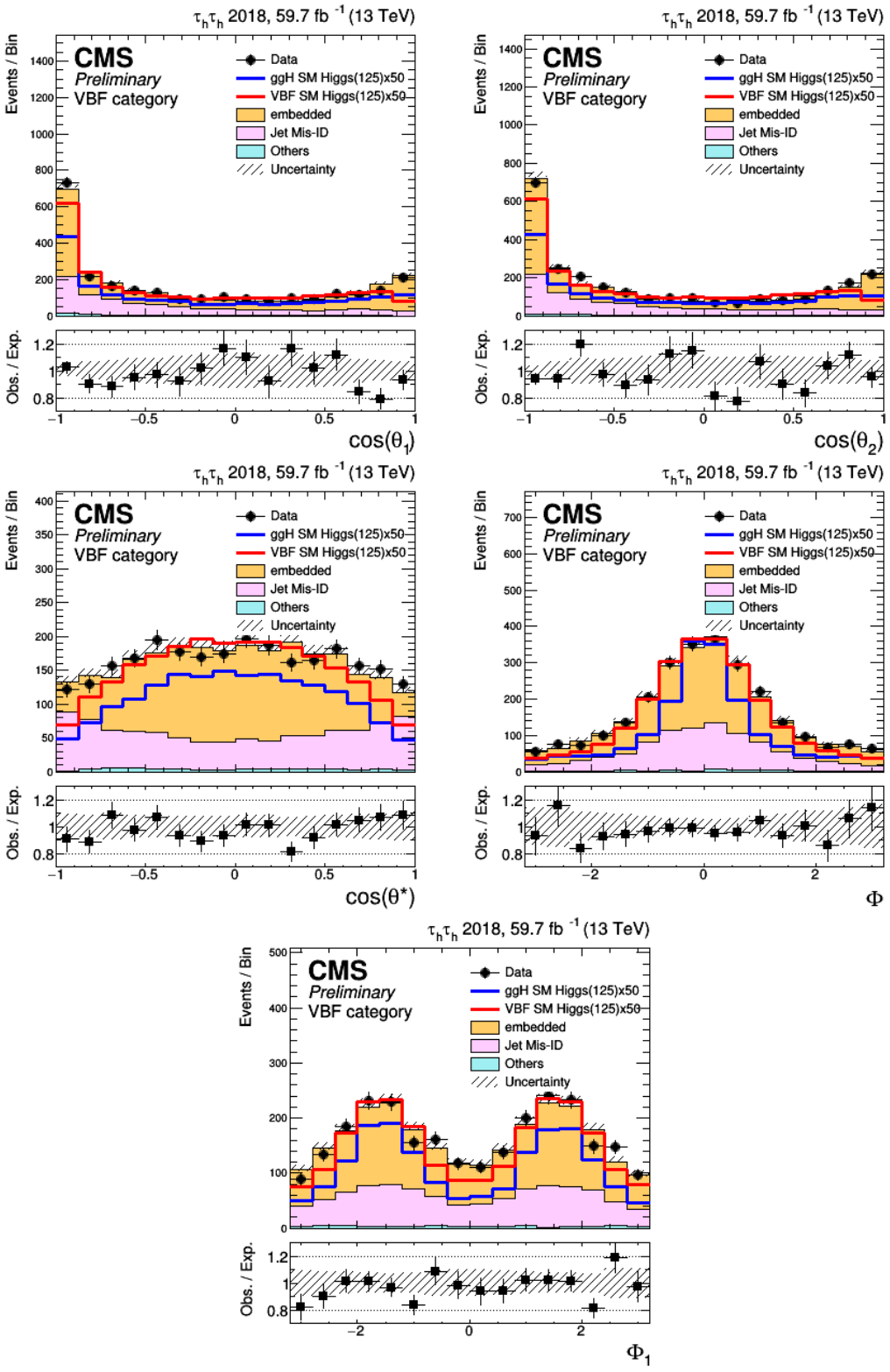

Figure 9.2: *The distribution of five angles characterizing kinematics of the Higgs + 2 jet system in the $\tau_h\tau_h$ channel, 2018. Only statistical uncertainties are included in the uncertainty band.*

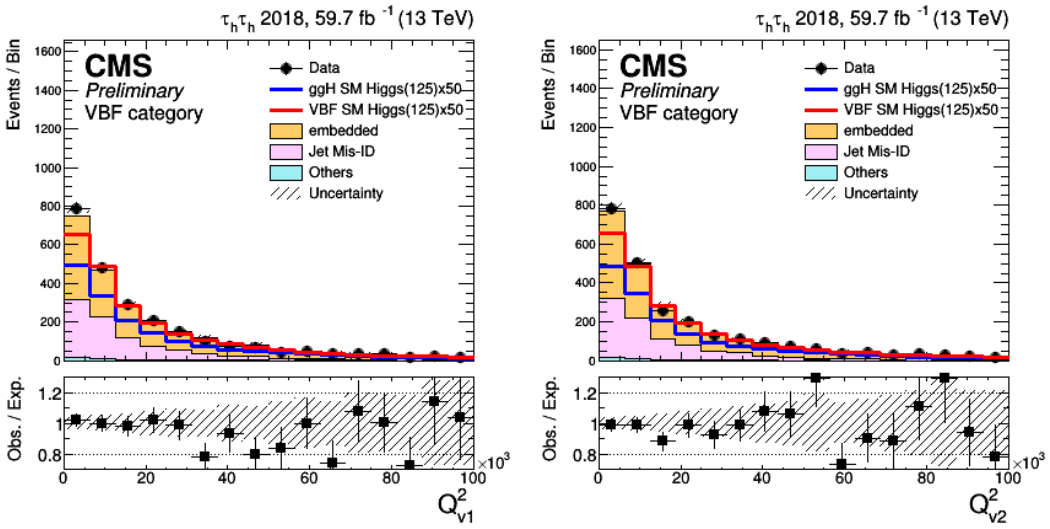

Figure 9.3: *The distribution of two momenta characterizing kinematics of the Higgs + 2 jet system in the $\tau_h\tau_h$ channel, 2018. Only statistical uncertainties are included in the uncertainty band.*

Coupling	Discriminant
a_3^{gg}	$\mathcal{D}_{0-}^{ggH}$
a_3	$\mathcal{D}_{0-}$
a_2	$\mathcal{D}_{0h+}$
κ_1	$\mathcal{D}_{\Lambda 1}$
$\kappa_2^{Z\gamma}$	$\mathcal{D}_{\Lambda 1}^{Z\gamma}$

Table 9.2: *List of discriminants for separating anomalous couplings (BSM) from the SM contribution.*[6]

where $\mathcal{P}^{\text{int}}_{\text{SM}-\text{BSM}}(\vec{\Omega})$ is the interference part of the probability distribution for a process with mixture of the SM and anomalous contributions. There are two discriminants for HVV and Hgg CP-odd parameters and they are called $\mathcal{D}_{CP}^{\text{VBF}}$ and $\mathcal{D}_{CP}^{\text{ggH}}$. Any asymmetry in their distributions would indicate CP violation.

The last type of MELA discriminant utilized in the analysis is called $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$. It is designed to separate ggH production under any kind of hypotheses from VBF production under SM hypothesis. This separation is crucial in the final fit because the signal strength parameters

for each production mode are allowed to float. $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ is defined as follow:

$$\mathcal{D}_{2jet}^{VBF} = \frac{\mathcal{P}_{\text{SM}}^{\text{ggH}}(\vec{\Omega}) + \mathcal{P}_{\text{PS}}^{\text{ggH}}(\vec{\Omega})}{\mathcal{P}_{\text{SM}}^{\text{ggH}}(\vec{\Omega}) + \mathcal{P}_{\text{PS}}^{\text{ggH}}(\vec{\Omega}) + \mathcal{P}_{\text{SM}}^{\text{VBF}}(\vec{\Omega})}. \tag{9.3}$$

9.3.2 Neural network

Three separate neural networks are used in this analysis. Each neural network designed for a specific di-tau decay channel with slightly different architecture because the background composition is different in each channel. The architectures are slightly different, but they are utilized in a similar fashion. The general goal of neural network discriminants is to separate Higgs signal from chosen dominant backgrounds.

$\tau_h\tau_h$ neural network

In $\tau_h\tau_h$ channel VBF category, about 55% of the background is from the $Z \to \tau\tau$ process. However, the second dominant background from the jets misidentified as hadronic tau leptons is non-negligible, contributing at about 30%. As a result, an effective neural network for this channel must be able to separate Higgs signal from both backgrounds simultaneously. In order to achieve this separation, a multi-class neural network is trained to discriminate Higgs signal vs. backgrounds with genuine di-τ (embedded) vs. backgrounds with jet $\to \tau_h$ fake (jet fakes).

The $\tau_h\tau_h$ neural network is constructed with a feed-forward architecture containing two hidden layers and three output nodes to accommodate three distinct classes. The nodes per layer is chosen based on the number of input variables; the first layer contains twice as many nodes as input variables while the second layer contains a number of nodes equal to the number of input variables. This choice is driven by the desire to keep the network as simple as possible without sacrificing performance. By keeping the network simple, we increase the understandability and decrease the probability of over-training. The nodes in each hidden layer utilize the ReLU activation function, while the softmax function is used to activate the output node. These choices are again based on simplicity and resiliency to

over-training. The softmax function also provides the useful ability to normalize the output discriminant between zero and one. Because of this normalization we may interpret the output discriminant as the probability the given event belongs to the target class. From these three output scores or "probabilities"we take the one that corresponds to the signal class as the signal vs background discriminating observable. This discriminant is used later as an observable for the fit. Categorical cross-ntropy is chosen as loss function.

Ten variables are used as input parameters for the neural network: two momenta and five angles that are calculated by MEAL package as shown in Figure 9.2 and Figure 9.3, the dijet mass (m_{jj}), the p_T of the Higgs boson, and the di-tau invariant mass ($m_{\tau\tau}$), which are shown in Figure 9.4. The MELA variables are used to completely describe the kinematics in the rest frame of the Higgs+di-jet system. All input variables are transformed independently to be of zero mean with a variance of one in order to reduce the effects of numeric instability. Fig. 9.5 shows the correlation between the ten input variables and neural network output. The network performance is driven by the complex relationship between $m_{\tau\tau}$ (m_mtt), p_T of the Higgs boson(higgs_pT), and m_{jj} (mjj). The network learns these correlations along with higher order correlations between these observables to drive performance. In addition to these correlations, we know that $m_{\tau\tau}$ is the main driver of performance.

The loss function is monitored for overtraining, which is shown in Figure 9.6. A divergence between the losses would indicate overtraining of the network. The training process is concluded when the validation loss has converged for 60 epochs.

The signal and background classes are scaled to an equal number of events. Prior to the scaling, all the typical corrections are applied per event to give the correct event weight relative to all other events of the same class. These event weights are also transformed to be between 1 and 2 allowing all events to be of use in training the network. The resulting $\mathcal{D}_{NN}$ distribution is shown in Figure 9.7.

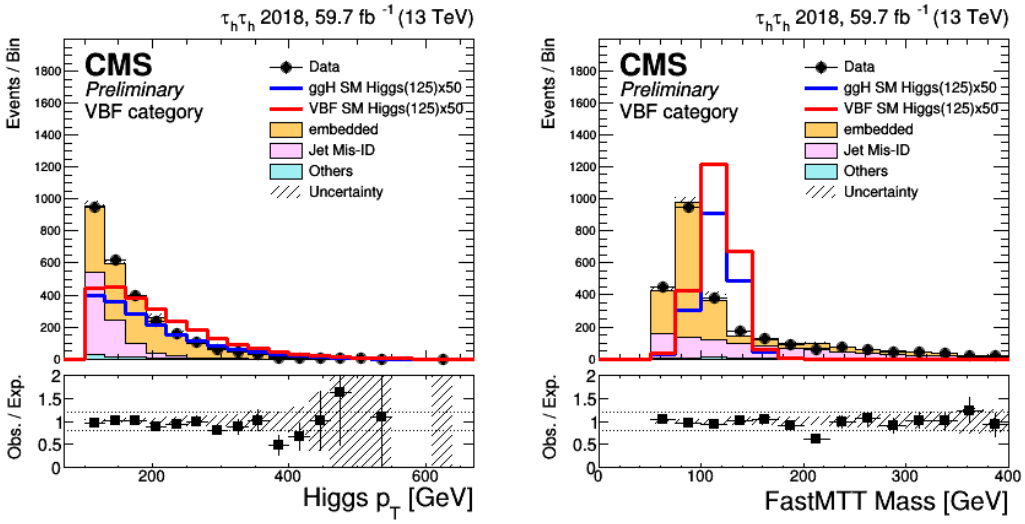

Figure 9.4: *The distribution of the p_T of the Higgs boson (left) and di-tau invariant mass (right) in the $\tau_h\tau_h$ channel, 2018. Only statistical uncertainties are included in the uncertainty band.*

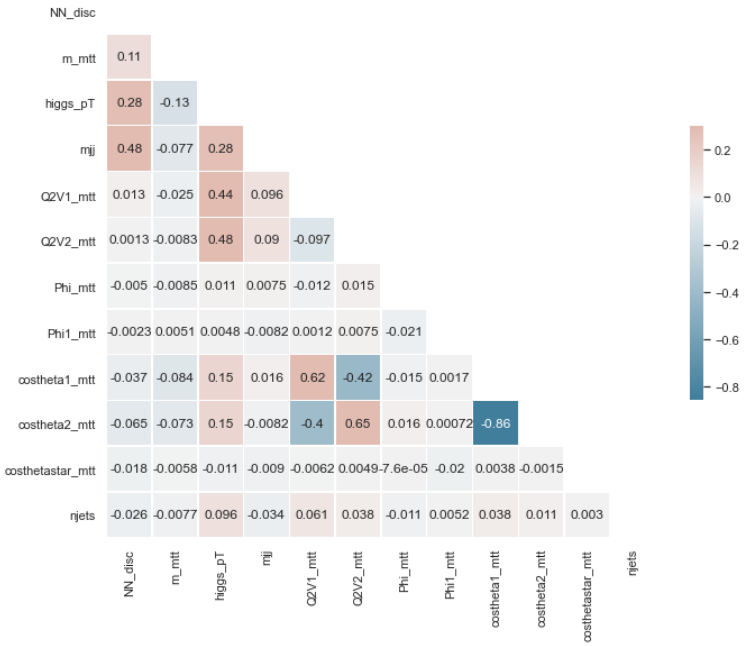

Figure 9.5: *Plots shows correlation between all neural network training input variables in the $\tau_h\tau_h$ channel, 2018.*

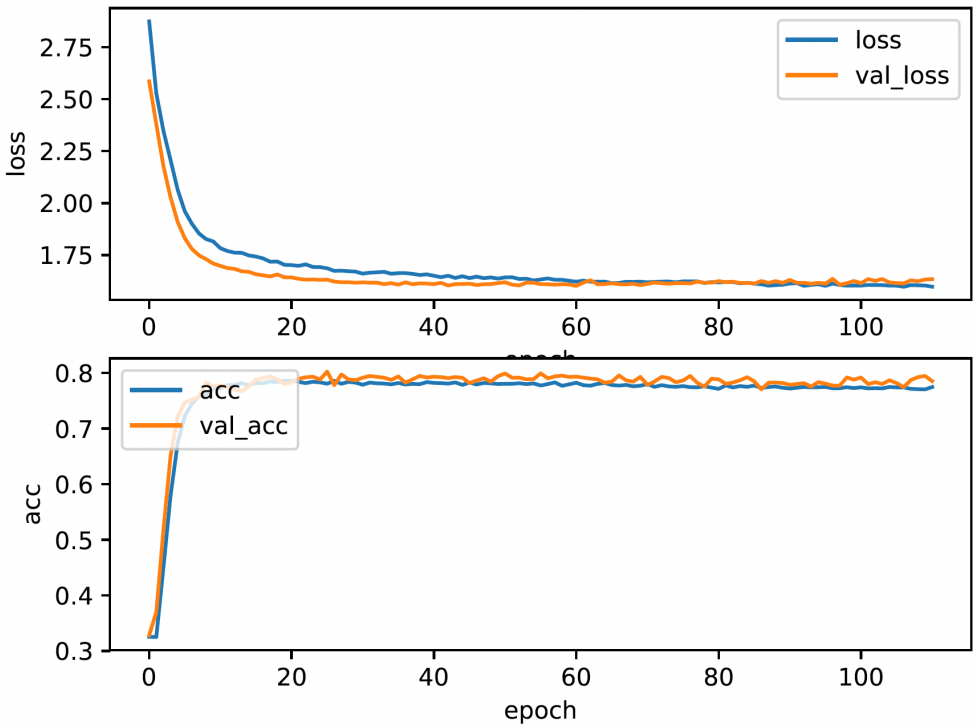

Figure 9.6: *Training and validation loss for 2018, the $\tau_h\tau_h$ channel. The training and validation loss converge to a similar value.*

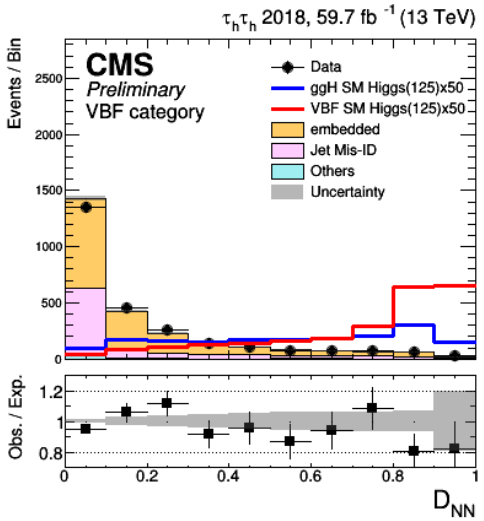

Figure 9.7: *$\mathcal{D}_{NN}$ distribution in the $\tau_h\tau_h$ channel, 2018. Only statistical uncertainties are included in the uncertainty band.*

$\tau_\mu\tau_h$ and $\tau_e\tau_h$ neural network

The $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels are combined for training a single neural network due to the fact that both channels have similar background compositions. The primary background in both channels is the decay of a Z boson to a pair of τ leptons, which has a peak in the di-tau mass spectrum overlapping with Higgs signal. As a result, a neural network designed for binary classification targeting Higgs signal vs. backgrounds with genuine di-τ (embedded).

The network architecture is identical to the network architecture in the $\tau_h\tau_h$ channel with the exception that the output layer is modified to accommodate a single class. The outputs are normalized to unity allowing them to be interpreted as probability of signal class. The normalization is done via sigmoid function. Binary cross-entropy is the obvious choice for the loss function in our network designed as a binary classifier. The $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels use the same set of input variables as the $\tau_h\tau_h$ channel.

$\tau_e\tau_\mu$ neural network

The $\tau_e\tau_\mu$ channel has a unique background composition necessitating a separate neural network. In the $\tau_e\tau_\mu u$ channel, the $Z \to \tau\tau$ and $t\bar{t}$ backgrounds are of roughly the same magnitude. Binary classifiers to separate Higgs signal from those backgrounds were tested, but resulted in sub-optimal discrimination. As a result, a multi-class neural network is used in the $\tau_e\tau_\mu$ channel. This network has an architecture identical to the network used in the $\tau_h\tau_h$ channel. In addition to the same set of input variables as the other three channels, the $\tau_e\tau_\mu$ channel uses the D_ζ, and the number of jets as input variables. The same method of scaling variables in the other channels is applied to the $\tau_e\tau_\mu$ channel.

Training the $\tau_e\tau_\mu$ channel neural network requires three samples: $H \to \tau\tau$, $Z \to \tau\tau$, and $t\bar{t}$. Conveniently, all three samples can be taken from MC simulation. All corrections are applied for each process. Then, the Higgs signal and $t\bar{t}$ samples are scaled to have an equal number of events. After this, $Z \to \tau\tau$ is scaled relative to $t\bar{t}$ to give the $Z \to \tau\tau$ to $t\bar{t}$ ratio observed in the VBF category. The final result is an appropriate balance of training events between all three classes.

9.4 Systematic uncertainties

In addition to the systematic uncertainties discussed in Chapter 6, an additional shape uncertainty is also considered to take into account differences between NLO ggH+2j event generators (comparing MadGraph5_aMC@NLO with MINLO). It is found that the inclusion of this uncertainty have a negligible impact on the final results, so the uncertainty is eventually neglected in the analysis.

For the VBF production, theory uncertainties affecting the acceptance is considered. The templates are produced with JHUGen then scaled to the yields predicted by powheg, therefor, the uncertainty on the powheg yield is taken into account. This uncertainty is estimated by varying the μ_R and μ_F scales in both the matrix element calculation and the parton shower independently by factors of 2 and 0.5, which give the largest differences with respect to the nominal to estimate the uncertainties.

The shape differences are found to be negligible and therefore the uncertainties are converted to normalization uncertainties which range between 0–4%.

Additional shape uncertainties due to the LO-only description by JHUGen compared to NLO by powheg is considered, but it is found that the inclusion of such uncertainties to have a negligible impact on the final results. Figure 9.8 shows distribution comparison between different generator and hypotheses of VBF process in the $\tau_h\tau_h$ channel in VBF category for 2018.

9.5 Signal extraction and statistical interpretation

In this section, we review the analysis methods employed to extract the HVV and Hgg anomalous coupling parameters.

For events entering the VBF category, a combination of multivariate observables introduced in Section 9.3 is used. MELA discriminants provide powerful handles to distinguish different signal hypotheses (shown in Table 9.2), while simple neural networks offer optimal separation of the dominant backgrounds for a given channel from the Higgs production.

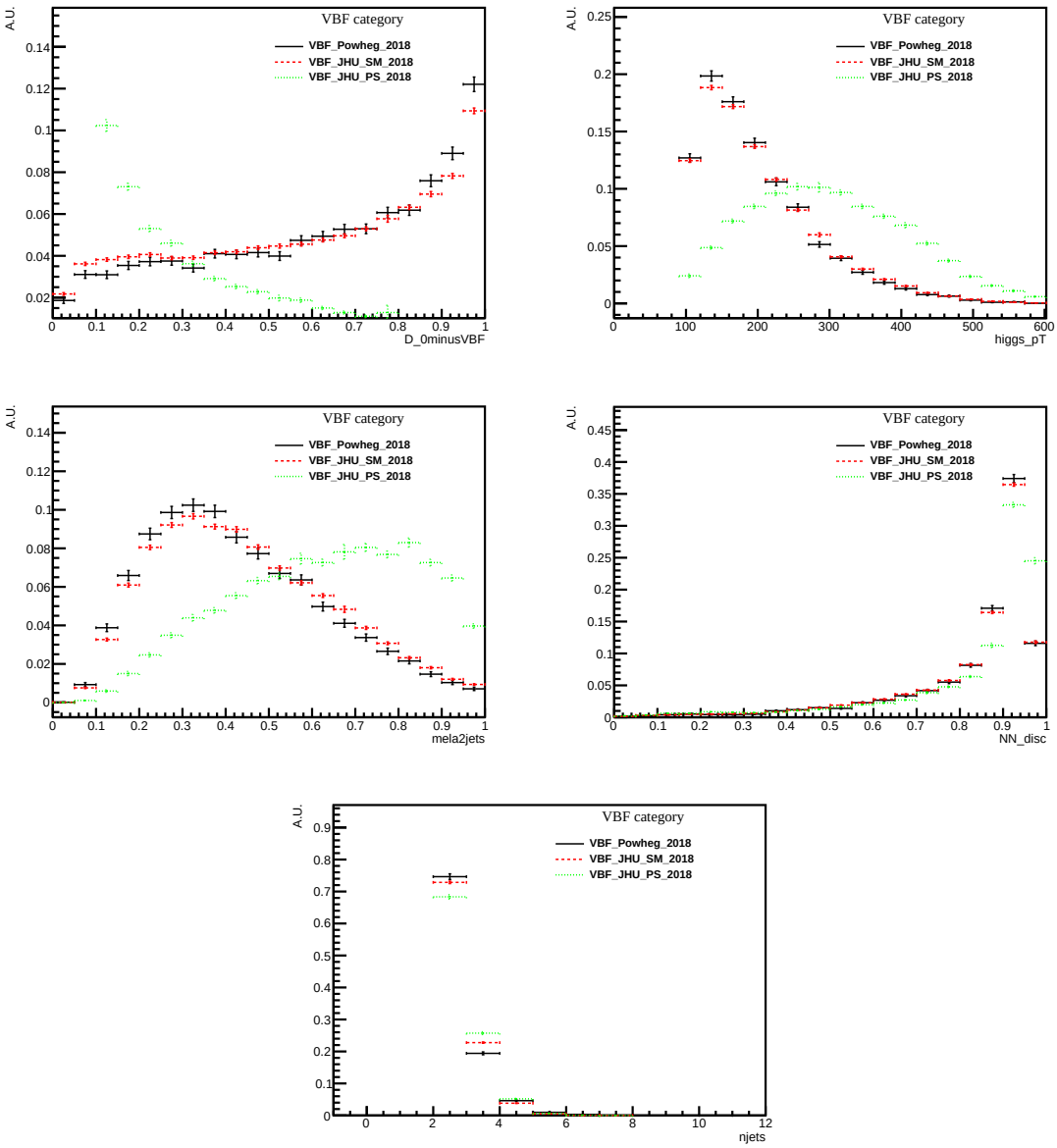

Figure 9.8: *Distribution comparison between different generator and hypotheses (*POWHEG *SM (black),* JHUGEN *SM (red), and* JHUGEN *PS (green)) of VBF process in the $\tau_h\tau_h$ channel in VBF category for 2018.*

Although VBF signal is used as the signal process for all neural networks, it can be the observable in the analyses of both ggH and VBF production processes because sensitivity to the Hgg anomalous coupling is maximal for events with kinematics similar to those of VBF production. The $\mathcal{D}_{2jet}^{VBF}$ defined in Equation 9.3 is also employed to separate ggH productions from the SM VBF production in every analyses. The MELA discriminants listed in Table 9.2 constructed to optimally separate the SM hypothesis from the potential anomalous couplings using Equation 9.1 as follow

$$\begin{aligned} \mathcal{D}_{0-}^{\mathrm{ggH}} = \frac{\mathcal{P}_{\mathrm{SM}}^{\mathrm{ggH}}}{\mathcal{P}_{\mathrm{SM}}^{\mathrm{ggH}} + \mathcal{P}_{0-}^{\mathrm{ggH}}}, \qquad \mathcal{D}_{0-} = \frac{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}}}{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}} + \mathcal{P}_{0-}^{\mathrm{VBF}}}, \qquad \mathcal{D}_{0h+} = \frac{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}}}{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}} + \mathcal{P}_{\mathrm{a2}}}, \\ \mathcal{D}_{\Lambda 1} = \frac{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}}}{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}} + \mathcal{P}_{\Lambda 1}}, \qquad \mathcal{D}_{\Lambda 1}^{\mathrm{Z}\gamma} = \frac{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}}}{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}} + \mathcal{P}_{\Lambda 1}^{\mathrm{Z}\gamma}} \end{aligned} \tag{9.4}$$

for each anomalous couplings parameter. Two pure CP-odd MELA discriminants are defined based on Equation 9.2, which are sensitive to interference effects between the SM and pseudoscalar Higgs contributions to directly probe for CP-violation in the HVV and Hgg vertices respectively. They are given as follows

$$\mathcal{D}_{CP}^{\mathrm{VBF}} = \frac{\mathcal{P}_{\mathrm{SM}-0-}^{\mathrm{VBF}}}{\mathcal{P}_{\mathrm{SM}}^{\mathrm{VBF}} + \mathcal{P}_{0-}^{\mathrm{VBF}}}, \qquad \mathcal{D}_{CP}^{\mathrm{ggH}} = \frac{\mathcal{P}_{\mathrm{SM}-0-}^{\mathrm{ggH}}}{\mathcal{P}_{\mathrm{SM}}^{\mathrm{ggH}} + \mathcal{P}_{0-}^{\mathrm{ggH}}} \tag{9.5}$$

and are used in measurements of CP-violating parameters for VBF and ggH productions.

$m_{\tau\tau}$ is chosen as an observable in both 0-jet and boosted categories, and $p_T^{\tau\tau}$ is also selected as another observable to construct templates in the boosted category. There are no dedicated MELA observables sensitive to anomalous couplings in these categories because the events belonging to these categories have less than the two jets, or do not display significant separation between different signal scenarios. Still, $p_T^{\tau\tau}$ brings some sensitivity to anomalous HVV couplings in boosted category because the BSM VBF events typically have larger $p_T^{\tau\tau}$. The 0-jet and boosted categories are important to include in the fit, as they constraint backgrounds in spite of not bringing much sensitivity.

Example distributions of the observables in the most sensitive $\tau_h\tau_h$ and $\tau_\mu\tau_h$ channels are given in Fig. 9.9.

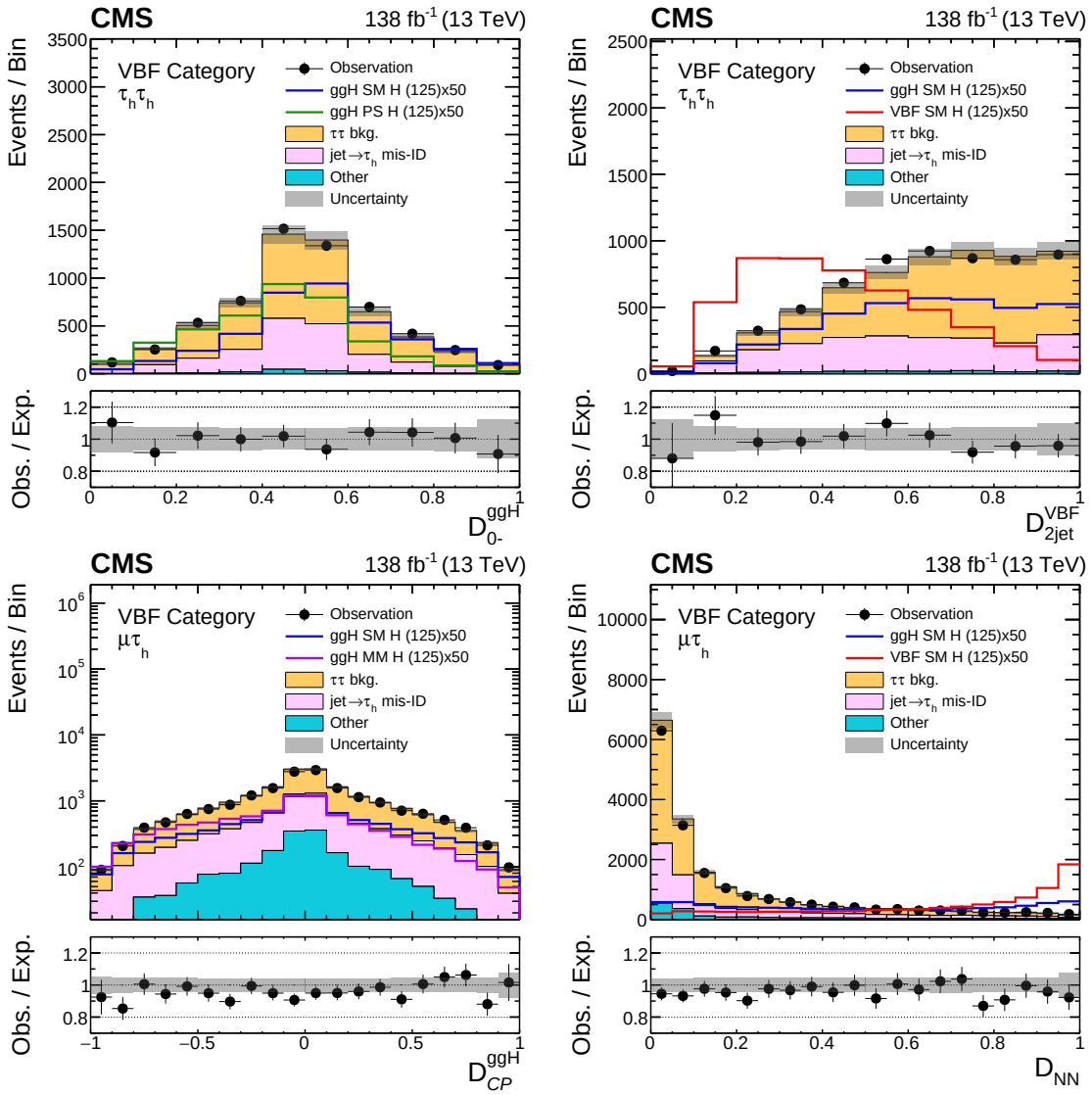

Figure 9.9: *Examples of data and signal and background predictions for* MELA *and neural network discriminants in the $\tau_h\tau_h$ ($\tau_h\tau_h$) and $\tau_\mu\tau_h$ ($\mu\tau_h$) channels. Statistical uncertainties and systematic uncertainties that affect the normalization of the background distribution are included in the uncertainty band. The expectation in the ratio panel is the sum of the estimated the SM Higgs signal and backgrounds. For the $\mathcal{D}_{0-}^{\mathrm{ggH}}$ discriminant the distribution expected for a pseudoscalar Higgs hypothesis (labeled "PS" in the legend) is overlaid to be compared to the SM signal. Similarly, for the $\mathcal{D}_{CP}^{\mathrm{ggH}}$ discriminant the distribution for a CP-violating scenario with the maximum-mixing between CP-even and CP-odd couplings (labeled "MM" in the legend) is shown.*[6]

Category	Observable	Goal
0-jet	$m_{\tau\tau}$	Separate Higgs signal from backgrounds
Boosted	$p_T^{\tau\tau}$, $m_{\tau\tau}$	Separate Higgs signal from backgrounds and BSM from SM HVV
VBF	$\mathcal{D}_{NN}$	Separate VBF-like Higgs signal from backgrounds
VBF	$\mathcal{D}_{2jet}^{\mathrm{VBF}}$	Separate ggH from VBF Higgs production
VBF	$\mathcal{D}_{0-}^{\mathrm{ggH}}(\mathcal{D}_{0-})$	Separate BSM from SM Hgg (HVV)
VBF	$\mathcal{D}_{CP}^{\mathrm{ggH}}(\mathcal{D}_{CP}^{\mathrm{VBF}})$	Sensitive to the interference between the CP-even and CP-odd contributions to the Hgg (HVV) coupling

Table 9.3: *List of observables used in each category.*

The results are extracted with a global maximum likelihood fit based on 4D or 3D distributions built in the VBF category, 2D distributions in the boosted category, and 1D distribution in 0-jet category. The observables utilized in the analysis are summarized in Table 9.3.

A binned maximum likelihood fit to the data is performed to extract results combining all categories for the different channels and data taking years. The likelihood function is defined as a product of conditional probabilities over all bins i:

$$\mathcal{L}(\mathrm{data}|\mu_{\mathrm{ggH}}, \mu_{\mathrm{qqH}}, \vec{f}, \theta) = \prod_i \mathrm{Poisson}(\mathrm{n_i}|s_i(\mu_{\mathrm{ggH}}, \mu_{\mathrm{qqH}}, \vec{f}, \theta) + b_i(\theta)) \cdot p(\tilde{\theta}|\theta), \tag{9.6}$$

where n_i is the observed number of data events in each bin. s_i and b_i are the signal and background expectations, which are functions of the full set of nuisance parameters corresponding to the systematic uncertainties (θ). Three parameters, μ_{ggH}, μ_{qqH}, and $\vec{f}$, modify the Higgs boson signal processes. Two Higgs boson signal strength modifiers μ_{ggH} and μ_{qqH} modify the ggH and VBF+VH cross sections with respect to the SM values respectively. The set of anomalous coupling parameters are defined as $\vec{f}$ and the parameters modify the distributions of the ggH and/or VBF+VH signals. For example, in the case of ggH nomalous coupling measurements $\vec{f} = (f_{a3}^{ggH}, f_{a3})$. For the VBF anomalous coupling measurements $\vec{f}$ is chosen depends on CP property of the parameter being measured. For f_{a3} measurment $\vec{f} = (f_{a3}, f_{a3}^{ggH})$, for all other CP-even parameter (f_{a2}, $f_{\Lambda 1}$, $f_{\Lambda 1}^{Z\gamma}$) measurments, $\vec{f} = (f_i)$. Lastly, the $p(\tilde{\theta}|\theta)$ term represents the entire set of probability density functions of the uncer-

tainties on the nominal values of the nuisance parameters $\tilde{\theta}$. The systematic uncertainties that modify only the normalizations are assigned log-normal pdfs, while the shape altering systematics are assigned Gaussian pdfs.

The negative log-likelihood is defined as

$$-2\Delta \ln \mathcal{L} = -2\Delta \ln \frac{\mathcal{L}(\text{data}|\mu_{\text{ggH}}, \mu_{\text{qqH}}, \vec{f}, \theta)}{\mathcal{L}(\text{data}|\hat{\mu}_{\text{ggH}}, \hat{\mu}_{\text{qqH}}, \vec{\hat{f}}, \hat{\theta})}, \tag{9.7}$$

with $\hat{\mu}_{\text{ggH}}$, $\hat{\mu}_{\text{qqH}}$, $\vec{\hat{f}}$, and $\hat{\theta}$ as the best-fit values of the signal modifiers and nuisance parameters. The 68 % and 95 % confidence intervals are corresponding to $-2\Delta \ln \mathcal{L} = 1.00$, and 3.84 respectively.

9.6 Results

9.6.1 Results of the HVV analysis

The four f_{ai} parameters describing anomalous HVV couplings, as defined in Equation 2.2, are tested against the data by maximizing the likelihood function defined in Equation 9.6. As it is described in Section 2.3.1, the event kinematics of the Higgs production in WW fusion and in ZZ fusion are very similar, and it is practically impossible to distinguish between a_i^{WW} and a_i^{ZZ} in the VBF production. In our measurements, two approaches are adopted to set the relationship between a_i^{WW} and a_i^{ZZ} couplings. In the first approach (Approach 1), they are analyzed together assuming $a_i^{WW} = a_i^{ZZ}$ and $\kappa_i^{ZZ}/(\lambda_1^{ZZ})^2 = \kappa_i^{WW}/(\lambda_1^{WW})^2$. In the second approach (Approach 2), the results for the CP-violating parameter a_3 is reinterpreted[65], applying additional considerations of custodial and SU(2)×U(1) symmetries in the relationships of anomalous couplings. With $a_3^{\gamma\gamma}$ and $a_3^{Z\gamma}$ set to zero, only a simple relationship between a_3^{WW} and a_3^{ZZ} is left as following:

$$a_3^{WW} = cos^2\theta_W a_3^{ZZ} \ . \tag{9.8}$$

An example of a pre-fit distributions in the most sensitive VBF category for the $\tau_h\tau_h$ channel is shown in Figure 9.10. The results of the likelihood scans using two approaches are shown in Figures 9.11 and 9.12, respectively, and listed Table 9.4. In the CP-odd parameter fit, f_{a3}^{ggH} left unconstrained, while the values of the other anomalous coupling parameters are set to zero in CP-even parameter fits. The signal strength parameters μ_{qqH} and μ_{ggH} are profiled in the fits, whose best-fit values are consistent with unity.

Approach	Parameter	Observed/(10^{-3})		Expected/(10^{-3})	
		68% CL	95% CL	68% CL	95% CL
Approach 1	f_{a3}	$0.26^{+0.38}_{-0.21}$	$[-0.01, 1.30]$	0.00 ± 0.06	$[-0.23, 0.23]$
	f_{a2}	$1.1^{+0.9}_{-0.9} \cup [-1.8, -0.1]$	$[-3.4, 3.2]$	$0.0^{+0.6}_{-0.5}$	$[-1.4, 1.5]$
	f_{Λ}	$-0.12^{+0.08}_{-0.10}$	$[-0.34, 0.01]$	$0.00^{+0.19}_{-0.05}$	$[-0.15, 0.55]$
	$f_{\Lambda}^{Z\gamma}$	2.5 ± 1.8	$[-3.6, 6.5]$	$0.0^{+1.5}_{-1.2}$	$[-3.2, 3.4]$
Approach 2	f_{a3}	$0.40^{+0.53}_{-0.33}$	$[-0.01, 1.90]$	0.00 ± 0.08	$[-0.33, 0.33]$

Table 9.4: *Allowed 68% (central values with uncertainties) and 95% CL (in square brackets) intervals on anomalous HVV coupling parameters. Approaches 1 and 2 refer to the choice of the relationship between the a_i^{WW} and a_i^{ZZ} couplings.*[6]

Two observed minima in the likelihood scan for f_{a2} (and to a lesser extent $f_{\Lambda 1}^{Z\gamma}$) occur due to the limited sensitivity to the sign of the interference between a_1 and a_2 ($\kappa_2^{Z\gamma}$) couplings, therefore, the sensitivity to the signs of the f_{a2} ($f_{\Lambda 1}^{Z\gamma}$) parameters are limited.

The VBF and VH production processes are sensitive to higher values of q_i^2 in Equation 2.1, which means the cross section of anomalous contributions increases rapidly with f_{ai}. The increase rates are different for production and decay, and relatively small values of f_{ai} correspond to a major anomalous contribution to the production cross section, which explains the plateau in the $-2\Delta\ln\mathcal{L}$ distributions for large values of f_{ai}.

9.6.2 Results of the ggH analysis

An example of a pre-fit distribution for the measurements of f_{a3}^{ggH} is given in Figure 9.13. The results of the likelihood scans are shown in Figure 9.14. Observed f_{a3}^{ggH} is found to be $0.08^{+0.35}_{-0.08}$ at 68% CL. Noticeably large $-2\Delta\ln\mathcal{L}$ for the observed scan occurred at

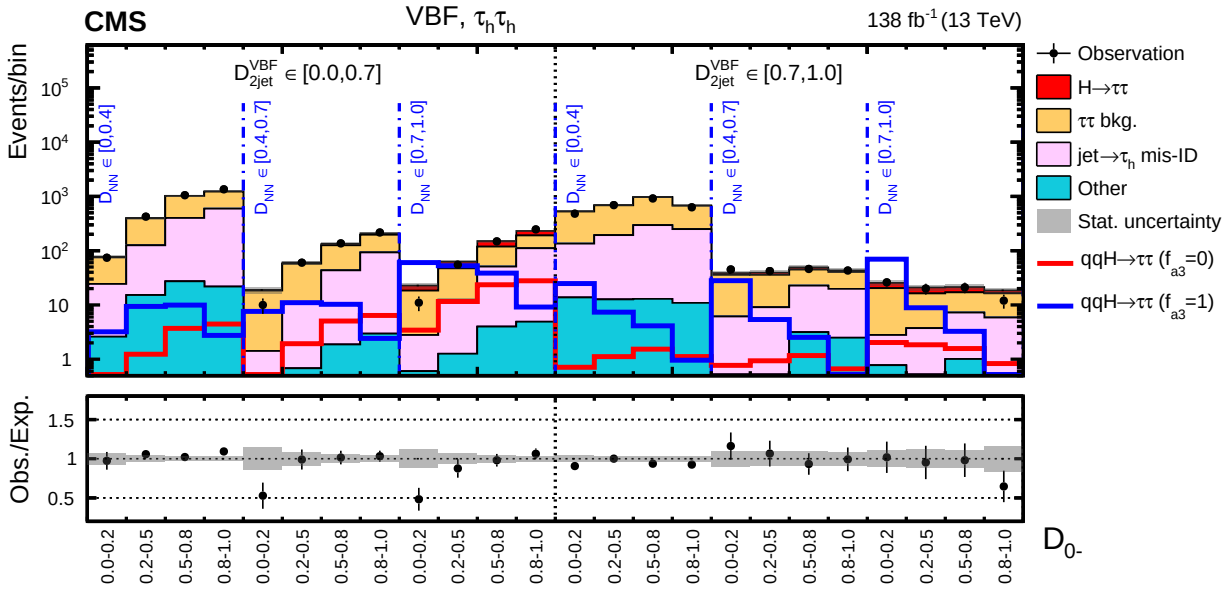

Figure 9.10: *The observed and predicted 2D distribution of ($\mathcal{D}_{0-}$, $\mathcal{D}_{NN}$, $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$) before the fit to data in the most sensitive VBF category in the $\tau_h\tau_h$ channel. The total Higgs signal is shown stacked on top of the background in the solid red histogram. The VBF+VH signal for the CP-even (CP-odd) scenario is shown by the red (blue) line. The uncertainty band include only the statistical uncertainties.*[6]

$f_{a3}^{ggH} \approx -0.8$ are affected by statistical fluctuations, but still compatible with the signal plus background model. It is proven by pseudo-experiments estimating the probability of obtaining a maximum value of the $-2\Delta \ln \mathcal{L}$ of the observed scan, which was determined to be 33 %.

In the case of ggH production, the quark loop generates the SM process, which is represented by the a_2^{gg} term in Equation 2.1. Therefore, it is more difficult to distinguish the anomalous a_3^{gg} contribution from the SM contribution because they are both generated by dimension-six operators. In addition, many of their kinematics has similar features. Most of the sensitivity to CP-odd couplings is predominantly in the azimuthal angle between two associated jets. This explains why the sensitivity is much worse compare to the case of VBF production, which could gain full information from the multivariate MELA treatment. For the VBF production, the SM process is generated by the tree-level coupling which is a dimension-four operator. This results in kinematic differences between the SM and the anomalous contributions in multiple observables.

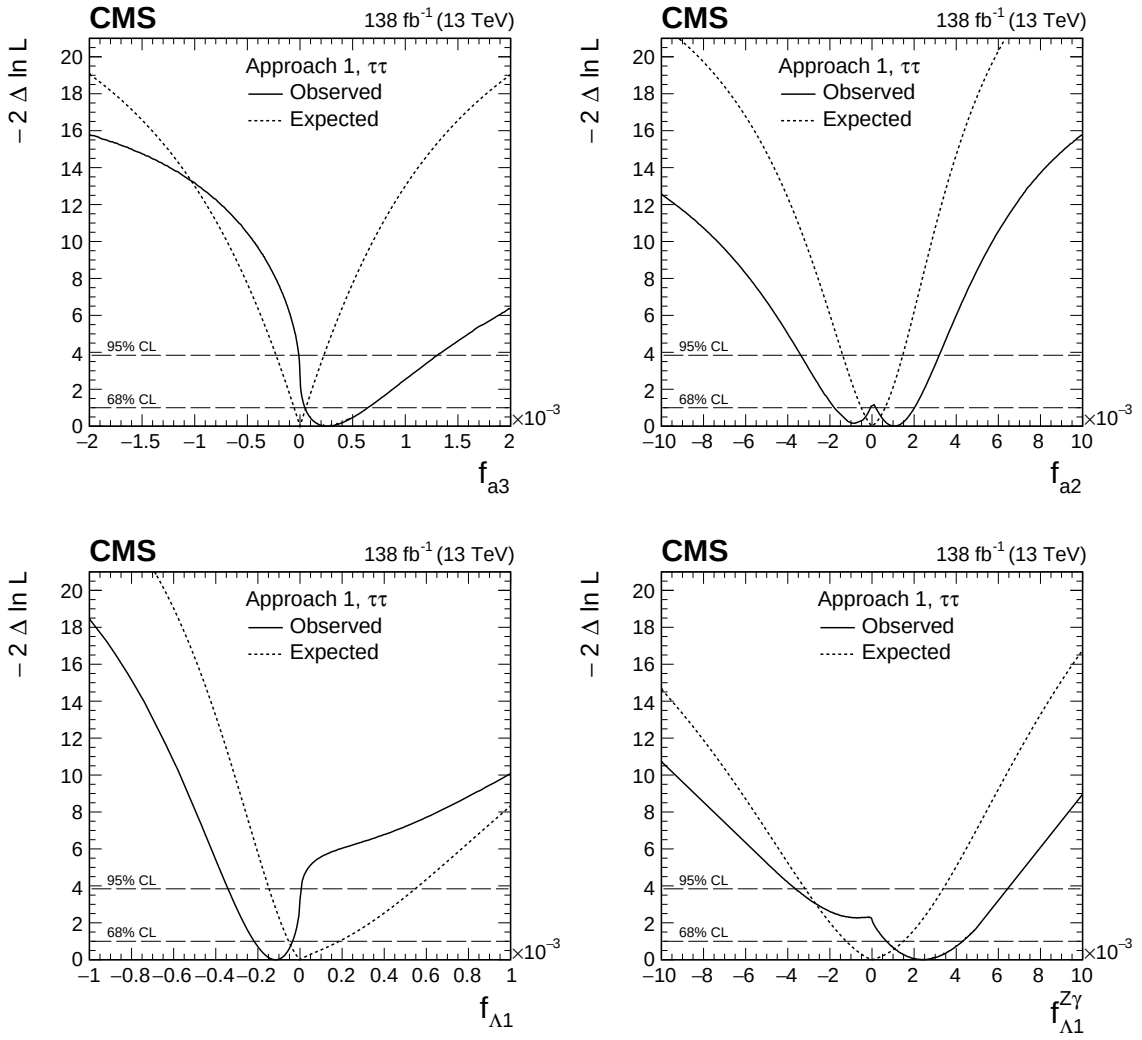

Figure 9.11: *Observed (solid) and expected (dashed) likelihood scans of f_{a3} (upper left), f_{a2} (upper right), $f_{\Lambda 1}$ (lower left), and $f_{\Lambda 1}^{Z\gamma}$ (lower right) in Approach 1 ($a_i^{WW} = a_i^{ZZ}$).*[6]

9.7 Chapter summary

A study is presented of anomalous interaction of the Higgs boson with vector bosons, including CP-violation, using its associated production with two hadronic jets in ggH, VBF and VH productions and subsequent decay to a pair of τ leptons. Constraints have been set on multiple the CP-violating and CP-conserving parameters in terms of the effective cross section ratio f_{a3}^{ggH} using multivariate techniques. In the VBF production analysis, constraints on the CP-violating parameter f_{a3} and the CP-conserving parameters f_{a2}, $f_{\Lambda 1}$ and $f_{\Lambda 1}^{Z\gamma}$ have been set. In the ggH production analysis, constraints on the CP-violating effects in ggH

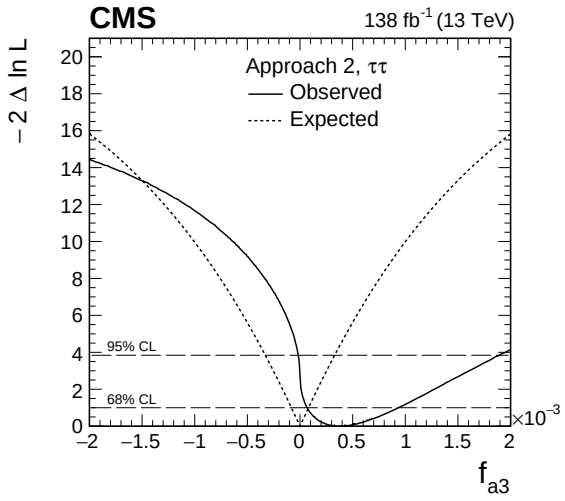

Figure 9.12: *Observed (solid) and expected (dashed) likelihood scans of f_{a3} in Approach 2.*[6]

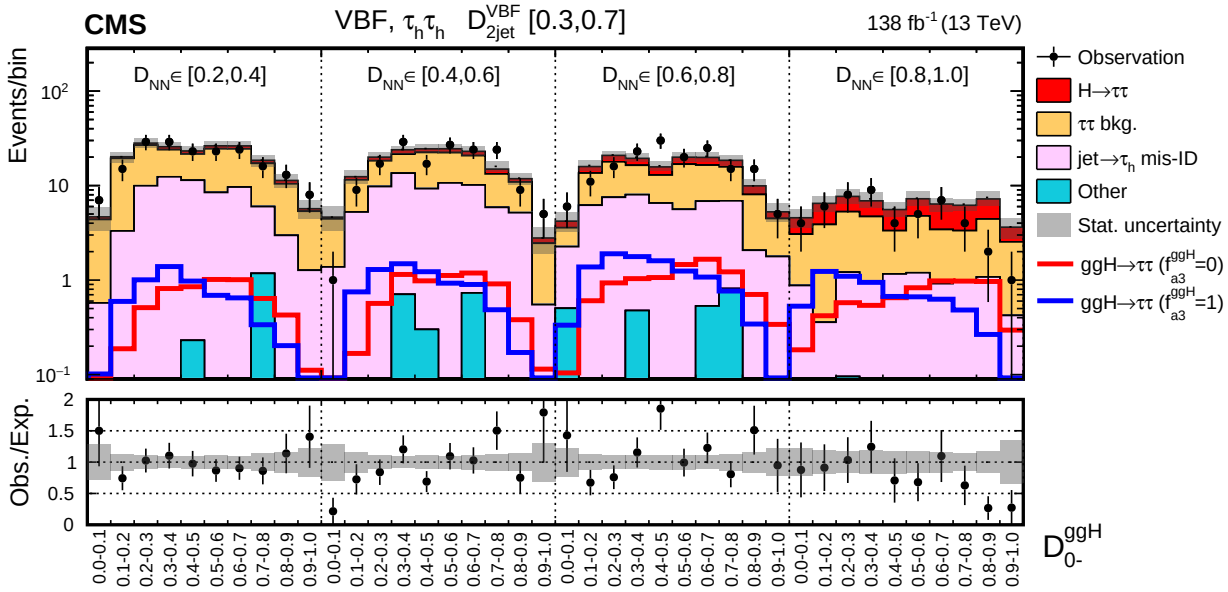

Figure 9.13: *The observed and predicted 2D distribution of ($\mathcal{D}_{0-}^{ggH}$, $\mathcal{D}_{NN}$) before the fit to data in the most sensitive VBF category region with $0.3 < \mathcal{D}_{\text{2jet}}^{\text{VBF}} < 0.7$ in the $\tau_h\tau_h$ channel. The total Higgs signal is shown stacked on top of the background in the solid red histogram. The ggH signal for the CP-even (CP-odd) scenario is shown by the red (blue) line. The uncertainty band include only the statistical uncertainties.*[6]

production have been set. This results set the most stringent limits on CP-violation in ggH production to date, excluding the pur CP-odd scenario of the Higgs coupling to gluons with

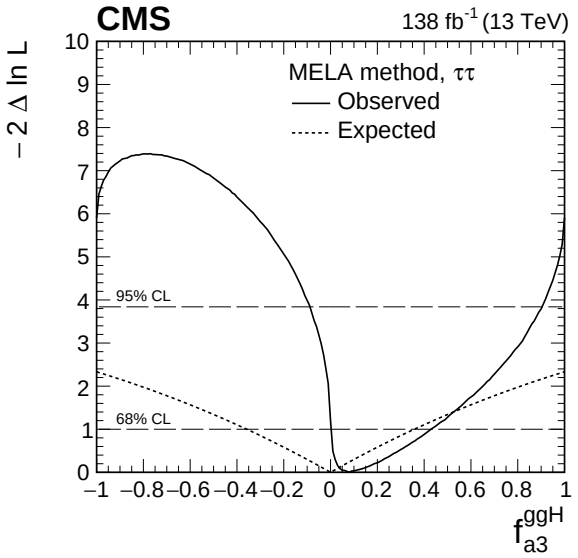

Figure 9.14: *Observed (solid) and expected (dashed) likelihood scans of* f_{a3}^{ggH} *obtained.*[6]

a significance of 2.4 standard deviations.

Chapter 10

Summary

I have presented a study of the Higgs boson properties using its decay to a pair of τ leptons using 138 fb^{-1} data collected at $\sqrt{s} = 13$ TeV by the CMS experiment.

The properties of the Higgs boson were studied by two complementary approaches to measure its cross section. Firstly, the cross sections of different Higgs productions were explored within each production mode using the Simplified Template Cross Section (STXS) framework. Secondly, the inclusive and differential fiducial cross sections for thie Higgs boson production decaying to two τ leptons were presented. In the third analysis, a search for anomalous couplings in the HVV and Hgg vertices is performed utilizing the correlation between jets produced in association with the Higgs boson.

Studying the nature of the properties of the Higgs boson is essential to completing our understanding of fundamental physics. All results are consistent with the SM prediction setting stringent limits on anomalous Higgs boson couplings to data and measuring precise cross sections in smaller phase spaces.

Bibliography

[1] Standard Model of Particle Physics [The Physics Travel Guide]. https://physicstravelguide.com/models/standard_model, Jun 2021. Accessed: 2022-03-29.

[2] David Barney. CMS Detector Slice. CMS Collection., Jan 2016. URL http://cds.cern.ch/record/2120661.

[3] Measurement of Higgs boson production in the decay channel with a pair of τ leptons. Technical report, CERN, Geneva, 2020. URL https://cds.cern.ch/record/2725590.

[4] A. Tumasyan, W. Adam, J. W. Andrejkovic, and et al. Measurement of the inclusive and differential higgs boson production cross sections in the decay mode to a pair of of τ leptons in pp collisions at $\sqrt{s} = 13$. *Physical Review Letters*, 128(8), feb 2022. doi: 10.1103/physrevlett.128.081805. URL https://doi.org/10.1103%2Fphysrevlett.128.081805.

[5] V. Khachatryan, A.M. Sirunyan, and et al. Constraints on the spin-parity and anomalousHVVcouplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Physical Review D*, 92(1), Jul 2015. ISSN 1550-2368. doi: 10.1103/physrevd.92.012004. URL http://dx.doi.org/10.1103/PhysRevD.92.012004.

[6] Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production with associated particles using the H $\rightarrow \tau\tau$ final state. Technical report, CERN, Geneva, 2021. URL https://cds.cern.ch/record/2784578.

[7] G. Cowan. *Statistical data analysis*. 1998. ISBN 978-0-19-850156-5.

[8] S Heinemeyer, editor. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. CERN Yellow Reports: Monographs. Jul 2013. doi: 10.5170/CERN-2013-004. URL http://cds.cern.ch/record/

1559921. Comments: 404 pages, 139 figures, to be submitted to CERN Report. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.

[9] D. de Florian, C. Grojean, and et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. CERN, Geneva, Oct 2016. doi: 10.23731/CYRM-2017-002. URL 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.

[10] A.M. Sirunyan, A. Tumasyan, and et al. Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector. *Physics Letters B*, 779:283–316, Apr 2018. ISSN 0370-2693. doi: 10.1016/j.physletb.2018.02.004. URL http://dx.doi.org/10.1016/j.physletb.2018.02.004.

[11] M. Aaboud, G. Aad, and et al. Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS detector. *Physical Review D*, 99(7), Apr 2019. ISSN 2470-0029. doi: 10.1103/physrevd.99.072001. URL http://dx.doi.org/10.1103/PhysRevD.99.072001.

[12] G. Aad, B. Abbott, and et al. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *The European Physical Journal C*, 75(10), Oct 2015. ISSN 1434-6052. doi: 10.1140/epjc/s10052-015-3685-1. URL http://dx.doi.org/10.1140/epjc/s10052-015-3685-1.

[13] Ian Anderson, Sara Bolognesi, and et al. Constraining anomaloushvvinteractions at proton and lepton colliders. *Physical Review D*, 89(3), Feb 2014. ISSN 1550-2368. doi: 10.1103/physrevd.89.035007. URL http://dx.doi.org/10.1103/PhysRevD.89.035007.

[14] Andrei V. Gritsan, Jeffrey Roskes, and et al. New features in the jhu generator framework: Constraining higgs boson properties from on-shell and off-shell production. *Phys-*

ical Review D, 102(5), Sep 2020. ISSN 2470-0029. doi: 10.1103/physrevd.102.056022. URL http://dx.doi.org/10.1103/PhysRevD.102.056022.

[15] Maximilien Brice. Aerial views: CERN. https://home.cern/news/news/cern/25-years-large-hadron-collider-experimental-programme, accessed 2021-12-01.

[16] CMS Collaboration. Public CMS luminosity information. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults?rev=162, accessed 2021-12-05.

[17] Tai Sakuma and Thomas McCauley. Detector and Event Visualization with SketchUp at the CMS Experiment. *Journal of Physics: Conference Series*, 513(2):022032, Jun 2014. ISSN 1742-6596. doi: 10.1088/1742-6596/513/2/022032. URL http://dx.doi.org/10.1088/1742-6596/513/2/022032.

[18] CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03):T03021–T03021, Mar 2010. ISSN 1748-0221. doi: 10.1088/1748-0221/5/03/t03021. URL http://dx.doi.org/10.1088/1748-0221/5/03/T03021.

[19] A Dominguez, Abbaneo, and et al. CMS Technical Design Report for the Pixel Detector Upgrade. Technical report, Sep 2012. URL https://cds.cern.ch/record/1481838. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

[20] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, Geneva, Aug 2011. URL http://cds.cern.ch/record/1379837.

[21] J. S. Conway. Incorporating nuisance parameters in likelihoods for multisource spectra, 2011.

[22] Alexander L. Read. Linear interpolation of histograms. *Nucl. Instrum. Meth. A*, 425: 357–360, 1999. doi: 10.1016/S0168-9002(98)01347-3.

[23] Roger Barlow and Christine Beeston. Fitting using finite monte carlo samples. *Computer Physics Communications*, 77(2):219–228, 1993. ISSN 0010-4655. doi: https://doi.org/10.1016/0010-4655(93)90005-W. URL https://www.sciencedirect.com/science/article/pii/001046559390005W.

[24] A L Read. Presentation of search results: theCLstechnique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693–2704, sep 2002. doi: 10.1088/0954-3899/28/10/313. URL https://doi.org/10.1088/0954-3899/28/10/313.

[25] Robert D. Cousins. Generalization of chisquare goodness-offit test for binned data using saturated models , with application to histograms. 2013.

[26] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009–P10009, Oct 2014. ISSN 1748-0221. doi: 10.1088/1748-0221/9/10/p10009. URL http://dx.doi.org/10.1088/1748-0221/9/10/P10009.

[27] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987. ISSN 0168-9002. doi: https://doi.org/10.1016/0168-9002(87)90887-4. URL https://www.sciencedirect.com/science/article/pii/0168900287908874.

[28] Measurement of Tracking Efficiency. Technical report, CERN, Geneva, 2010. URL https://cds.cern.ch/record/1279139.

[29] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, 1998. doi: 10.1109/5.726788.

[30] Wolfgang Waltenberger, Rudolf Frühwirth, and Pascal Vanlaer. Adaptive vertex fitting. *Journal of Physics G: Nuclear and Particle Physics*, 34(12):N343–N356, nov 2007. doi:

10.1088/0954-3899/34/12/n01. URL https://doi.org/10.1088/0954-3899/34/12/n01.

[31] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008. ISSN 1029-8479. doi: 10.1088/1126-6708/2008/04/063. URL http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[32] A.M. Sirunyan, A. Tumasyan, and et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(05):P05011–P05011, may 2018. doi: 10.1088/1748-0221/13/05/p05011. URL https://doi.org/10.1088/1748-0221/13/05/p05011.

[33] A.M. Sirunyan, A. Tumasyan, and et al. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, Oct 2017. ISSN 1748-0221. doi: 10.1088/1748-0221/12/10/p10003. URL http://dx.doi.org/10.1088/1748-0221/12/10/P10003.

[34] A.M. Sirunyan, A. Tumasyan, and et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV. *Journal of Instrumentation*, 13(06):P06015–P06015, Jun 2018. ISSN 1748-0221. doi: 10.1088/1748-0221/13/06/p06015. URL http://dx.doi.org/10.1088/1748-0221/13/06/P06015.

[35] S Chatrchyan, Hmayakyan, and et al. The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment. *JINST*, 3:S08004. 361 p, 2008. doi: 10.1088/1748-0221/3/08/S08004. URL https://cds.cern.ch/record/1129810. Also published by CERN Geneva in 2010.

[36] A.M. Sirunyan, A. Tumasyan, and et al. Performance of the cms muon detector and muon reconstruction with proton-proton collisions at s=13 tev. *Journal of Instrumentation*, 13(06):P06015–P06015, Jun 2018. ISSN 1748-0221. doi: 10.1088/1748-0221/13/06/p06015. URL http://dx.doi.org/10.1088/1748-0221/13/06/P06015.

[37] *Journal of Instrumentation*, 10(06):P06005–P06005, Jun 2015. ISSN 1748-0221. doi: 10.1088/1748-0221/10/06/p06005. URL http://dx.doi.org/10.1088/1748-0221/10/06/P06005.

[38] Gavin P. Salam. Towards jetography. *The European Physical Journal C*, 67(3–4): 637–686, May 2010. ISSN 1434-6052. doi: 10.1140/epjc/s10052-010-1314-6. URL http://dx.doi.org/10.1140/epjc/s10052-010-1314-6.

[39] A.M. Sirunyan, A. Tumasyan, and et al. Pileup mitigation at cms in 13 tev data. *Journal of Instrumentation*, 15(09):P09018–P09018, Sep 2020. ISSN 1748-0221. doi: 10.1088/1748-0221/15/09/p09018. URL http://dx.doi.org/10.1088/1748-0221/15/09/P09018.

[40] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002–P11002, Nov 2011. ISSN 1748-0221. doi: 10.1088/1748-0221/6/11/p11002. URL http://dx.doi.org/10.1088/1748-0221/6/11/P11002.

[41] The CMS collaboration. Identification of b-quark jets with the cms experiment. *Journal of Instrumentation*, 8(04):P04013–P04013, Apr 2013. ISSN 1748-0221. doi: 10.1088/1748-0221/8/04/p04013. URL http://dx.doi.org/10.1088/1748-0221/8/04/P04013.

[42] Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector. Technical report, CERN, Geneva, 2016. URL https://cds.cern.ch/record/2205284.

[43] M Tanabashi, K Hagiwara, and et al. Review of Particle Physics, 2018-2019. *Phys. Rev. D*, 98:030001. 1898 p, 2018. doi: 10.1103/PhysRevD.98.030001. URL https://cds.cern.ch/record/2636832.

[44] A.M. Sirunyan, A. Tumasyan, and et al. Performance of reconstruction and identification of leptons decaying to hadrons and v in pp collisions at s=13 tev. *Journal of Instru-*

mentation, 13(10):P10005–P10005, Oct 2018. ISSN 1748-0221. doi: 10.1088/1748-0221/13/10/p10005. URL http://dx.doi.org/10.1088/1748-0221/13/10/P10005.

[45] Andy Buckley, Jonathan Butterworth, and et al. General-purpose event generators for lhc physics. *Physics Reports*, 504(5):145–233, Jul 2011. ISSN 0370-1573. doi: 10.1016/j.physrep.2011.03.005. URL http://dx.doi.org/10.1016/j.physrep.2011.03.005.

[46] Keith Hamilton, Paolo Nason, , and et al. NNLOPS simulation of Higgs boson production. *Journal of High Energy Physics*, 2013(10), Oct 2013. ISSN 1029-8479. doi: 10.1007/jhep10(2013)222. URL http://dx.doi.org/10.1007/JHEP10(2013)222.

[47] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5: going beyond. *Journal of High Energy Physics*, 2011(6), Jun 2011. ISSN 1029-8479. doi: 10.1007/jhep06(2011)128. URL http://dx.doi.org/10.1007/JHEP06(2011)128.

[48] J. Alwall, S. Höche, F. Krauss, and et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *The European Physical Journal C*, 53(3):473–500, Dec 2007. ISSN 1434-6052. doi: 10.1140/epjc/s10052-007-0490-5. URL http://dx.doi.org/10.1140/epjc/s10052-007-0490-5.

[49] Rikkert Frederix and Stefano Frixione. Merging meets matching in mc@nlo. *Journal of High Energy Physics*, 2012(12), Dec 2012. ISSN 1029-8479. doi: 10.1007/jhep12(2012)061. URL http://dx.doi.org/10.1007/JHEP12(2012)061.

[50] Torbjörn Sjöstrand, Stefan Ask, and et al. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015. ISSN 0010-4655. doi: 10.1016/j.cpc.2015.01.024. URL http://dx.doi.org/10.1016/j.cpc.2015.01.024.

[51] V. Khachatryan, A. M. Sirunyan, and et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *The European Physical Journal C*, 76(3), Mar 2016. ISSN 1434-6052. doi: 10.1140/epjc/s10052-016-3988-x. URL http://dx.doi.org/10.1140/epjc/s10052-016-3988-x.

[52] Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. Technical report, CERN, Geneva, 2018. URL https://cds.cern.ch/record/2636284.

[53] Ian Anderson, Sara Bolognesi, and et al. Constraining anomaloushvvinteractions at proton and lepton colliders. *Physical Review D*, 89(3), Feb 2014. ISSN 1550-2368. doi: 10.1103/physrevd.89.035007. URL http://dx.doi.org/10.1103/PhysRevD.89.035007.

[54] Richard D. Ball, Valerio Bertone, and et al. Parton distributions for the lhc run ii. *Journal of High Energy Physics*, 2015(4), Apr 2015. ISSN 1029-8479. doi: 10.1007/jhep04(2015)040. URL http://dx.doi.org/10.1007/JHEP04(2015)040.

[55] S. Agostinelli, J. Allison, and et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(03)01368-8. URL https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[56] Lorenzo Bianchini, John Conway, and et al. Reconstruction of the Higgs mass in H $\rightarrow \tau\tau$ Events by Dynamical Likelihood techniques. *Journal of Physics: Conference Series*, 513(2):022035, jun 2014. doi: 10.1088/1742-6596/513/2/022035. URL https://doi.org/10.1088/1742-6596/513/2/022035.

[57] Albert M Sirunyan et al. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. Submitted to EPJC, 2021.

[58] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$. CMS Physics Analysis Summary CMS-PAS-LUM-17-004, CERN, 2017. URL https://cds.cern.ch/record/2621960.

[59] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at

$\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-002, CERN, 2019. URL https://cds.cern.ch/record/2676164.

[60] Nicolas Berger, Claudia Bertella, and et al. Simplified template cross sections - stage 1.1, 2019.

[61] Albert M Sirunyan, Armen Tumasyan, and et al. Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton–proton collisions at $\sqrt{s} =$ 13 TeV. *Eur. Phys. J. C*, 81:488. 66 p, Mar 2021. doi: 10.1140/epjc/s10052-021-09200-x. URL https://cds.cern.ch/record/2753947. Replaced with the published version. Added the journal reference and the DOI. All the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-19-001 (CMS Public Pages).

[62] S. Chatrchyan, V. Khachatryan, and et al. Evidence for the 125 gev higgs boson decaying to a pair of leptons. *Journal of High Energy Physics*, 2014(5), May 2014. ISSN 1029-8479. doi: 10.1007/jhep05(2014)104. URL http://dx.doi.org/10.1007/JHEP05(2014)104.

[63] Christian Bierlich, Andy Buckley, and et al. Robust Independent Validation of Experiment and Theory: Rivet version 3. *SciPost Physics*, 8(2), Feb 2020. ISSN 2542-4653. doi: 10.21468/scipostphys.8.2.026. URL http://dx.doi.org/10.21468/SciPostPhys.8.2.026.

[64] S Schmitt. Tunfold, an algorithm for correcting migration effects in high energy physics. *Journal of Instrumentation*, 7(10):T10003–T10003, Oct 2012. ISSN 1748-0221. doi: 10.1088/1748-0221/7/10/t10003. URL http://dx.doi.org/10.1088/1748-0221/7/10/T10003.

[65] A.M. Sirunyan, A. Tumasyan, and et al. Measurements of the higgs boson width and anomalous hvv couplings from on-shell and off-shell production in the four-lepton final

state. *Physical Review D*, 99(11), Jun 2019. ISSN 2470-0029. doi: 10.1103/physrevd.99.112003. URL http://dx.doi.org/10.1103/PhysRevD.99.112003.