

Search for the associated production of the Higgs boson with a single top quark in multi-lepton final states and studies with p_T^{miss} and jets at $\sqrt{s} = 13$ TeV with the CMS detector

A Thesis

Submitted to the

Tata Institute of Fundamental Research, Mumbai

for the degree of

Doctor of Philosophy in Physics

by

Pallabi Das

School of Natural Sciences

TATA INSTITUTE OF FUNDAMENTAL RESEARCH, MUMBAI



November, 2019



DECLARATION

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. The work was done under the guidance of Professor Kajari Mazumdar, at the Tata Institute of Fundamental Research, Mumbai.

Pallabi Das
(Pallabi Das)

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Kajari Mazumdar
(Prof. Kajari Mazumdar)

Date: 21.11.19

Dedicated to my mother

Acknowledgements

My doctoral journey has reached its end and I wish to express my gratitude to all the people who have supported and encouraged me for the last five years. I cannot do justice to the task of naming each and every one of you that influenced me within this span of my life in this short piece of writing, but I'm thankful toward all the goodwill.

I would first like to thank the Tata Institute of Fundamental Research, Mumbai for its superb infrastructure, taking care of many practical needs of the research scholars. The serene beauty of the sea-shore and the colonnade will remain very close to my heart.

I'd like to thank my supervisor, Prof. Kajari Mazumdar for her guidance. She has pushed me in the right research directions and supported me in times of need. I have also received help many times from Prof. Monoranjan Guchait, Prof. Gobinda Majumder and Prof. Sudeshna Banerjee. Specially, I want to acknowledge Prof. Guchait's truly great advices in both academic and non-academic situations. I have had great support from the scientific officers Prashant Shinghade, Brij Jashal and Puneet Patel. I also take this opportunity to thank few other staff members of the Department of High Energy Physics: Mandakini Patil, Pramod Pathare, Minal Rane and Ravindra Shinde who have helped me in different occasions.

I thank Aurobindo Holur, Soham, Suman, Muzamil and Arnab who have shared the office room D212 with me for most of these five years in TIFR. We have had many colorful discussions on a myriad of topics, listened to and made fun of each others predicaments, all the while maintaining a serious work environment. My seniors Sanmay, Rajdeep, Saranya, Gouranga, Sanjeev, Ramkrishna, Soureek, Nairit, Deepanwita, Niladri, Siddhesh, and my juniors Sarvesh, Soumya, Mintu, Saikat, Debashish, Soumen, Uttiya have all been great company, thank you.

I must mention Sandhya separately as a great senior and friend who has always patiently listened to my troubles. I owe her for all the help in the beginning of my research work. I am very much grateful to Benjamin Stieger for teaching me many things in the last three years. I look up to him for his very straight forward and practical approach to work. I'd also like to thank Marco Peruzzi who has also helped me many times for my research. Jose Monroy, Nicolas Tonon, Andrew Gilbert, Maria Cepeda Hermida, Leonora Vesterbacka, Jordan Dagmov, Loukas Gouskous, Laurent Thomas, Isabell Pellman, Henning Flecher are among the many people that I have directly interacted with and benefited from within the CMS collaboration. Most of the work in my thesis have been done together with Benjamin, Jose and Nicolas. I specially thank Prof. Ken Bloom, Zeynep Demiragli and Roberto Salerno who have written recommendation letters for my post-doctoral applications. I thank Priyanka, Preeti, Samrangny, Jhuma, Manisha, Saswati, Ankita, Anamika, Archana, Sudha, Shilpi, Nishu, Arun, Raman, Monika, Varun, Ramkrishna, Subhankar, Ashim, Kuntal, Sandeep and many more for making my stays at CERN fun and memorable.

My life in Mumbai would have been mundane, had it not been for Sucheta, Ahana, Angana, Simli, Madhusree and Disha whom I mention among many others as they accompanied me through several ups and downs along the way. Complaining about life in general would not be much fun without them.

Swagata, Juhi, Promita and Madhurima have been my true support group and I cannot thank them enough. I'm also grateful to Anasua, Puja, Nivedita, Ushasi, Tithi for being friends since school years and still keeping in touch. Finally, I mention Pritam for being the default listening ear to all my woes and frustrations. I thank him from the bottom of my heart and hope to have been equally supportive for him.

I thank my parents and family: maa, baba, pisi, jemma, didi, Pinaki da and Ishan. Jethu who passed away just before I came to TIFR; boro jethu and jethima as well, with whom I could not share this achievement. I only hope to have made them proud and whoever that believed in me happy.

Contents

List of Publications	xiii
List of Figures	xx
List of Tables	xxiii
1 Introduction	1
1.1 The Standard Model	2
1.1.1 Particles and symmetries	2
1.1.2 Gauge interactions	3
1.1.3 The Higgs mechanism	4
1.1.4 Higgs coupling to fermions	6
1.2 Shortcomings of Standard Model	6
1.3 Possible existence of new physics	7
1.4 Higgs physics at the LHC	8
1.4.1 Higgs couplings measurement	13
1.4.2 The top-Yukawa coupling	13
1.4.3 Associated production of the Higgs boson with a single top quark	16
2 The LHC and the CMS experiment	18
2.1 The Large Hadron Collider	18
2.2 The Compact Muon Solenoid	21
2.2.1 Overview	22
2.2.2 Particle kinematics	23

2.2.3	Sub-detectors	24
2.2.4	Trigger and data acquisition system	31
2.2.5	CMS computing and software	34
3	Simulation and reconstruction	35
3.1	Theoretical structure of collision events	36
3.1.1	Hard scattering	36
3.1.2	Parton shower and hadronization	37
3.1.3	Underlying event and pileup	38
3.1.4	Detector simulation	39
3.2	Event reconstruction	39
3.3	The PF algorithm	40
3.3.1	Track and vertex	42
3.3.2	Muon	43
3.3.3	Electron	44
3.3.4	Photon and hadrons	45
3.3.5	Jet	45
3.3.6	p_T^{miss}	48
4	Studies with p_T^{miss} and jets	50
4.1	Measurement of early trigger rates in HCAL	50
4.1.1	Event selection	51
4.1.2	Data samples	52
4.1.3	Results	53
4.1.4	Effect of beam halo filter	55
4.2	Performance of algorithms for p_T^{miss} measurements	56
4.2.1	Corrections and uncertainties	57
4.2.2	Data and simulated samples	58
4.2.3	Objects	61
4.2.4	Event selection	65
4.2.5	Data/MC comparison	66

4.2.6	Performance of reconstruction	79
4.2.7	Studies with reconstruction recipes	88
4.2.8	Conclusions	94
5	Associated production of the Higgs boson with single top quark	95
5.1	Measurements of Higgs properties	97
5.1.1	Relative sign of top-Yukawa coupling	98
5.1.2	CP properties of the Higgs	98
5.2	Analysis strategy	100
5.2.1	Background estimation	101
5.2.2	Multivariate analysis	104
5.2.3	Statistical interpretation	106
5.2.4	Analysis workflow	109
6	tH multi-lepton at $\sqrt{s} = 13$ TeV	110
6.1	Analysis using 2016 data	111
6.1.1	Data and simulated samples	111
6.1.2	Triggers	113
6.1.3	Identification of leptons	116
6.1.4	Lepton selection	121
6.1.5	Taus	122
6.1.6	Jets and b-tagging	124
6.1.7	Event selection	125
6.1.8	Signal discrimination	126
6.1.9	Signal extraction	128
6.1.10	Systematic uncertainties	133
6.1.11	Results	135
6.2	Analysis using 2017 data	140
6.2.1	Data and simulated samples	140
6.2.2	Triggers	141
6.2.3	Identification of leptons	143

6.2.4	Lepton selection	145
6.2.5	Taus	145
6.2.6	Jets and b-tagging	146
6.2.7	Event selection	148
6.2.8	Signal discrimination	148
6.2.9	Signal extraction	150
6.2.10	Systematic uncertainties	152
6.2.11	Results	154
6.3	Combination of 2016 and 2017 results	159
7	Combination of tH final states and future prospects	163
7.1	Results from combination based on 2016 data	164
7.2	Future prospects of tH search	167
7.2.1	Projection scenarios	168
7.2.2	Results	170
7.2.3	Significance of measurement	172
8	Conclusions and outlook	173
	Appendices	197
A	Training of BDT discriminators	199
A.1	Training with 2016 samples	200
A.2	Training with 2017 samples	208
B	Control region distributions	217
B.1	Two leptons same-sign	217
B.2	Three leptons	217
B.3	Two leptons opposite-sign	221

LIST OF PUBLICATIONS

Internal CMS Documents

- J. Monroy, P. Das, B. Stieger et al., “**Search for tHq production in multilepton final states at 13 TeV**”, CMS AN-2016/378
- P. Das, “**Jet Pre-firing Studies in 2016 p-p Collision Data**”, CMS DN-2017/001
- B. Stieger, P. Das, J. Monroy et al., “**Combination of tHq analyses with 2016 dataset**”, CMS AN-2018/082
- P. Das, K. Mazumdar, “**Potential for the study of tHq multilepton at HL-LHC**”, CMS AN-2018/156
- P. Das, N. Tonon, B. Stieger et al., “**tHq multilepton analysis with 2017 data**”, CMS AN-2019/078
- P. Das, S. Kumar, K. Mazumdar, “**Study of the performance of the p_T^{miss} algorithms in 2017 and 2018 data using γ +jets events**”, CMS AN-2019/115

Papers and Public Notes

- CMS Collaboration, “**Search for production of a Higgs boson and a single top quark in multilepton final states in proton collisions at $\sqrt{s} = 13$ TeV**”, 2017, [CMS-PAS-HIG-17-005](#)
- CMS Collaboration, “**Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s} = 13$ TeV**”, 2018, [Phys. Rev. D 99, 092005 \(2019\)](#), [arXiv:1811.09696](#)

- CMS Collaboration, “**Sensitivity Projections for Higgs boson properties measurements at HL-LHC**”, 2018, [CMS-PAS-FTR-18-011](#)
- Cepeda, M. and others, “**Higgs Physics at the HL-LHC and HE-LHC**”, 2019, [CERN-LPCC-2018-04](#), [arXiv:1902.00134](#)

Conference Proceedings

- P. Das, “**New physics searches in the top quark sector in hadron colliders**”, [PoS CKM2016 \(2017\) 129](#)
- P. Das, “**Study of the Associated Production of Higgs Boson with a Single Top Quark**”, [Springer Proc. Phys. 203 \(2018\) 833-835](#)
- P. Das, K. Mazumdar, “**Measurement of Missing Transverse Energy in CMS Experiment**”, [Springer Proc. Phys. 203 \(2018\) 521-524](#)
- P. Das, “**Searches for the associated production of a Higgs boson with a single top quark at $\sqrt{s} = 13$ TeV at CMS**”, [PoS EPS-HEP2017 \(2018\) 270](#)

Publications as an author of CMS Collaboration can be found in INSPIRE publication database author profile [INSPIRE-00451710](#).

List of Figures

1.1	The SM Higgs boson production cross sections and decay fractions.	9
1.2	The SM Higgs boson production cross sections and decay fractions.	9
1.3	Reconstructed mass spectrum showing the Higgs resonance.	10
1.4	Main leading order Feynman diagrams contributing to the Higgs boson production at the LHC.	12
1.5	Run 1 combination of Higgs coupling modifiers measurement.	14
2.1	Overview of CERN accelerator complex.	20
2.2	Total integrated luminosity and average pileup distribution recorded by CMS.	22
2.3	Schematic view of the CMS detector.	23
2.4	Schematic view of the inner tracking system.	26
2.5	Location of the ECAL and HCAL around the magnet.	29
2.6	Schematic diagram depicting the CMS muon detectors.	30
3.1	Steps of event reconstruction in the CMS analyses.	40
3.2	Slice view of CMS detector.	41
3.3	Performance of b-tagging algorithms.	48
4.1	Bunch pattern of fill number 5078.	52
4.2	Leading jet E_T distributions in bunch fill 5355.	53
4.3	Pre-firing rates in bunch fill 4947.	54
4.4	Pre-firing rates in bunch fill 5078.	54
4.5	Pre-firing rates in bunch fill 5211.	54

4.6 Pre-firing rates in bunch fill 5355.	55
4.7 Pre-firing jets without and with beam halo filter.	55
4.8 Nominal BX jets without and with beam halo filter.	56
4.9 Illustration of γ +jets event kinematics.	57
4.10 Number of vertices distributions in 2017 data.	67
4.11 p_T and η distributions of the tight photon in 2017 data.	68
4.12 p_T and η distributions of the leading jet in 2017 data.	68
4.13 Distributions of the hadronic recoil components in 2017 data.	69
4.14 Raw p_T^{miss} distributions in 2017 data.	70
4.15 Type 1 p_T^{miss} distributions in 2017 data.	71
4.16 Type 1 PUPPI p_T^{miss} distributions in 2017 data.	72
4.17 Number of vertices distributions in 2018 data.	73
4.18 p_T and η distributions of the tight photon in 2018 data.	74
4.19 p_T and η distributions of the leading jet in 2018 data.	74
4.20 Distributions of the hadronic recoil components in 2018 data.	75
4.21 Raw p_T^{miss} distributions in 2018 data.	76
4.22 Type 1 p_T^{miss} distributions in 2018 data.	77
4.23 Type 1 PUPPI p_T^{miss} distributions in 2018 data.	78
4.24 Response curves for raw p_T^{miss} in years 2017 and 2018.	80
4.25 Response curves for Type 1 PF p_T^{miss} in years 2017 and 2018.	81
4.26 Response curves for Type 1 PUPPI p_T^{miss} in years 2017 and 2018.	81
4.27 Resolution as a function of q_T for the parallel and perpendicular compo- nents of raw p_T^{miss} in years 2017 and 2018.	82
4.28 Resolution as a function of q_T for the parallel and perpendicular compo- nents of Type 1 PF p_T^{miss} in years 2017 and 2018.	83
4.29 Resolution as a function of q_T for the parallel and perpendicular compo- nents of Type 1 PUPPI p_T^{miss} in years 2017 and 2018.	84
4.30 Resolution as a function of N_{vtx} for the parallel and perpendicular compo- nents of raw p_T^{miss} in years 2017 and 2018.	85

4.31 Resolution as a function of N_{vtx} for the parallel and perpendicular compo-	
nents of Type 1 PF p_T^{miss} in years 2017 and 2018.	86
4.32 Resolution as a function of N_{vtx} for the parallel and perpendicular compo-	
nents of Type 1 PUPPI p_T^{miss} in years 2017 and 2018.	87
4.33 Type 1 PF p_T^{miss} distributions with nominal reconstruction in 2017 data. . .	89
4.34 Type 1 PF p_T^{miss} distributions with EENoiseFix recipe in 2017 data.	90
4.35 Type 1 PF p_T^{miss} distributions with Ultra Legacy candidate dataset.	91
4.36 Response of the Type 1 PF p_T^{miss} algorithm with Ultra Legacy candidate	
dataset.	91
4.37 p_T^{miss} distributions with pileup and N_{vtx} reweighting.	92
4.38 Type 1 PF p_T^{miss} distributions with corrections in x and y components. . . .	93
5.1 Feynman diagrams contributing to the t-channel tH production.	96
5.2 Feynman diagrams contributing to the s-channel tH production.	96
5.3 Feynman diagrams contributing to the W -associated tH production.	96
5.4 Interference of tHW production in the 4FS scheme with the $t\bar{t}H$ process. .	97
5.5 Variation of tH and $t\bar{t}H$ inclusive cross sections as a function of κ_t/κ_V	99
5.6 Variation of tH and $t\bar{t}H$ inclusive cross sections as a function of α	100
5.7 Schematic view of a decision tree and multivariate classification.	105
5.8 Illustration of hypothesis testing.	108
5.9 Illustration of analysis workflow.	109
6.1 BDT output distributions in 2016 analysis for 3ℓ channel.	129
6.2 BDT output distributions in 2016 analysis for $e^\pm\mu^\pm$ channel.	129
6.3 BDT output distributions in 2016 analysis for $\mu^\pm\mu^\pm$ channel.	130
6.4 BDT classifier output planes in 2016 analysis.	131
6.5 BDT classifier output planes in 2016 analysis.	132
6.6 Binning overlaid on the S/B ratio in 2016 analysis.	132
6.7 Pre-fit and post-fit BDT bin distributions in 2016 analysis for 3ℓ channel. .	136
6.8 Pre-fit and post-fit BDT bin distributions in 2016 analysis for $2\ell ss e^\pm\mu^\pm$	
channel.	137

6.9 Pre-fit and post-fit BDT bin distributions in 2016 analysis for $2\ell ss \mu^\pm \mu^\pm$ channel.	137
6.10 Upper limits on $\sigma \times BR$ of $tH+t\bar{t}H$ signal in 2016 analysis.	138
6.11 Upper limits on $\sigma \times BR$ of tH signal as a function of κ_t in 2016 analysis.	139
6.12 BDT output distributions in 2017 analysis for 3ℓ channel.	149
6.13 BDT output distributions in 2017 analysis for $e^\pm \mu^\pm$ channel.	149
6.14 BDT output distributions in 2017 analysis for $\mu^\pm \mu^\pm$ channel.	151
6.15 Binning overlaid on the S/B ratio in 2017 analysis.	151
6.16 Pre-fit and post-fit BDT bin distributions in 2017 analysis for 3ℓ channel.	154
6.17 Pre-fit and post-fit BDT bin distributions in 2017 analysis for $e^\pm \mu^\pm$ channel.	155
6.18 Pre-fit and post-fit BDT bin distributions in 2017 analysis for $\mu^\pm \mu^\pm$ channel.	155
6.19 Upper limits on $\sigma \times BR$ of tH signal and NLL scan as a function of κ_t in 2017 analysis.	157
6.20 Upper limits on $\sigma \times BR$ of $tH+t\bar{t}H$ signal and NLL scan as a function of $\cos \alpha$ in 2017 analysis.	158
6.21 Combined (2016+2017) upper limits on $\sigma \times BR$ of tH signal and NLL scan as a function of κ_t .	160
7.1 Observed and background-only expected $\sigma \times BR$ limits on combined tH production.	164
7.2 NLL scan of κ_t for the combined tH production.	166
7.3 Timeline for the LHC and HL-LHC data taking and upgrade.	167
7.4 NLL scan of κ_t for the Asimov datasets corresponding to $\kappa_V=1.0$ with scenarios S1 and S2.	170
7.5 The variation of expected upper limit on μ_{tH} with scenarios S1 and S2.	171
A.1 BDT inputs as seen by TMVA for the 3ℓ channel, discriminated against $t\bar{t}$ in 2016 analysis.	201
A.2 BDT inputs as seen by TMVA for the 3ℓ channel, discriminated against $t\bar{t}V$ in 2016 analysis.	202

A.3 BDT inputs as seen by TMVA for the $2\ell ss$ channel, discriminated against $t\bar{t}$ in 2016 analysis.	203
A.4 BDT inputs as seen by TMVA for the $2\ell ss$ channel, discriminated against $t\bar{t}V$ in 2016 analysis.	204
A.5 Correlation matrices for the input variables for the 3ℓ channel in 2016 analysis.	205
A.6 Correlation matrices for the input variables for the $2\ell ss$ channel in 2016 analysis.	205
A.7 ROC curves and classifier output distributions for the 3ℓ channel in 2016 analysis.	206
A.8 ROC curves and classifier output distributions for the $2\ell ss$ channel in 2016 analysis.	207
A.9 BDT inputs as seen by TMVA for the 3ℓ channel, discriminated against $t\bar{t}$ in 2017 analysis.	209
A.10 BDT inputs as seen by TMVA for the 3ℓ channel, discriminated against $t\bar{t}V$ in 2017 analysis.	210
A.11 BDT inputs as seen by TMVA for the $2\ell ss$ channel, discriminated against $t\bar{t}$ in 2017 analysis.	211
A.12 BDT inputs as seen by TMVA for the $2\ell ss$ channel, discriminated against $t\bar{t}V$ in 2017 analysis.	212
A.13 The SumJetsPt variable input to BDT training as seen by TMVA showing more discrimination power against the prompt background.	213
A.14 Correlation matrices for the input variables for the 3ℓ channel in 2017 analysis.	213
A.15 Correlation matrices for the input variables for the $2\ell ss$ channel in 2017 analysis.	214
A.16 ROC curves and classifier output distributions for the 3ℓ channel in 2017 analysis.	214
A.17 ROC curves and classifier output distributions for the $2\ell ss$ channel in 2017 analysis.	215

B.1	Contributions from signal and background in $t\bar{t}$ control region for $\mu^\pm\mu^\pm$ channel in the 2016 analysis.	218
B.2	Contributions from signal and background in $t\bar{t}$ control region for $\mu^\pm\mu^\pm$ channel in the 2017 analysis.	218
B.3	Contributions from signal and background in $t\bar{t}$ control region for $e^\pm\mu^\pm$ channel in the 2016 analysis.	219
B.4	Contributions from signal and background in $t\bar{t}$ control region for $e^\pm\mu^\pm$ channel in the 2017 analysis.	219
B.5	Distributions of input variables to the BDT for signal discrimination, in Z enriched 3ℓ selection in the 2016 analysis.	220
B.6	Distributions of input variables to the BDT for signal discrimination, in Z enriched 3ℓ selection in the 2017 analysis.	220
B.7	Kinematic distributions in the $t\bar{t}$ -enriched opposite-sign $e\mu$ selection in 2016 analysis.	221
B.8	Pseudorapidity of the light forward jet in the $t\bar{t}$ -enriched opposite-sign $e\mu$ selection for the three p_T cut values in 2016 analysis.	222
B.9	Pseudorapidity of the light forward jet in the $t\bar{t}$ -enriched opposite-sign $e\mu$ selection for the four options in 2017 analysis.	224

List of Tables

1.1	Quantum numbers of the three fermion generations.	3
1.2	Decay branching fractions of the Higgs boson.	12
1.3	Couplings of the fermions to the scalar and pseudoscalar Higgs bosons in 2HDM model.	15
2.1	Types of silicon strip detector used in CMS in Runs 1 and 2.	26
2.2	Design parameters of the CMS trigger and DAQ system.	32
4.1	Data samples used to determine pre-firing rates and the dates of collection.	53
4.2	Data samples used in the determination of p_T^{miss} reconstruction performance.	59
4.3	List of simulated samples considered for the γ +jets final state, along with the cross sections.	60
4.4	High-level triggers used to select events with photons.	61
4.5	Effective areas for photon isolation.	62
4.6	PF jet identification criteria used to select γ +jets events.	63
4.7	Loose identification criteria for electrons.	64
4.8	Loose identification criteria for muons.	65
6.1	tH and $t\bar{t}H$ samples with their cross sections in the ITC scenario used in 2016 analysis.	112
6.2	LO $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$ samples used in the BDT training in 2016 analysis.	112
6.3	List of background MC samples used in 2016 analysis.	114
6.4	Table of high-level triggers used in 2016 analysis for overlap removal.	115
6.5	List of HLT paths considered in the 2016 analysis for event selection.	116

6.6	Trigger efficiency scale factors in 2016 analysis.	117
6.7	Effective areas for electrons and muons used in 2016 analysis.	119
6.8	Requirements on each of the three muon selections in 2016 analysis	122
6.9	Criteria for each of the three electron selections in 2016 analysis.	123
6.10	Approximate branching fractions of 1- and 3- prong hadronic decay of τ .	124
6.11	PF jet identification criteria used in 2016 analysis.	124
6.12	Data yields and expected backgrounds at 35.9 fb^{-1} .	127
6.13	List of input variables used for the MVA training.	130
6.14	Production cross sections for the tH and $t\bar{t}H$ samples in the ITC scenario used in 2017 analysis.	140
6.15	LO $t\bar{t}W$ and $t\bar{t}Z$ samples and NLO $t\bar{t}$ samples used in the signal BDT train- ing in 2017 analysis.	141
6.16	List of background MC samples used in 2017 analysis.	142
6.17	Table of high-level triggers used in 2017 analysis for overlap removal.	143
6.18	List of HLT paths considered in the 2017 analysis for event selection.	144
6.19	Trigger efficiency scale factors in 2017 analysis.	144
6.20	Effective areas for electrons and muons used in 2017 analysis.	145
6.21	Criteria for each of the three muon selections in 2017 analysis.	146
6.22	Criteria for each of the three electron selections in 2017 analysis.	147
6.23	PF jet identification criteria used in 2017 analysis.	147
6.24	Data yields and expected backgrounds at 41.5 fb^{-1} .	150
6.25	95% CL observed (expected) upper limits on $tH, H \rightarrow WW^*/ZZ^*/\tau\tau$ signal strength obtained from multi-lepton channel with the SM and ITC scenarios.	161
6.26	95% CL observed (expected) upper limits on $tH, H \rightarrow WW^*/ZZ^*/\tau\tau$ cross section obtained from multi-lepton channel with the SM and ITC scenarios.	161
6.27	95% CL observed (expected) upper limits on $\sigma \times BR$ of combined $tH+t\bar{t}H$ production with $H \rightarrow WW^*/ZZ^*/\tau\tau$ decay in SM and ITC scenarios.	162
7.1	95% CL observed (expected) upper limits, best-fit signal strength and sig- nificance.	165

7.2	95% CL observed (expected) upper limits on the tH production cross section times $H \rightarrow WW/ZZ/\tau\tau/b\bar{b}/\gamma\gamma$ branching fraction.	165
7.3	Summary of the main sources of systematic uncertainty.	166
7.4	The sources of systematic uncertainty for which minimum values are applied in S2.	169
7.5	The $\pm 1\sigma$ uncertainties on expected $\mu_{tH}=1$ for scenarios S1 and S2.	172
A.1	MVA input discriminating variables.	199
A.2	TMVA configuration used in the training.	200
A.3	TMVA input variables ranking for BDTG method in the 3ℓ channel.	207
A.4	TMVA input variables ranking for BDTG method in the $2\ell_{ss}$ channel.	208
A.5	TMVA configuration used in the training.	208
A.6	TMVA input variables ranking for BDTG method in the 3ℓ channel.	216
A.7	TMVA input variables ranking for BDTG method in the $2\ell_{ss}$ channel.	216
B.1	Data/MC scale factors for η distribution of light forward jet with three different p_T cuts.	223
B.2	Data/MC scale factors for different requirements on the light forward jet.	224

Chapter 1

Introduction

Mankind is driven by questions concerning the universe since time immemorial. As our ancestors looked above at the stars, some imagined the glimmering lights to have supernatural origins whereas some built instruments to actually peer into the truth. In current society, both oriental and occidental, scientific ignorance still prevails. However, in stark contrast, the achievements of modern science are extra-ordinary in all fields of fundamental knowledge. This group of people that had started with humble apparatus such as a piece of polished glass have come a long way to present day, where conditions of the early universe are re-created by smashing sub-atomic particles together. Experimental particle physics studies the building blocks of Nature reaching into smallest-ever length scales with highest-ever energies with the help of particle colliders. The experimental and theoretical findings of the minute details of matter are best described by the Standard Model (SM) of particle physics [1-3].

The electromagnetic interactions of particles were the first to be explained in terms of quantum electrodynamics (QED) within the framework of a quantum theory in the 1940's. Following the success of QED, weak and strong interaction theories had also been proposed based on local gauge symmetries. The unification of QED and weak interactions predicted massless gauge bosons, but no such particles were detected mediating the weak force. It was realized then that spontaneously broken symmetries due to the existence of a massive neutral scalar particle, the Higgs boson, can give mass to gauge bosons [4-7]. The massive gauge particles, $W^\pm$ and Z^0 , have been discovered in experiment in the

1980's [8–11], with the predicted properties. This strengthened the case of the Higgs boson and experiments at the Large Hadron Collider (LHC) were designed to search for it. One of the major achievements of Physics in recent history is the discovery of the Higgs boson in 2012 [12, 13], 50 years after it was postulated, completing the spectrum of fundamental particles according to the SM. The current mandate of the experiments at the LHC is to characterize the newly-discovered particle by measuring its properties and search for new physics phenomena beyond the SM prediction. This thesis is a continuation towards this global effort as presented in the following.

1.1 The Standard Model

1.1.1 Particles and symmetries

As the smallest constituents of matter at high energies are governed by quantum laws, their fields and interactions are most elegantly described by Quantum Field Theory (QFT). Using this approach SM interprets quantum excitations of a field as particles, categorized into two groups: *fermions* (spin 1/2) which constitute matter and *vector bosons* (spin 1) which mediate interactions.

Fermions are further divided into *leptons* and *quarks*, each of which come in three generations with increasing mass. The weak and electromagnetic interactions are described by the gauge symmetry group $SU(2)_L \times U(1)_Y$ while the strong interaction or quantum chromodynamics (QCD) is described by $SU(3)_c$ symmetry group. Each fermion or matter particle has a quantum number corresponding to each of the symmetries, which remains conserved according to Noether's theorem [14]. These quantum numbers are electric charge, weak isospin corresponding to electroweak interaction and color charge corresponding to QCD interaction. Table 1.1 shows the electroweak quantum numbers of the fermions as described by SM. Leptons do not have any color charge and hence they do not experience strong interaction, whereas quarks can have three possible colors: red, green, blue.

The vector bosons mediate the fundamental interactions relevant for elementary particles

Table 1.1: Quantum numbers of the three fermion generations. Left-handed fermions, denoted using L , are doublets under weak interaction whereas right-handed fermions, denoted using R , are singlets. Right handed neutrinos are not included in the SM. Leptons constitute the electron (e), the muon (μ) and the tau (τ) along with corresponding flavors of neutrinos (ν). Left handed quarks are paired in isospin as: up (u) and down (d), charm (c) and strange (s), top (t) and bottom (b). Here T_3 stands for the third component of the weak isospin, Q is electric charge and Y weak hypercharge, defined as $Q = T_3 + \frac{Y}{2}$.

Particles	T_3	Q	Y
Leptons $\nu_{eL}, \nu_{\mu L}, \nu_{\tau L}$	$+\frac{1}{2}$	0	-1
	$-\frac{1}{2}$	-1	-1
	0	-1	-2
Quarks	$+\frac{1}{2}$	$+\frac{2}{3}$	$+\frac{1}{3}$
	0	$+\frac{2}{3}$	$+\frac{4}{3}$
	$-\frac{1}{2}$	$-\frac{1}{3}$	$+\frac{1}{3}$
	0	$-\frac{1}{3}$	$-\frac{2}{3}$

– electromagnetic, strong and weak. The infinite range electromagnetic force is carried by the massless photon, described by QED. Gluons, mediator of the strong force, are also massless. But unlike photons, gluons self-interact as they carry the color charge, leading to *confinement* which results in the limited range ($\sim 10^{-15}$ m) of the strong force. Weak nuclear interactions are also short range ($\sim 10^{-18}$ m), mediated by W and Z bosons.

The widespread acceptance of SM as a theory of the building blocks of the physical universe lies in the sheer success of explaining the existing and broken symmetries of the fundamental forces of Nature, except gravity.

1.1.2 Gauge interactions

The gauge invariant kinetic term of the SM interaction Lagrangian is written as:

$$\mathcal{L}_{gauge} = -\frac{1}{4}W_{\mu\nu}^a W^{\mu\nu,a} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu} - \frac{1}{4}G_{\mu\nu}^a G^{\mu\nu,a} \quad (1.1)$$

where

$$\begin{aligned} W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon^{abc} W_\mu^b W_\nu^c \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ G_{\mu\nu}^a &= \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c. \end{aligned} \quad (1.2)$$

W_μ^a , $a = 1, 2, 3$ and B_μ are field strengths of the electroweak interaction and G_μ^a , $a = 1, \dots, 8$ corresponds to gluon field strength. The g and g_s are gauge coupling strengths of weak and strong interactions respectively, and ϵ^{abc} and f^{abc} are structure constants of the $SU(2)_L$ and $SU(3)_c$ symmetry groups respectively. The gauge index a indicates different number of gauge fields: in general, for $SU(N)$ group it will run from 1 to $N^2 - 1$.

The kinetic and interaction parts of the Lagrangian describing fermions is expressed in terms of covariant derivative D_μ which preserves the gauge invariance:

$$D_\mu \equiv \partial_\mu + ig \frac{\sigma_a}{2} W_\mu^a P_L + ig' \frac{Y}{2} B_\mu + ig_s \frac{\lambda_a}{2} G_\mu^a \quad (1.3)$$

where $P_L = (1 - \gamma_5)/2$ is the left-handed projection operator, σ_a are the Pauli matrices and $T_a = \sigma_a/2$ are the generators of $SU(2)_L$ symmetry, g' is the coupling strength of electromagnetic interaction and λ_a are Gell-Mann matrices, $t_a = \lambda_a/2$ being the generators of $SU(3)_c$ symmetry. The complete gauge invariant Lagrangian describing electroweak(EW) and QCD interaction is thus:

$$\mathcal{L}_{EW+QCD} = -\frac{1}{4} W_{\mu\nu}^a W^{\mu\nu,a} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} G_{\mu\nu}^a G^{\mu\nu,a} + i\bar{\psi}\gamma_\mu D^\mu \psi \quad (1.4)$$

where ψ denotes the fermion field. Explicit mass terms of the vector boson fields are prohibited as they violate gauge symmetry. Fermionic mass terms mix the left-handed and right-handed components and are thus forbidden, as by above formulation their transformation properties under $SU(2)_L \times U(1)_Y$ clearly differentiates chirality (handedness).

1.1.3 The Higgs mechanism

Masses of the vector bosons are generated in the SM by the Electroweak Symmetry Breaking (EWSB) mechanism [15–19]. In the Glashow-Salam-Weinberg (GSW) model, a weak isospin doublet of a complex scalar field (Φ) spontaneously breaks the $SU(2)_L$ symmetry by acquiring a non-zero vacuum expectation value (VEV). In this process the elec-

tromagnetic gauge group $U(1)_{EM}$ remains unbroken and electric charge continues to be conserved. The interaction Lagrangian of Φ is written as:

$$\mathcal{L}_{Higgs} = |(\partial_\mu + ig \frac{\sigma_a}{2} W_\mu^a + ig' \frac{Y}{2} B_\mu) \Phi|^2 - V(\Phi) \quad (1.5)$$

which includes the scalar potential of the form:

$$V(\Phi) = -\mu^2 |\Phi|^2 + \lambda |\Phi|^4. \quad (1.6)$$

By choosing a VEV v and expanding about the vacuum such that

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \quad (1.7)$$

where $v = \sqrt{\mu^2/\lambda}$ with $\mu^2 > 0$, $\lambda > 0$ and H is the Higgs field, the electroweak symmetry is broken. Using Eq.(1.7) in Eq.(1.5) the mass terms of the vector bosons are generated from the kinetic terms:

$$M_W = \frac{1}{2} v g; \quad M_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}; \quad M_\gamma = 0. \quad (1.8)$$

The Higgs field also gains a mass term $m_H = \sqrt{2\lambda}v$, if we substitute Eq.(1.7) in Eq.(1.6). The m_H is thus generated by the self-interaction of H and depends on the self-coupling parameter λ .

The bosonic sector of the GSW model is solely described by the four parameters: g , g' , μ and λ . The weak coupling constant g is evaluated using the muon lifetime measurement experiment [20–23] based on the Fermi's theory of weak interaction [24] and the low-energy value is 0.652. Precise masses of W and Z bosons are known to be 80.385 GeV and 91.1876 GeV respectively [25], thus using Eq.(1.8) the value of $v \approx 246$ GeV and $U(1)$ coupling constant g' is determined to be 0.357. The Higgs boson was discovered in the experiments having mass near 125 GeV [26], implying $\lambda \approx 0.13$ and $|\mu| \approx 88.8$ GeV at the electroweak scale. The Higgs couplings to the vector boson particles ($V = W, Z$) are proportional to m_V^2/v . Hence, determination of the physical masses of various particles gives an exact picture of the underlying theory, a very important feature of the SM.

1.1.4 Higgs coupling to fermions

Fermion masses are generated by the Higgs field through the Yukawa interaction, the most general form of which can be written as:

$$\mathcal{L}_{Yukawa} = -y_d \bar{\psi}_L \Phi \psi_R - y_u \bar{\psi}_L \tilde{\Phi} \psi_R + \text{h.c.} \quad (1.9)$$

Here y_d, y_u are Yukawa coupling strengths to quarks and leptons with quantum numbers $T_3 = -1/2$ and $1/2$ respectively, and $\tilde{\Phi} = i\sigma_2 \Phi^\dagger$ is the charge conjugate of the field Φ . Neutrinos do not take part in this interaction since there exists no right handed neutrinos within the framework of SM. After the VEV is generated the fermions gain masses of the form: $m_f = y_f v / \sqrt{2}$ from the interaction terms of Eq. (1.9), neutrinos remain massless. The strength of the individual interactions with the scalar field essentially determines the mass of each particle. Given the Higgs boson mass, then, each of the couplings is predicted by the SM.

1.2 Shortcomings of Standard Model

The SM can explain most of the experimental observations with spectacular success but quite a few open questions still remain. Various parameters such as the Higgs boson mass itself is not predicted by the theory, neither are the other particle masses: they are inputs to the theory. The appearance of particles in several generations and disparate mass spectrum cannot be explained nor can they be predicted a-priori. The physical Higgs boson mass diverges in this theory due to higher order radiative corrections from particle loops, commonly referred to as the hierarchy problem [27]. An unsatisfactory explanation can be given in terms of fine tuning of the mass correction terms, but this has to be done order-by-order in a perturbation theory. SM does not incorporate the fourth fundamental interaction, viz. gravity and therefore breaks down at very high energy scale $\sim 10^{19}$ GeV where gravity becomes strong. Equally critical is the lack of understanding of the nature of dark matter [28, 29], origin of neutrino masses and oscillations [30–33] which have been observed in experiments. Several other known issues include matter-antimatter asymmetry observed in the universe, extent of CP violation etc. [34] that render the for-

malism incomplete. It is possible that SM is only a low-energy manifestation of a theory encompassing several other symmetries relevant at higher energy scales. Before LHC, experiments were carried out at an energy scale of several hundred GeV. LHC has opened the door to studying fundamental interactions at centre-of-mass energies ($\sqrt{s}$) ten times higher and hence test the validity of the SM at TeV energies.

1.3 Possible existence of new physics

Beyond Standard Model or BSM theories try to address the open questions by proposing existence of new symmetries, higher dimensional operators etc. The most popular theory alternative that answers some of the drawbacks of SM is Supersymmetry (SUSY). It provides an elegant solution to the hierarchy problem by postulating bosonic superpartners of fermions and fermionic superpartners of gauge bosons which contribute to cancel the diverging correction terms [27]. It can even incorporate possible dark matter candidates in the theory. However, the beauty of this mathematical construct is not supported by current observation, unlike SM. SUSY doubles the particle content of SM, yet none of the superpartner particles have been observed. This leads to the conclusion that SUSY must be a broken symmetry, where the superpartner particle masses are much greater than the corresponding SM particle.

The most simple supersymmetric model that extends the SM is the Minimal Supersymmetric Standard Model (MSSM) [35–40]. In particular, MSSM theory includes two complex scalar Higgs boson doublets under $SU(2)$ symmetry:

$$H_1 = \begin{pmatrix} h_1^0 \\ h_1^- \end{pmatrix}, H_2 = \begin{pmatrix} h_2^+ \\ h_2^0 \end{pmatrix} \quad (1.10)$$

with VEVs v_1, v_2 respectively such that $v^2 = v_1^2 + v_2^2$ where v is the VEV from SM, and defining the parameter $\tan \beta = v_2/v_1$. h_1^0, h_1^- are isospin $\pm 1/2$ components of H_1 and h_2^+, h_2^0 are isospin $\pm 1/2$ components of H_2 .

An important facet of MSSM is that there can be mixing between scalar states resulting in different mass eigenstates after EWSB. The physical particle spectrum of MSSM Higgs boson sector contains two CP-even scalars h and H , one CP-odd pseudoscalar A , and two

charged scalars $H^\pm$ and described by the two parameters: pseudoscalar mass m_A and $\tan\beta$, the ratio of the VEVs. In contrast to SM which cannot predict the mass of the Higgs boson, at least one neutral scalar Higgs boson mass, which is a linear combination of h_1^0 and h_2^0 , is predicted to be lighter than 135 GeV in the MSSM [41, 42]. It is interesting that the observed 125 GeV particle falls nicely under this description.

1.4 Higgs physics at the LHC

Before the discovery of the Higgs boson, a lower bound of 114.4 GeV [43] was set on its mass using data from the Large Electron-Positron Collider (LEP) with highest collision energy reaching 209 GeV. An upper bound of 1 TeV [44] was set much earlier on the grounds of unitarity violation above this scale. The relative contributions of different Higgs boson production modes as well as the decay branching fractions depend strongly on the Higgs boson mass. Figures 1.1 and 1.2 show the theoretical predictions for variations of production cross sections and the decay fractions as a function of the Higgs boson mass at $\sqrt{s} = 13$ TeV.

At the LHC, the strongest evidence of the resonance was obtained around 125 GeV using the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels by probing the reconstructed mass range, shown in Fig. 1.3.

Incidentally these two decay modes also provide the best mass resolution for the Higgs boson. Currently the Higgs boson mass is measured with better than 2% precision using the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ [49, 50]:

$$m_H = 125.35 \pm 0.15 \text{ GeV}.$$

Given this information, the predicted natural width of the Higgs boson is $\Gamma_H \sim 4$ MeV which is much smaller than the experimental resolution achievable as seen in Fig. 1.3. Thus Γ_H can only be constrained by studying various production and decay modes of Higgs boson. For the measured mass, the main production processes at the LHC are [51]:

- **Gluon-gluon fusion (ggF):** It is the dominating production mode at the LHC since gluons carry a large fraction of the proton momentum. The leading Feynman dia-

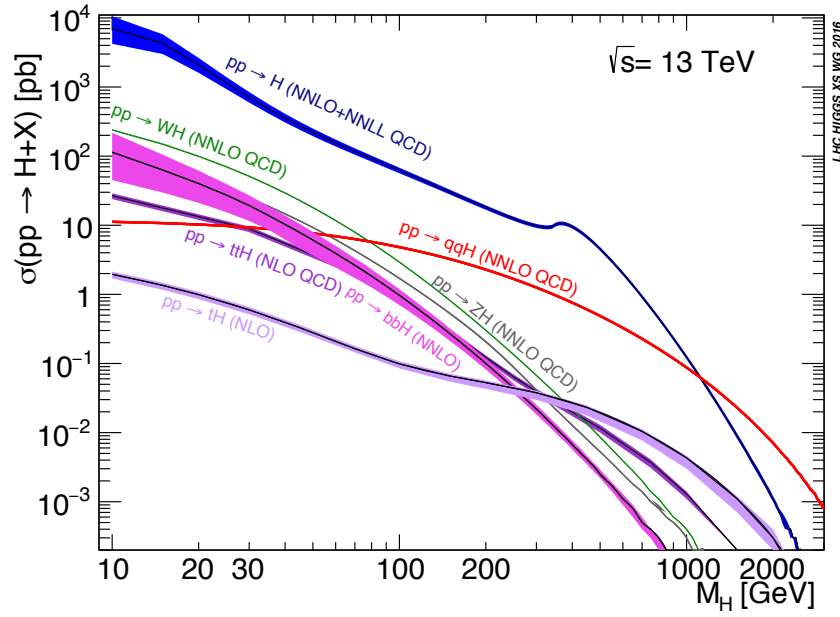


Figure 1.1: The SM Higgs boson production cross sections as a function of the Higgs boson mass [45]. The VBF process is indicated as qqH.

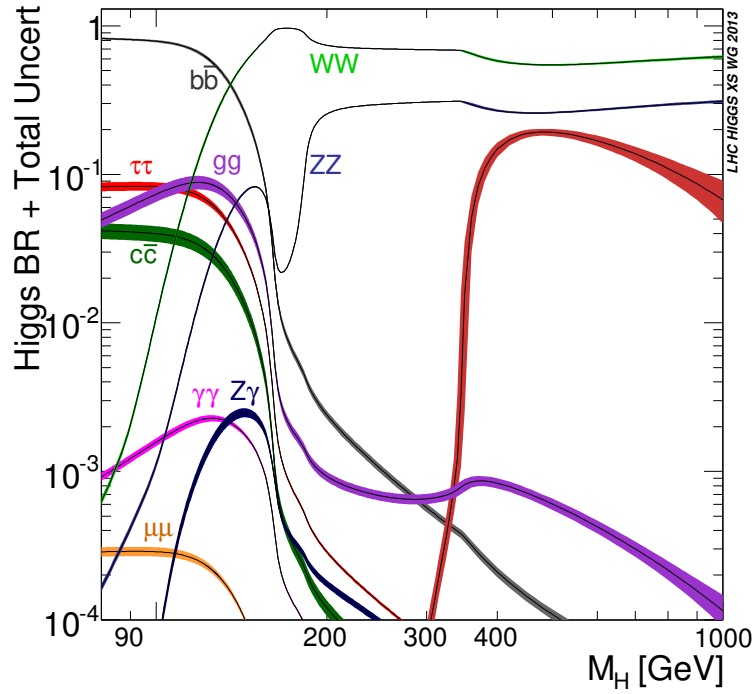


Figure 1.2: The SM Higgs boson decay fractions as a function of the Higgs boson mass [46].

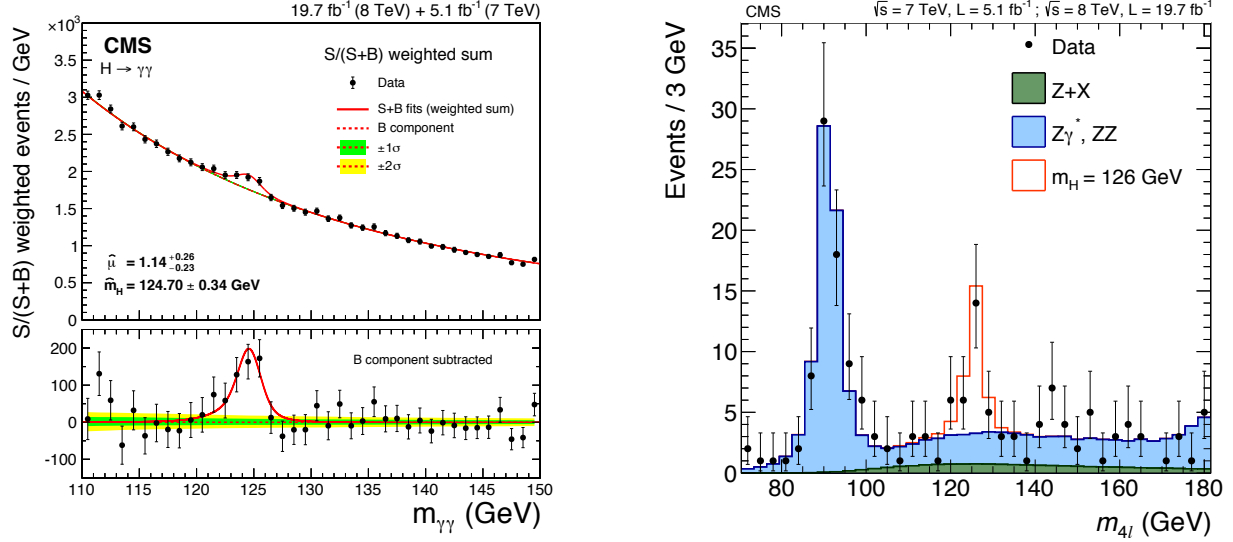


Figure 1.3: Reconstructed mass spectrum showing the Higgs resonance in the $H \rightarrow \gamma\gamma$ (left) and $H \rightarrow ZZ^* \rightarrow 4\ell$ (right) channels [47,48].

gram involves a quark loop through which the massless gluons couple to the colorless Higgs boson. The contribution to the loop varies as the square of quark-Higgs coupling and hence it is dominated by the top quark. The latest best predicted cross section at $\sqrt{s} = 13$ TeV is

$$\sigma_{ggF} = 48.61 \text{ pb }^{+4.27\%}_{-6.49\%} (\text{theory}) \pm 1.85\% (\text{PDF})^{+2.59\%}_{-2.62\%} (\alpha_s)$$

with an accuracy of three higher order corrections or next-to-next-to-next leading order (N3LO) in QCD and next-to-leading order (NLO) correction in EW interaction. Here α_s is the strong interaction coupling strength given by $\alpha_s = g_s^2/4\pi$. PDF refers to parton density functions (see Chapter 3).

- **Vector boson fusion (VBF):** This production mode has the second largest cross section at the LHC:

$$\sigma_{VBF} = 3.766 \text{ pb }^{+0.43\%}_{-0.33\%} (\text{scale}) \pm 2.1\% (\text{PDF} + \alpha_s)$$

at $\sqrt{s} = 13$ TeV, considering NNLO QCD and NLO EW corrections. The momentum transfer between the two scattering quarks being small, they appear as jets with

high rapidity in the final state. Further, the di-jet system has large rapidity gap and high invariant mass. Due to this unique event signature VBF production mode is useful to search for rare decay modes of the Higgs boson, such as $H \rightarrow \text{invisible}$, $H \rightarrow \mu\mu$ etc.

- **Associated production with a vector boson or Higgstrahlung (VH):** In this process the Higgs boson is emitted by W or Z bosons. The leptonic decay products of the vector bosons are useful to identify the events of interest from multi-jet backgrounds in case of hadronic decay modes of Higgs boson, e.g., $H \rightarrow b\bar{b}$. The production cross section is small compared to ggF and VBF:

$$\sigma_{WH} = 1.358 \text{ pb} \pm 0.51\% (\text{scale}) \pm 1.35\% (\text{PDF} + \alpha_s)$$

and

$$\sigma_{ZH} = 0.880 \text{ pb} {}^{+3.50\%}_{-2.68\%} (\text{scale}) \pm 1.65\% (\text{PDF} + \alpha_s)$$

at $\sqrt{s} = 13 \text{ TeV}$, considering NNLO QCD and NLO EW corrections.

- **Associated production with a top-quark pair ($t\bar{t}H$):** This is an important production mode of the Higgs boson where the top-Higgs coupling is involved in the tree level diagram, unlike in the loop in case of ggF. The decay products of the top quarks lead to a variety of final states, making it difficult to reconstruct. Further, the cross section is

$$\sigma_{t\bar{t}H} = 0.5065 \text{ pb} {}^{+5.8\%}_{-9.2\%} (\text{scale}) \pm 3.6\% (\text{PDF} + \alpha_s)$$

at $\sqrt{s} = 13 \text{ TeV}$ at NLO QCD+EW accuracy, roughly two orders of magnitude less than that of ggF production.

Feynman diagrams for the above processes are shown in Fig. 1.4.

The decay fractions of the 125.09 GeV Higgs boson are given in Table 1.2 along with the theoretical uncertainties. $H \rightarrow \tau\tau$ has been recently observed at the LHC [52] and it gives first direct evidence of Higgs boson coupling to fermions. The dominant decay mode of the Higgs boson is $H \rightarrow b\bar{b}$ with a branching ratio (BR) of 58.4%. Due to the hadronic activity in proton-proton (p-p) collisions, this decay mode may get swamped in with a

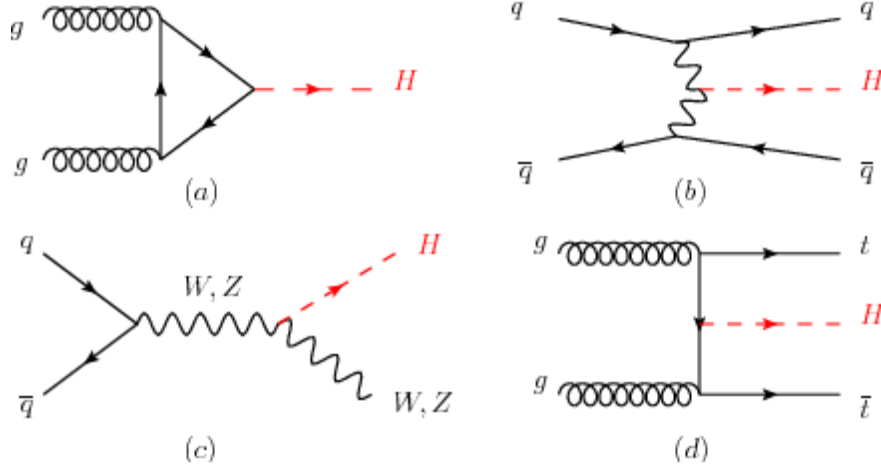


Figure 1.4: Main leading order Feynman diagrams contributing to the Higgs boson production at the LHC: (a) gluon-gluon fusion, (b) vector boson fusion, (c) Higgstrahlung, (d) associated production with top quark pair [25].

Table 1.2: Decay branching fractions of the Higgs boson with theoretical uncertainties [45].

Decay mode	BR	Th. uncertainty
$H \rightarrow \gamma\gamma$	2.270×10^{-3}	+1.73% -1.72%
$H \rightarrow ZZ^*$	2.641×10^{-2}	+0.99% -0.99%
$H \rightarrow WW^*$	2.152×10^{-1}	+0.99% -0.99%
$H \rightarrow \tau\tau$	6.256×10^{-2}	+1.17% -1.16%
$H \rightarrow b\bar{b}$	5.809×10^{-1}	+0.65% -0.65%
$H \rightarrow c\bar{c}$	2.884×10^{-2}	+1.20% -1.20%
$H \rightarrow Z\gamma$	1.541×10^{-3}	+5.71% -5.71%
$H \rightarrow \mu\mu$	2.171×10^{-4}	+1.23% -1.23%

large QCD background. $H \rightarrow b\bar{b}$ decays have also been observed during Run 2 [53, 54]. The next largest BR is for $H \rightarrow WW^*$ at 21.4%. This decay leads to multi-leptonic final states 55% of the time, making it useful to study Higgs boson productions with relatively lower background. The $H \rightarrow c\bar{c}$, $H \rightarrow Z\gamma$ and $H \rightarrow \mu\mu$ have limited sensitivity with currently accumulated data.

1.4.1 Higgs couplings measurement

In the absence of any specific knowledge about beyond SM physics, it is assumed that effectively the Higgs couplings to vector bosons and fermions can be modified. At the LHC, the measurement of the couplings is performed using the simplified kappa (κ) framework [46, 55], where scale factors or coupling modifiers are used to determine deviations from SM couplings. The coupling modifiers appear in the expressions for individual production and decay modes with zero-width approximation:

$$\kappa_i^2 = \frac{\sigma_i}{\sigma_i^{SM}}, \quad \kappa_f^2 = \frac{\Gamma_f}{\Gamma_f^{SM}} \quad (1.11)$$

where σ_i denotes cross section and Γ_f denotes decay width for i^{th} production mode and f^{th} decay mode. Using this framework, the Higgs boson coupling modifiers are obtained from the Run 1 LHC data combining results from ATLAS and CMS experiments [56]. Figure 1.5 shows the observed best-fit values and 65% confidence level (CL) contour for fermion and vector boson coupling modifiers, κ_F and κ_V respectively, obtained from different decay modes of Higgs boson. The fit allows negative values of κ_F if no assumption on its sign is made, also it has large uncertainty for individual decay modes. By convention κ_V values are considered to be positive. The best fit value gives $\kappa_F > 0$ which is driven by the large asymmetry between the positive and negative coupling ratios in case of $H \rightarrow \gamma\gamma$. Results from LHC Run 1 indicate that couplings of the Higgs boson to W and Z bosons are compatible with the predictions of SM [56]. In contrast, Higgs boson to fermion couplings have been constrained only with large uncertainties.

1.4.2 The top-Yukawa coupling

Top quark being the heaviest fermion, has the largest Yukawa coupling, $y_t = \sqrt{2}m_t/v \sim \mathcal{O}(1)$. It is highly possible that top quark also plays a significant role in the EWSB due to its large mass [58–61]. Deviations of y_t from the SM prediction would unambiguously indicate presence of BSM physics.

The determination of the sign of the top-Yukawa is an important handle to test several BSM theories, including MSSM. In general the two Higgs doublet model (2HDM) [62, 63]

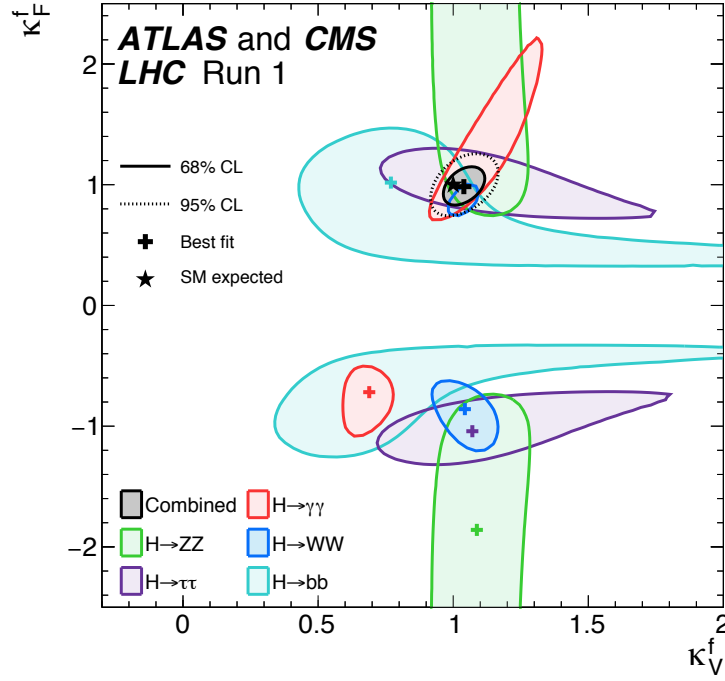


Figure 1.5: Run 1 combination of Higgs coupling modifiers measurement [56]. Negative log-likelihood contours are shown at 65% CL in the κ_F vs κ_V plane for the combination of ATLAS and CMS showing individual decay channels (colored). The best-fit combination of all κ is shown in black with 65% and 95% CL contours.

extensions of the SM allow several possibilities of the Higgs boson to fermion couplings. The couplings are such that a fermion can couple to only one of the doublets to avoid existence of tree level flavor changing neutral currents (FCNCs) in the theory. For example, first doublet (ϕ_1) cannot couple to any fermions whereas the second doublet (ϕ_2) does (Type I). Another valid choice may be ϕ_1 coupling to down-type quarks and leptons, ϕ_2 coupling to up-type quarks (Type II). The second model is the required choice of MSSM theories, as discussed in a previous section. The physical mass eigenstates of a 2HDM model are given by:

$$\begin{aligned} H &= \sqrt{2}[(\text{Re}\phi_1^0 - v_1)\cos\alpha + (\text{Re}\phi_2^0 - v_2)\sin\alpha] \\ h &= \sqrt{2}[-(\text{Re}\phi_1^0 - v_1)\sin\alpha + (\text{Re}\phi_2^0 - v_2)\cos\alpha] \end{aligned} \quad (1.12)$$

where α is the mixing angle between the CP even Higgs bosons. The couplings of the

light (heavy) higgs boson, h (H), with W and Z is the same as that in SM multiplied by a factor $\sin(\beta - \alpha)$ ($\cos(\beta - \alpha)$). The couplings of fermions to the scalar and pseudoscalar Higgs bosons relative to the value m_f/v are given in Table 1.3.

Table 1.3: Couplings of the fermions to the scalar and pseudoscalar Higgs bosons in 2HDM model, relative to the value m_f/v . Relative couplings to charged leptons are same as that for the down type quarks.

	Type I 2HDM			Type II 2HDM		
	h	H	A	h	H	A
Up type quarks	$\cos \alpha / \sin \beta$	$\sin \alpha / \sin \beta$	$\cot \beta$	$\cos \alpha / \sin \beta$	$\sin \alpha / \sin \beta$	$\cot \beta$
Down type quarks	$\cos \alpha / \sin \beta$	$\sin \alpha / \sin \beta$	$-\cot \beta$	$-\sin \alpha / \cos \beta$	$\cos \alpha / \cos \beta$	$\tan \beta$

Assuming $0 \leq \beta - \alpha \leq \pi$, the light CP-even Higgs boson h is SM like in the limit $\cos(\beta - \alpha) \rightarrow 0$. Depending on the $\tan \beta$ value, fermion couplings can be enhanced or suppressed for a given $\cos(\beta - \alpha)$.

Inverted sign of Yukawa couplings are possible with some choice of the parameters α and β in the 2HDM theory [64]. By convention the coupling of h to vector bosons is always non-negative: $\sin(\beta - \alpha) \geq 0$. In the Type II 2HDM the relative fermion couplings can be re-written for convenience (using 3rd generation quark notation) as:

$$\begin{aligned}
 g_{ht\bar{t}} &= \frac{\cos \alpha}{\sin \beta} = \sin(\beta + \alpha) + \cot \beta \cos(\beta + \alpha), \\
 g_{hb\bar{b}} &= -\frac{\sin \alpha}{\cos \beta} = -\sin(\beta + \alpha) + \tan \beta \cos(\beta + \alpha).
 \end{aligned}
 \tag{1.13}$$

For the case of $\sin(\beta + \alpha) = -1$ the $ht\bar{t}$ coupling normalized to the SM value is -1 , whereas $g_{hb\bar{b}} = +1$. In this limit $\sin(\beta - \alpha) = \cos 2\beta$ which implies that the inverted or wrong sign of the coupling is only possible for $\tan \beta < 1$. Similarly, for Type I 2HDM both $g_{ht\bar{t}}$ and $g_{hb\bar{b}}$ are equal to -1 if $\tan \beta < 1$. Even though it has been pointed out that only inverted down-type couplings in Type II 2HDM are phenomenologically viable, global fits to experimental data still leave some room for the inverted up-type couplings [65,66]. This motivates the precise determination of both the magnitude and the sign of Yukawa couplings from all possible processes involving the Higgs boson.

1.4.3 Associated production of the Higgs boson with a single top quark

One of the significant highlights of the LHC Run 2 is the observation of Higgs boson production with top quark pair [67,68], providing direct access to the top-Yukawa coupling. Even though y_t is involved in Higgs boson production in ggF process and in $H \rightarrow \gamma\gamma$ decay through particle loops, any deviation from SM may be compensated by unknown particles in the said loops. In case of $t\bar{t}H$ production, the cross section varies as y_t^2 , i.e., it is only sensitive to the magnitude of the coupling. Hence in the scenario of inverted top coupling (ITC) where $\kappa_t = -1$, $t\bar{t}H$ cross section will not change, although some indirect effect may be observed in inclusive $H \rightarrow \gamma\gamma$ final state measurement where the interference is also present.

Apart from the $t\bar{t}H$ production, y_t can also be directly probed with the single top associated production of the Higgs boson (tH). This process is sub-dominant at the LHC; however it brings to light important information about the relative sign of the top-Higgs and W-Higgs couplings due to an interference effect between leading order diagrams, discussed in Chapter 5.

The single top associated production of the Higgs boson can resolve this sign ambiguity and also provide a model independent search of new physics in the top-Higgs sector. Due to the subtle interplay of the leading order diagrams it is possible to study the Higgs boson CP properties and FCNC scenarios as well (more detail in Chapter 5), making this channel very much interesting for analysis. This thesis is based on the studies of this rare production process with p-p collision data collected at the Run 2 in 2016 and 2017 by the CMS detector, with $\sqrt{s} = 13$ TeV. From Run 1 data at $\sqrt{s} = 8$ TeV a 95% CL upper limit on t-channel tH production in the ITC scenario was extracted for the first time [69], which yielded an upper limit twice that of the expectation. With the increase in $\sqrt{s}$ from 8 TeV to 13 TeV in Run 2, the cross section increases from 18 fb to 71 fb for the t-channel, thus more number of signal events are accessible. The Run 2 analysis also employs more sophisticated techniques for signal extraction, resulting in a more stringent upper limit on ITC production of tH process as well as restricting the allowed range of κ_t . Further studies in this thesis include probe of CP phase scenarios for the Higgs boson.

Chapters [2](#) and [3](#) of this thesis briefly discuss the experimental setup of the LHC and CMS detector. Work done on trigger and reconstruction performance studies as part of service work for CMS collaboration is described in Chapter [4](#). Chapter [5](#) gives outline of the tH analysis and detailed results from studies with multi-lepton final states are given in Chapter [6](#). The work has been carried out in collaboration with others, in fact a parallel analysis exists with the 2017 data employing matrix element method techniques that arrives at a similar result. In the case of 2016 data, significant contribution is made toward studying the di-lepton channel. In Chapter [7](#) the combination results with other final states and future prospect of the analysis in the context of high luminosity LHC (HL-LHC) are discussed. The conclusions and final results are summarized in Chapter [8](#).

Chapter 2

The LHC and the CMS experiment

Historically, the first experimental observation of a sub-atomic particle was made when high voltage difference between two electrodes caused unknown rays to be emitted that had electromagnetic properties [70]. In the early days of particle physics, spontaneous decays of radioactive sources were used to produce low energetic particles and study various properties of matter, e.g., the discovery of the neutron. It was becoming clear that in order to understand finer structures, higher energy particles would be needed since they correspond to effective probe of very short wavelengths in the quantum realm. J. Cockcroft and E. Walton were the first to attempt *splitting the atom* using an accelerated proton beam in 1932 [71]. Although extra terrestrial sources of highly energetic particles can be utilized, neither the energy nor the flux can be controlled. Systematic study using such cosmic particles will require investment of longer time. This motivated the rapid growth in accelerator physics – Van de Graaf generator, cyclotron, linear accelerators, cavity resonators etc. enabled access to TeV energy ranges in laboratories. There has been remarkable technological advancement in particle detectors as well, from hand-counting scintillations on a ZnS screen to real-time digitized electronic pulses in computers.

2.1 The Large Hadron Collider

The LHC, located at the Conseil Européen pour la Recherche Nucléaire (CERN) in Geneva, Switzerland, is the most powerful particle collider today which uses superconducting

magnets to accelerate proton bunches through radio-frequency (RF) cavities. The proton beams are accelerated inside a 27 km long ring, designed to reach up to 7 TeV energy per beam and peak instantaneous luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ in a collision. Particles from two beams travel in opposite directions and collide at four interaction points where the rings intersect. The first beam was injected in the LHC in 2008 with first collisions occurring in 2009, quickly surpassing the record of highest collision energy of $\sqrt{s} = 1.96$ TeV previously held by the Tevatron accelerator in Fermilab in that same year. The high energy run in 2009 were recorded with $\sqrt{s} = 2.36$ TeV. The collision energy was increased to 7 TeV in 2010 and to 8 TeV in 2012. This data taking period spanning 2008–2013 is referred to as Run 1 of the LHC. Run 2 started with 13 TeV collision energy in 2015 and a record amount of luminosity is delivered till 2018.

Figure 2.1 shows the schematic of the CERN accelerator complex. The life of protons at the LHC starts from a bottle of hydrogen gas. In the first part of the accelerator electric field in a duoplasmatron strips electrons from the hydrogen atoms and produces protons. They are then accelerated to 50 MeV using a linear accelerator (LINAC2) and injected into Proton Synchrotron Booster (PSB). The booster accelerates the protons to 1.4 GeV and injects into the Proton Synchrotron (PS) where they reach an energy of 25 GeV. Protons then enter the Super Proton Synchrotron (SPS) and accelerate to 450 GeV after which they are fed into the LHC ring. Two proton beams circulate in clockwise and anti-clockwise directions in ultra-high vacuum (10^{-13} atm) to reach up to 6.5 TeV energies each and then brought to four pre-determined collision points. Four major experiments operate at each of the points: A Large Ion Collider Experiment (ALICE) [72], A Toroidal LHC Apparatus (ATLAS) [73], the Compact Muon Solenoid (CMS) [74] and the Large Hadron Collider beauty (LHCb) [75] experiment.

An important aspect of the accelerator is the magnetic field. The LHC has 9593 magnets in total. Largest of them are the 1232 superconducting dipole bending magnets operated at very low temperature of 1.9 K (-271.3°C). In addition there are correcting magnets with larger number of poles in order to focus the beam. RF cavities are used to actually accelerate the protons and maintain constant beam energy. Energy gained per second by a proton is 0.18 TeV.

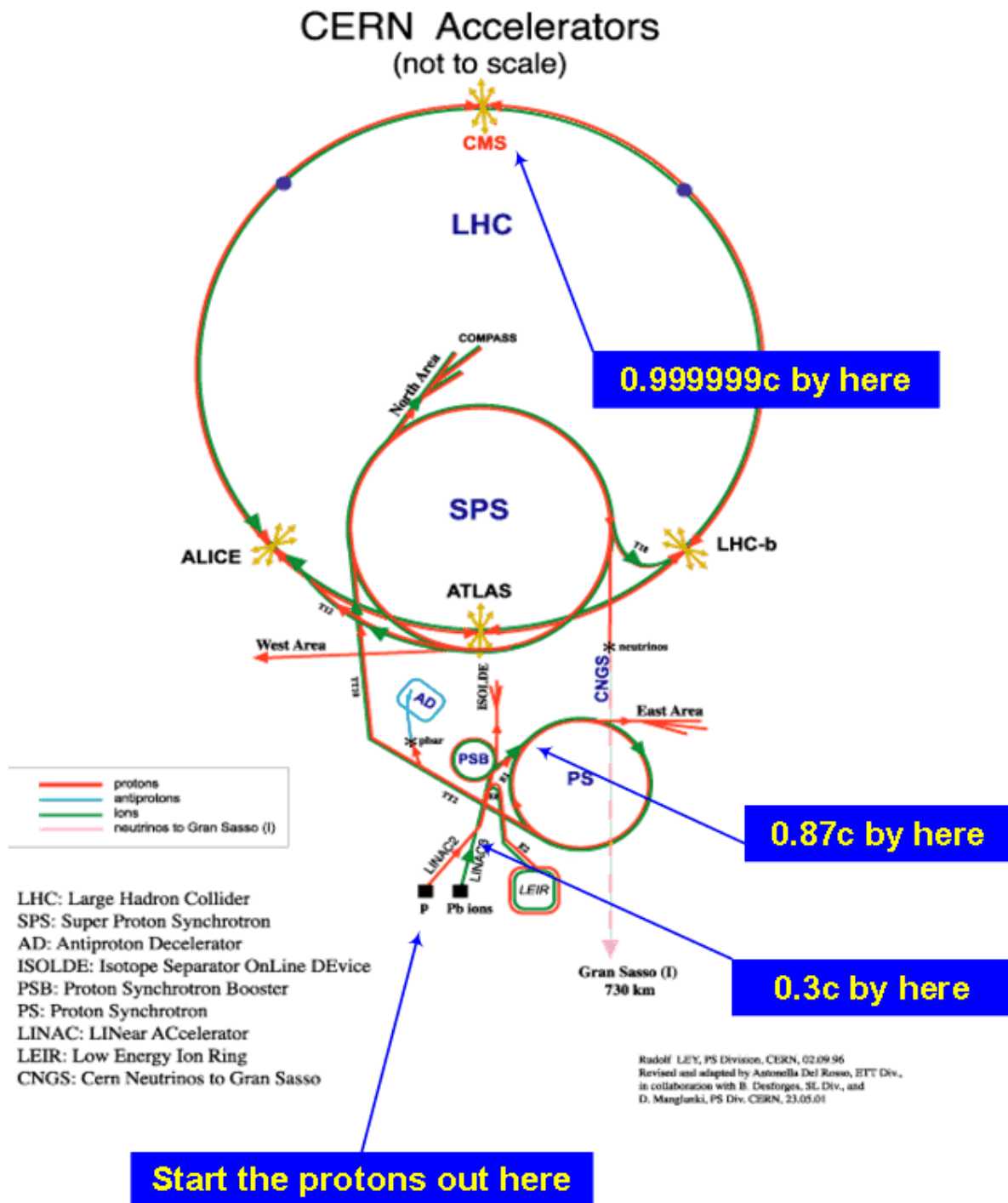


Figure 2.1: Overview of CERN accelerator complex.

The performance of the collider depends on the design parameters, most important quantity being the instantaneous luminosity (L) which is a measure of the collision rate:

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

where N_1 and N_2 are the number of particles in each colliding bunch, f is the frequency of revolution, N_b is the number of colliding bunches per beam, σ_x and σ_y are the bunch spread in x , y directions, assuming Gaussian beams. Typical longitudinal width of the beam is 3-4 cm. The integrated luminosity $L_{tot} = \int L dt$ corresponds to the total data collected and is related to the number of events produced for a given process of cross section σ :

$$N = L_{tot}\sigma. \quad (2.2)$$

At the LHC, $N_1 = N_2 = 1.2 \times 10^{11}$, $f = 11245$ Hz, $N_b = 2808$ were chosen as the design parameters. Quadrupole magnets are used to squeeze the beams to smallest possible size to have large L . During Run 2, the measured values of σ_x and σ_y are $14 \mu\text{m}$ and $12.5 \mu\text{m}$ respectively and peak instantaneous luminosity reached a value $\sim 2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, twice larger than what was designed. L_{tot} at the end of 2018 is $\sim 150 \text{ fb}^{-1}$. By design, minimum separation between two proton bunches is 7.5 m or 25 ns, leading to maximum bunch crossing rate of 40 MHz. In practice, bunch crossing rate : $N_b \times f \simeq 32 \text{ MHz}$.

Figure 2.2 shows L_{tot} as a function of time and number of interactions per crossing or *pileup* distributions for each year of the LHC operation, as seen by CMS. The number of pileup interactions (N_{PU}) are calculated considering the total inelastic cross section (σ_{in}^{pp}) at a particular energy: $N_{PU} = (L \times \sigma_{in}^{pp}) / (N_b \times f)$ with varying L over time.

In addition to p-p collisions, the LHC also accelerates Lead ions using Low Energy Ion Ring (LEIR). Both Pb-Pb and Pb-p collision data were recorded during Run 1 and 2, used to study the physics of quark-gluon plasma.

2.2 The Compact Muon Solenoid

In this thesis, analysis of the LHC Run 2 collision data recorded by the CMS detector is presented. CMS is a general purpose detector, similar to ATLAS, that records many

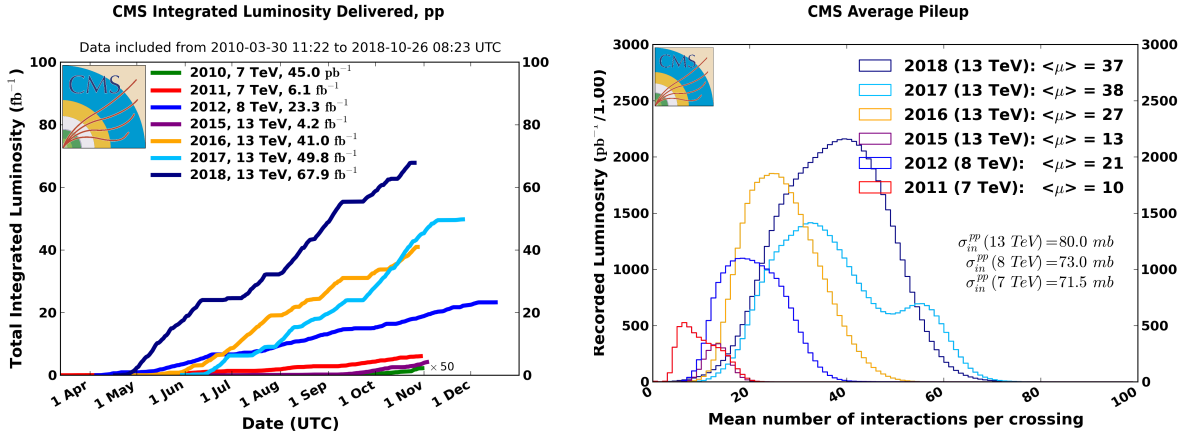


Figure 2.2: Total integrated luminosity (left) and average pileup distribution (right) recorded by CMS in different years from 2010 to 2018.

different kinds of events. The data are analyzed to test the validity of the SM and look for new physics phenomena in a multitude of physics processes.

2.2.1 Overview

The CMS detector is located near Cessy, France inside a large underground cavern approximately 100 m below surface. It is an almost hermetic detector built around a large solenoidal superconducting magnet of internal diameter 6 m and length 12.5 m. The magnetic field with axis aligned with the beam can reach up to 3.8 Tesla which essentially helps detect charged particles with very high momenta by bending the trajectories. The detector design is based on the shape of the superconducting magnet, with tracking sub-systems placed at regions with the highest transverse field. A steel yoke is used to return the magnetic flux which also supports the body of the whole detector.

The detector itself is cylindrical in shape, 21.6 m long and 14.6 m in diameter. The central region is called *barrel* and disks covering the area near the beam pipe are called *endcaps*. The name of the detector comes from the compact design which puts most of the calorimetry inside the magnet and dedicated muon sub-detectors placed outside the magnet. Unlike other detectors of the LHC, CMS was built on the surface and pieces were lowered inside the cavern and assembled. This feature is exploited to easily disassemble

and access different parts of the detector for maintenance and upgrade. Figure 2.3 shows the schematic diagram of the full detector.

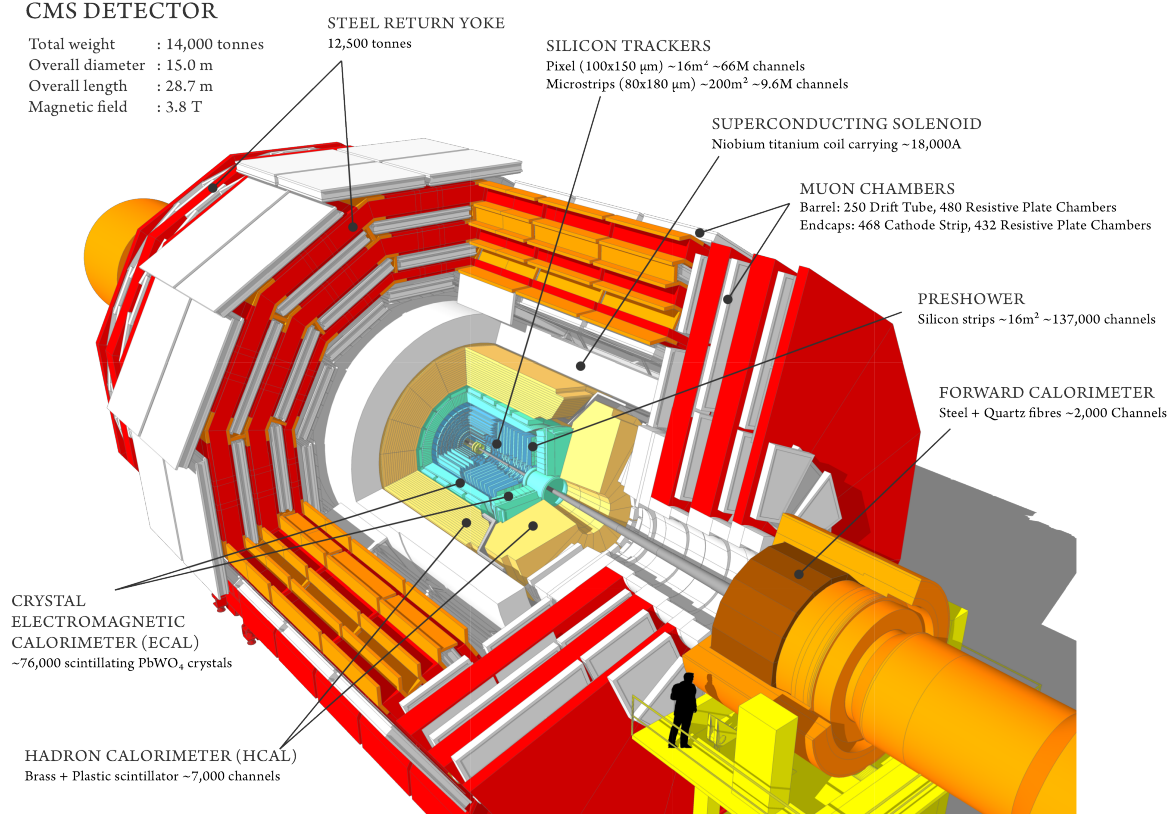


Figure 2.3: Schematic view of the CMS detector.

2.2.2 Particle kinematics

Right handed co-ordinate system is used in CMS with the origin at the centre of the detector, y axis points vertically upward and x axis towards the centre of the LHC ring. The z axis points along the anti-clockwise direction of the beam. The azimuthal angle ϕ is measured from the x axis in the x - y plane and the radial co-ordinate is denoted as r . The polar angle θ is measured from the z axis. *Rapidity* (w) of a particle having four-momentum $p = (E, \vec{p})$ is defined as:

$$w = \ln \left(\frac{E + |\vec{p}| \cos \theta}{E - |\vec{p}| \cos \theta} \right). \quad (2.3)$$

Rapidity difference is invariant under Lorentz transformation along the z axis, which makes it more useful compared to θ for a hadronic machine. However, for highly energetic particles ($E \gg m$) coming from a collision, rapidity reduces to *pseudorapidity* defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \quad (2.4)$$

Often angular co-ordinates of a particle are denoted in terms of (η, ϕ) and angular separation between two particles as:

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

The 4-dimensional space-time vector is thus expressed as (E, p_T, η, ϕ) where p_T is the transverse component of $\vec{p}$ in the x - y plane. For collisions of proton beams moving along z direction, the initial longitudinal momentum of the interacting sub-atomic particles, which carry some fraction of the proton energy, are not known. However, the initial momentum in the transverse plane is known to be zero. Thus to characterize highly inelastic events, p_z of the particles is not as useful as p_T in terms of studying event kinematics.

2.2.3 Sub-detectors

CMS employs different types of detectors to identify and measure different particles. Sub-systems are placed in radially outward layers with various services forming dead regions in-between. Closest to the interaction point the silicon pixel and strip detectors are positioned constituting the tracker sub-system, followed by lead-tungstate crystal electromagnetic calorimeter (ECAL) and a brass/scintillator sampling hadronic calorimeter (HCAL). Muon detectors are the outermost sub-system since muons are mostly minimum ionizing that traverse through the detector material without losing much energy.

Inner tracking system

In order to accurately determine the positions of interaction vertices and reconstruct trajectories (tracks) of charged particles before they lose their energy, the tracker [76] is placed at the innermost position of CMS. Momenta of the emerging particles are determined using the curvature of the tracks, hence minimization of interaction with the

detector material is required. Since near the interaction point there is maximum flux of energetic particles, a radiation hard detector material and electronics is required. As protons collide in every 25 ns, good timing resolution is required to identify collisions from the same bunch with that from previous bunch. Silicon semiconductors were chosen as they perfectly fit the requirements of being extremely fast and radiation hard. Two types of detectors are used by CMS in the tracker:

- **Pixels:** These are placed closest to the interaction point. Each pixel has the area $100 \times 150 \mu\text{m}^2$ and thickness of $285 \mu\text{m}$. There are 66M pixels arranged in 1440 modules. The initial design of the pixel detector had three 53 cm long coaxial layers (BPIX) at the distances 4.4, 7.3 and 10.2 cm in the barrel ($|\eta| < 0.9$) and $2 + 2$ disks (FPIX) in the endcap ($0.9 < |\eta| < 2.5$) region. From 2017 onwards number of barrel layers have increased to four, at radii 3.0, 6.8, 10.2 and 16.0 cm centred at the interaction point [77]. Number of disks in each endcap is also increased to three. This upgrade gives *four-hit* coverage which is an improvement to the track reconstruction quality as previously only three hits were possible. Each disk is divided into 24 blades, rotated by 20° in a turbine geometry to improve the resolution. A single hit in the pixel system can be resolved within 10-20 μm . Operating temperature for pixel detectors were changed from 0°C to -10°C in Run 2 to handle higher instantaneous luminosity.
- **Strips:** Around the pixels structure, the silicon microstrip detectors or strips are arranged. Within the barrel, strips are divided into inner barrel (TIB) and outer barrel (TOB) regions covering radial distances 20 to 55 cm and 55 to 116 cm respectively. TIB has four layers with varying pitch. In TOB the particle flux is lower and strips of larger pitch are arranged in six layers. The thickness is also more to maintain good signal to noise ratio in the larger strips. In both TIB and TOB the first two layers provide stereo measurement, i.e., hits in both $r - \phi$ plane and z direction can be resolved. The endcaps (TEC) contain nine disks between $|z|$ of 124 and 282 cm. Three inner disks (TID) are located in the gap between TIB and TEC, spanning the length $58 < |z| < 124$ cm. The pitch sizes and thickness of strips in each of these

regions are listed in Table 2.1. Position resolution in $r - \phi$ is 13-38 μm in TIB and TID, 18-47 μm in TOB and TEC. Operating temperature for strips were changed from $+4^\circ\text{C}$ to -15°C in Run 2 to deal with radiation damage.

Table 2.1: Types of silicon strip detector used in CMS in Runs 1 and 2.

Part	No. of detectors	Thickness (μm)	Mean pitch (μm)
TIB	2724	320	81/118
TOB	5208	500	81/183
TID	816	320	97/128/143
TEC	2512	320	97/126/128/143
TEC(2)	3888	500	143/158/183

The tracker layout as in Run 1 can be seen in Fig. 2.4.

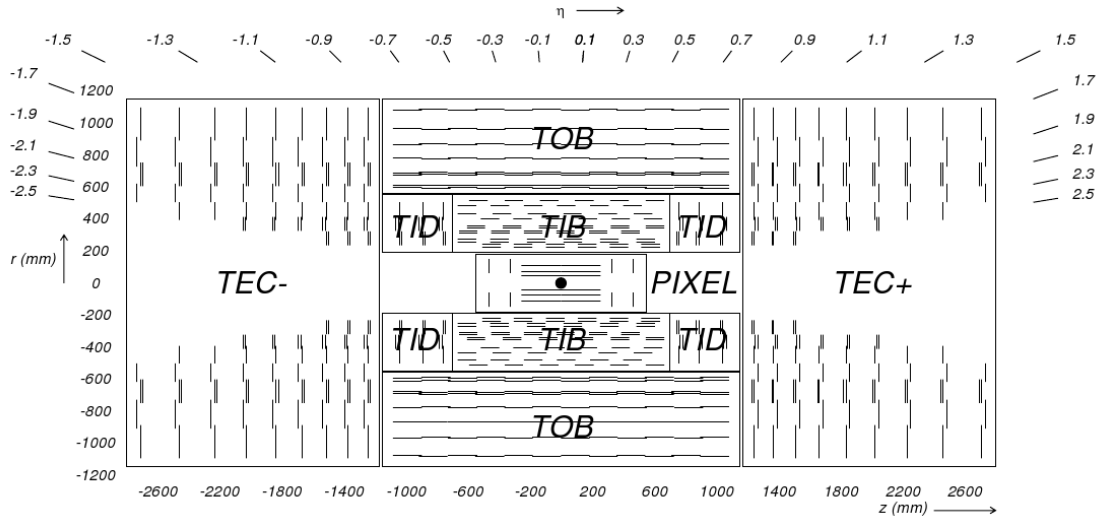


Figure 2.4: Schematic view of the inner tracking system in $r - z$ plane during Run 1. In Run 2 pixel sub-system layers were adjusted to include a fourth layer of silicon detectors.

Electromagnetic calorimeter

ECAL [78] is located just outside the tracker, designed to identify e/γ objects. It is made of ~ 76000 lead-tungstate (PbWO_4) crystals and homogeneous in nature. PbWO_4 has a

small Moliere radius (2.2 cm) which makes high granularity (lateral) possible in this sub-detector. Short radiation length (0.89 cm) of this material helps contain the longitudinal shower development, crucial for calorimetry. The response time to radiation is also very fast ~ 15 ns. Overall, these crystals are economical both in terms of production cost and application purpose, having reasonable radiation hardness as well. When an electron or photon passes through the crystal, electromagnetic interaction causes shower of secondary particles that produce scintillation light. The light is collected by photo-sensors that convert it to electrical pulse. The chosen photo-sensor for the ECAL system in CMS is the Avalanche Photo Diode (APD) in the barrel region and Vacuum Photo Triode (VPT) in the endcap. VPTs are specially developed for the CMS experiment, able to withstand high magnetic field and resilient to radiation.

The barrel (EB) section ($|\eta| < 1.479$) is positioned at an inner radius of 129 cm. The crystals are of dimension $2.2 \times 2.2 \times 23$ cm³ pointing toward the interaction point, tilted slightly by an angle of 3° to avoid particles escaping through the gaps between crystals. The granularity is 0.0175×0.0175 in $\eta - \phi$ plane corresponding to an angular width of 1° . The endcap (EE) extends to $|\eta| = 3$, however precision electromagnetic measurement is only possible up to $|\eta| = 2.6$. The crystals here have size $2.86 \times 2.86 \times 22$ cm³ and located at a distance of 314 cm from the interaction point. Crystals are preceded by a silicon based fine grained detector called *preshower* in the region $1.653 < |\eta| < 2.6$, intended to resolve prompt photons from pions and help in electron identification against minimum ionizing particles. It is a sampling calorimeter with lead radiators and silicon strip sensors with a total thickness of 20 cm. The high granularity of this sub-detector improves also the position determination of the electrons and photons. The operating temperature of the ECAL is $+18^\circ\text{C}$, while the preshower is operated at -5°C .

The resolution (σ) of the ECAL is given as a function of energy (E) deposited in an array of 3×3 crystals:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (2.6)$$

where, S is the stochastic term, N the noise term and C the constant term, measured values of which are $S = 2.8\% \sqrt{\text{GeV}}$, $N = 0.12$ GeV and $C = 0.3\%$ [79].

Hadronic calorimeter

The evidence of strongly interacting particles such as quarks and gluons are picked up by the HCAL [80] by measuring hadronic activity. Like ECAL, HCAL is also placed inside the magnet, the third sub-detector layer from the interaction point. It is made of alternating layers of absorbing (brass, 70% Cu and 30% Zn) and active (plastic scintillator) materials between the radial distance of 1.77 m and 2.95 m. The HCAL is composed of two main sections: barrel (HB) within $|\eta| < 1.3$ and endcap (HE) within $1.3 < |\eta| < 3.0$. The absorber layers are 5 cm thick in the barrel and 8 cm thick in the endcap. The innermost and outermost absorber layers are not brass, but 7 cm thick stainless steel which also provide structural strength. Active layers are 4 mm thick and contain wavelength shifting fibres (WLS) to collect the scintillation light and carry it to the readout electronics. A forward calorimeter (HF) between $3.0 < |\eta| < 5.2$ is also part of the HCAL, increasing the coverage for jets at high rapidity and crucial for measuring missing transverse energy. The active material for HF are quartz fibres embedded in iron absorber.

The nuclear interaction length (λ_I) of the brass absorber is 16.42 cm. Sampling depth of the HCAL varies between $5.8\lambda_I$ and $10.6\lambda_I$ within the barrel, ECAL crystals adding another $1.1\lambda_I$. Since this is not sufficient to contain the hadronic shower, an additional layer of scintillator tiles called outer HCAL (HO) is placed outside the magnet up to $|\eta| < 1.3$. Including HO, the total depth of the calorimeter is extended up to $11.8\lambda_I$.

The granularity of HCAL is provided by the scintillator tiles which is 0.087×0.087 for $|\eta| < 1.74$. This corresponds to 5° granularity in ϕ , which is reduced to 10° for $|\eta| \geq 1.74$. Along η the granularity ranges between 0.09 to 0.35 within HE. In the HF, the sections are mostly 0.175×0.175 in $\eta - \phi$, except for $|\eta| > 4.7$ where the segmentation is 0.175×0.35 . The readout electronics of HB, HE and HO system were multi-channel hybrid photodiodes (HPD) during Run 1 and most of Run 2, they have been replaced in phases with silicon photomultiplier (SiPM) devices since 2013 [81]. Photomultiplier tubes (PMTs) are used in HF due to lower magnetic field in that region. Figure 2.5 shows the configuration of ECAL and HCAL sub-detectors.

The energy resolution for the hadronic activity of HCAL and ECAL combined is parametrized

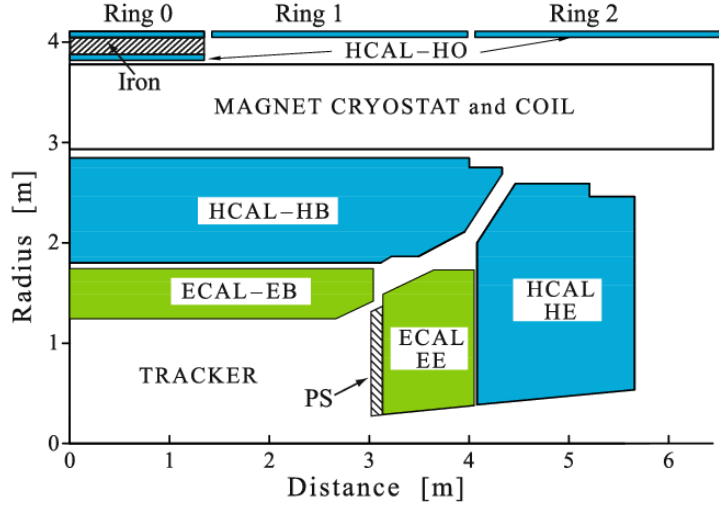


Figure 2.5: Location of the ECAL and HCAL around the magnet [82]. PS refers to the preshower detector.

as follows [83]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{84.7\%\sqrt{\text{GeV}}}{\sqrt{E}}\right)^2 + (7.4\%)^2 \text{ for HB, HE and HO}, \quad (2.7)$$

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{198\%\sqrt{\text{GeV}}}{\sqrt{E}}\right)^2 + (9\%)^2 \text{ for HF}. \quad (2.8)$$

Muon system

The muon detectors in CMS [84] are placed at the outermost layer, beyond the superconducting solenoid. As a result, the magnetic field experienced by the particles here are opposite in direction to that in the inner layers. This helps in independent momentum measurement for the muons once within the tracker and the second time in the muon system. The detectors are embedded inside the iron return yoke for the magnetic field. The iron absorber helps to reduce hadronic shower leakage from the calorimeters and only minimum ionizing particles like muons can reach the outermost layer. The structure of the muon sub-system includes five separate wheels in the barrel having four layers of detectors, and four independent disks in each endcap. Detectors in the fourth disk were installed during the long shutdown period of 2013-14 after Run 1. The arrangement of the detectors can be seen in Fig. 2.6.

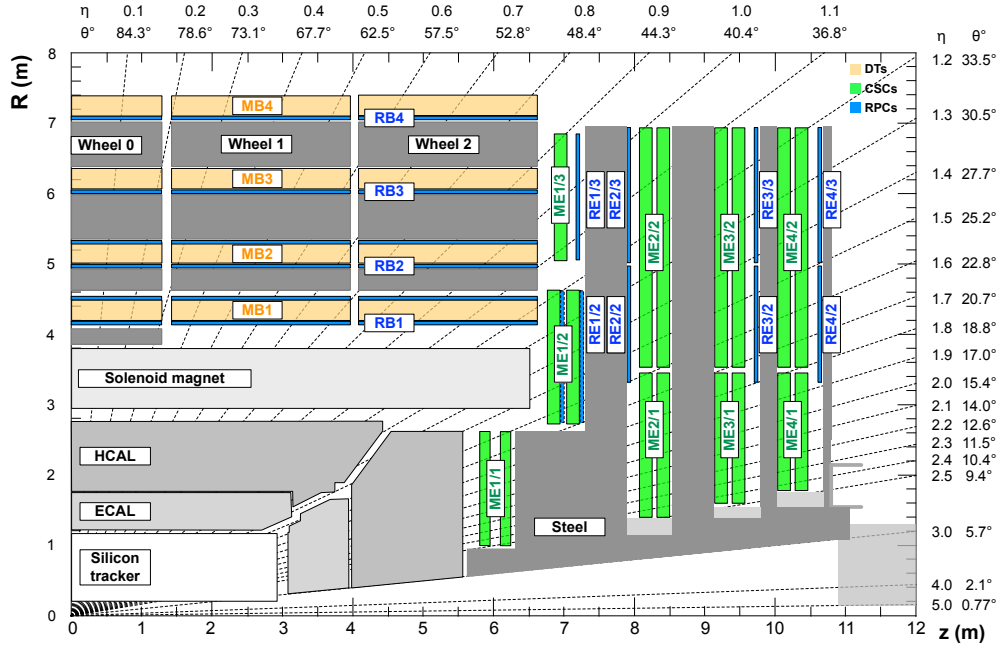


Figure 2.6: Schematic diagram depicting the CMS muon detectors (colored) in one quadrant as in Run 2.

Three types of gas-ionizing detectors are used in this sub-system:

- **Drift tubes (DT):** DTs are used in the barrel ($|\eta| < 1.2$) where the neutron flux is low, muon rates and residual magnetic field is small. They are organized in four stations, each station contains superlayers which consist of four chambers. First three stations have three superlayers each, two measuring the co-ordinates in the $r - \phi$ plane, one in the z direction. The outermost station does not contain the superlayer measuring the z co-ordinate. The four chambers in a superlayer are staggered by half-a-cell to provide dis-ambiguity in particle direction. This also provides time-tagging capability, with a resolution of few ns.
- **Cathode strip chambers (CSC):** In the endcap region ($0.9 < |\eta| < 2.4$) where the radiation and magnetic field are high, CSCs are used. There are four stations perpendicular to the beam line, each station containing six CSC chambers. Six layers of anode wires are sandwiched between radial cathode strips, providing position measurement in $r - \phi$ plane and η direction.

- **Resistive Plate Chambers (RPC):** RPC detector layers are placed in both barrel and endcap ($|\eta| < 2.1$), providing additional position measurement. However, its main advantage is the very fast timing response which is used to trigger signal events. There are six RPC layers in barrel, four layers in the endcap.

2.2.4 Trigger and data acquisition system

With an instantaneous luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ and a bunch crossing rate 40 MHz, about 10^9 p-p interactions happen every second at the collision points of the LHC, assuming total cross section is about 100 mb. Each fully reconstructed event having an average size of 1 MB requires a data archiving rate of 1 PB/s which goes beyond the current capacity of the electronics to transfer and store data. However, a large fraction of these events are not-so-interesting, hence a judicious selection is applied at the time of data acquisition itself to collect the events on time. Keeping these in mind, the CMS triggering and data acquisition system (*TriDAS*) [85,86] is designed to continuously process and record events with maximum possible rate.

Energetic particles produced in collisions fly out of the interaction point and traverse the detector layers. Along the way they deposit energy or produce accumulation of charge (hits) that are converted into electrical pulse by the front-end electronics. A fraction of this information is used to obtain a preliminary trigger decision (Level-1 or L1) only after which the full event information is forwarded to the back-end by the data acquisition (DAQ) system. In the next stage, a second trigger decision (high level trigger or HLT) reduces the event rate even further and the selected data is transferred to offline storage of Worldwide LHC Computing Grid (WLCG). The design parameters for the CMS trigger and DAQ system are listed in Table 2.2.

CMS detector was originally designed to study high p_T physics, mainly motivated to discover the Higgs boson and new physics at TeV energy scale. However, it was realized that the excellent particle detection and reconstruction capability can also be used in various other analyses involving B-mesons, quarkonia etc. Apart from searches for particles, there are also electroweak precision measurements (e.g., W , Z production cross-section) that can be performed. All of these require high efficiency of selecting trigger objects,

Table 2.2: Design parameters of the CMS trigger and DAQ system.

Parameter	Value
L1 rate	100 kHz
HLT rate	100 Hz
Event size	1 MB
Event builder bandwidth	100 GB/s
Data production	10 TB/day

mainly leptons and jets in both high and low p_T regimes. The system must also be able to trigger on different objects at once and be able to handle higher multiplicity of each object; this requirement is more aptly applicable in case of heavy-ion collisions which have a very busy environment. In addition a two-level trigger system of CMS experiment offers the flexibility to have overlapping triggers and lowering of thresholds with decreasing luminosities to adjust with the bandwidth of data transfer.

Trigger

The event filtering process at CMS is done in two steps to retain high efficiency of selecting interesting physics phenomena since the rejection factor is very large (10^6). At the first level, all collision data is stored in readout buffer or pipeline for $3.2 \mu\text{s}$. A global trigger decision is issued in the presence of local “trigger primitive” objects for photons, electrons, muons, and jets, using information from calorimeters and muon systems in a given $\eta - \phi$ region. This information can also trigger on the global sum of energy deposit or E_T and missing energy or E_T^{miss} . All the objects are required to pass a set of pre-determined p_T or E_T thresholds. The time taken by the L1 trigger decision is less than $1 \mu\text{s}$ and the final output rate is 100 kHz. The L1 trigger electronics consists of custom hardware components like Application Specific Integrated Circuits (ASICs), Field Programmable Gate Arrays (FPGAs) etc. housed in the counting room away from the cavern. Combined with the propagation delay through the electronics from the detector to the counting room and back, the total latency of the L1 decision is $3.2 \mu\text{s}$.

After receiving the L1 accept signal, the data enters front end readout buffers. An event builder “switch” transfers all information of one single event every $10\ \mu\text{s}$ to the HLT. This time is sufficient to store all front-end data to commercial electronics processors or computer farm in the counting rooms. Sophisticated algorithms similar to offline reconstruction are used, including the tracker information in addition to information from calorimeter and muon chambers. The algorithms, also called *paths*, look for specific event signatures and have multiple thresholds for flexibility. It also has access to L1 trigger decision and can make further complicated combinations of L1 reconstructed objects. For HLT paths leading to very high rates, events are *prescaled* i.e., one event is chosen at random from many similar events. The prescale rate is adjusted with luminosity, making it possible to study events with low energy objects. Typically HLT processing takes about 100 ms and the reduction factor is 10^3 . In the end the HLT system writes the data at the rate of several hundred events/s in mass storage, separating datasets with different signatures. The maximum capacity is 1 kHz, corresponding to reduction factor of 10^6 .

DAQ

The DAQ system of CMS takes care of the above functionality of reading and writing the collision data through the trigger system. Data is read from the front-end by the readout system that sends the data to Filter system or processors via switches or Builder network. The Filter system are the actual elements of the HLT farm. The flow of data is managed by the Event Manager and looks after the overall system synchronization. In addition, DAQ provides an interface to control and monitor the entire process of data-taking with individual access to sub-detectors. During the operation, it is the responsibility of the DAQ to maintain the deadtime of storing data below 2% of overall latency. It can also raise alarms to protect the detector from hardware failures and can automatically power off sub-systems in case of emergency situation.

2.2.5 CMS computing and software

The various datasets are stored in WLCG sites, primarily in “Tier-0” centre at CERN, supplemented by several “Tier-1” and “Tier-2” centres across the globe. Tier-0 resources store and reconstruct the primary collision data. These are then replicated in Tier-1 sites for further reconstruction with simplified data structures. The computing grid allows seamless access to any dataset and also provides resources to run simulation and analysis jobs. Most of the jobs are run in Tier-2 sites which provide user storage. Smaller clusters or “Tier-3” are provided to carry out final stage of the analysis and submit the jobs to the grid using an user interface. The overall collection of software used for analyzing the data is referred to as CMSSW, based on C++ and python programming languages. It consists of an application framework which are a collection of classes that describe the detector and event features. There are specific physics modules used for reconstruction while running the software and they can be easily replaced with different versions as required by the analysis. Finally, the output is integrated with physics type services like histograms, fits and other calculations.

The processed results of the trigger system is delivered in a particular format to analyzers using CMSSW, referred to as the *Event Data Model* or *EDM*. An “event” contains all information that was recorded in a single beam crossing and is defined by a class that represents the observed readout and inferred objects. Additional information about detector calibration, alignment and the configuration of software used for reconstruction of that event is also stored, called “metadata”. The EDM structure keeps the history of the previous reconstruction phases of a given event.

The analysis jobs can be run on WLCG clusters using CMS Remote Analysis Builder or CRAB that manages resource availability and storage across the sites. Within a Tier-3 site, moderately long interactive jobs can be managed through condor submission, which have taken over the batch processing method used till Run 2. The next chapter describes briefly the information structure of the collision events and simulations as presented to the analyzers.

Chapter 3

Simulation and reconstruction

The flow of information available for analysis is in the EDM format which is the result of interpreting the digitized signal into physics signatures. In CMS experiment this interpretation is done with the help of the Particle Flow (PF) algorithm [87]. This step is performed offline after HLT accept, combining detector layers to best identify physical objects like e , γ , μ , jets etc. In any analysis, observation is compared against SM expectation provided by Monte Carlo (MC) simulations having similar reconstruction and presented in an equivalent format to data. The comparison is done in terms of cross section on the basis of total event yield in a particular final state, as well as the *shape* of kinematic variable distributions. MC events are generated using the best physics knowledge of the particle interactions as well as the detector condition. An essential prerequisite of the physics interpretation is studying the detector performance and monitoring the calibration for better understanding of the experiment. In a well understood physics model the MC yields and shapes match to that of the data within *uncertainties* or *nuisances*. The total uncertainty has two components: statistical and systematic. The systematic uncertainties can be both theoretical (lack of higher order predictions of cross section, dependence on interaction energy scale etc.) or experimental (effect of pileup events, luminosity measurement etc.) and can have a variety of sources depending on the final states considered in an analysis.

3.1 Theoretical structure of collision events

In order to test the predictions of the SM, theoretical calculations need to be implemented in a certain way to compare with the observation. Given the initial state particles and the centre-of-mass energy, interactions take place according to the probability of a given process. The final state particles are the outputs with their kinematics dictated by the phase space. For hadron collisions the situation is more complex since the fundamental interaction takes place among the constituents of the hadrons, i.e. quarks and gluons, called partons. The *event generators* provide the four-momenta of final state stable particles, taking into account the intermediate phases, e.g. parton shower, hadronization, decay etc. For individual particles, the four-momentum in multi-dimensional phase space can be obtained via Monte Carlo method. This is primarily used to decide “cuts” or selection strategy to separate events of interest with high efficiency and reject background processes. The expected physics can then be compared to observation, after accounting for detector effects and calibration, necessary to draw a conclusion on the feasibility of the particle interactions considered.

3.1.1 Hard scattering

If the partonic interaction in a p-p collision involves large momentum transfer resulting in high energy (greater than few GeV) final state particles, it is referred to as a hard scattering process. At the LHC, fractional energy carried by the interacting partons can be as high as several TeV, enabling to test known interactions at these energy scales. The exact momentum information of the partons cannot be known but a probabilistic solution is given in terms of parton density functions or PDFs. The PDFs depend on the momentum fraction (x) carried by the parton and also the interaction scale energy (Q). They are extracted mainly from deep inelastic scattering experiment data at the hadron collision experiments at lower energy. The most accurate extrapolation to the LHC energies is given by Neural Network PDF or NNPDF [88] parametrization. The cross section of parton interaction can be calculated using the *matrix element* (ME) for the given transition which takes into account momentum transfer and couplings between initial and

final state particles, as well as interaction propagators. For a particular initial and final state, all possible Feynman diagrams need to be considered along with their interference effects. The inclusive cross section at collision energy $\sqrt{s}$ can be written as:

$$\sigma_{total}(s, Q) = \sum_{ij} \int_{x_1=0}^1 \int_{x_2=0}^1 f_i(x_1) f_j(x_2) \sigma_{partonic}(x_1, x_2, s, Q) dx_1 dx_2 \quad (3.1)$$

where Q is the energy scale of the interaction, x_1 and x_2 are the momentum fraction carried by the two initial partons, $f_i(x_1)dx_1$ denotes the probability that i^{th} parton carries a momentum fraction x_1 of the proton energy or the PDF. The partonic cross section $\sigma_{partonic}$ depends on the sub-process energy defined as $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$.

Cross sections calculated considering only the lowest order of the coupling constant of a particle interaction is termed as leading order or LO. However, initial, final state radiation and loop interactions consisting of higher orders of the coupling constant significantly add to the LO cross section and need to be taken into account in order to precisely determine rate of any process. This is achieved using perturbative method implemented in event generators, such as next-to-leading order (NLO), next-to-next leading order (NNLO) and so on. Within CMS, simulated samples are generated at LO or NLO accuracy using ME generators such as MADGRAPH [89], POWHEG [90] etc.

3.1.2 Parton shower and hadronization

In a similar manner to bremsstrahlung, quarks and gluons can lose energy by QCD radiation depending on the momentum. Gluons split into parton pairs and quarks emit soft gluons, in this way many particles are produced, a phenomena referred to as parton shower (PS). In MC simulations this is implemented using QCD evolution equations [91–93]. In case of final states with have more than two hard partons, the perturbative shower needs to be matched with the ME calculation to avoid double counting. There are dedicated algorithms used for this matching, e.g., MLM [94] implemented for LO interactions, FxFx [95] for NLO etc. An energy scale is chosen as the cut-off, partons having greater energy than the cut-off are simulated by the ME, otherwise they are treated as radiation in the PS.

Hadronization or fragmentation refers to the process where quarks and gluons in the final state produce the color singlet hadrons that are finally observed in the data. The evolution of energy scale is described by the Lund string model [96], where all outgoing partons are assumed to be connected by color field lines called strings. As the color forces between the outgoing partons increase with distance, the strings break apart to produce quark anti-quark pairs. Hadrons are produced as quarks and anti-quarks from adjacent string break combine to form color neutral particles. The combined energy of the many hadrons produced in such a way correspond to the initial energy scale of the partonic interaction Q . Parton shower and hadronization is implemented using PYTHIA [97], HERWIG [98] etc. in case of CMS analyses.

3.1.3 Underlying event and pileup

Other partons of the colliding protons not taking part in the hard interaction, appear as beam remnants. These partons may hadronize, or undergo soft (low Q) interactions known as *underlying event* (UE). According to different models adopted by generators, e.g., PYTHIA or HERWIG, the features of UE may vary significantly. Thus understanding of these soft QCD interactions is mostly driven by experimental observations. In order to better match the simulation with data, significant effort is given to *tune* the parameters of existing models [99].

Event pileup refers to overlap of more than one interaction within the detector integration time. Two types of pileup are possible: in-time and out-of-time. In-time pileup are due to interactions in the same bunch crossing, whereas out-of-time pileup are tracks and energy deposits from earlier bunch crossing still remaining in the detector due to timing resolution effects. Both of these pileups in event generation are controlled by configuration modules in CMSSW which take as input the mean number of primary vertices during the data taking period. Pileup events are separately generated and added to the hard interaction for realistic simulation of the p-p collisions at LHC.

3.1.4 Detector simulation

Detector effects may change the particle kinematics to a significant extent as it introduces matter-particle interactions. To compare the simulated events from event generators to the observed data, the effect of various detector material needs to be taken into account. An accurate description of the CMS detector is implemented using the GEANT4 package [100]. Energy and momentum resolutions are *smeared* with detector effects and electronics signals within readout sensors are replicated. The output of the detector simulation step is simulated hits or “SimHits” which undergo *digitization* where the detector electronics response is reproduced. In the next step, energy deposits are converted into charges and shaped into a waveform similar to electronics signal, noise is also added. Trigger primitives required for L1 is generated for calorimeter and muon chamber energy deposits. The digitized MC events are then converted to the raw format and undergo HLT reconstruction to arrive at the same structure of information as in the recorded collision data in the CMSSW EDM framework.

3.2 Event reconstruction

For both collision data and simulated events, the same flow of reconstruction is undertaken for CMS analyses. Figure 3.1 shows pictorially the steps to arrive at the analysis level format of the data and simulation.

At the first step, *local reconstruction* algorithms in sub-detectors take “RAW” data from the DAQ system as input and produce “RecHits” or reconstructed hits. The RecHits are hit positions in case of tracker and muon chambers, energy deposits in case of ECAL and HCAL. Track segments or energy clusters are built by grouping together nearby RecHits. Using these low level reconstructed objects, the PF algorithm proceeds to identify different particles as PF candidates in the reconstructed or “RECO” events. Further selection based on the RECO information is used to make concise collections of particles for Analysis Object Data or “AOD” format which is used in the analyses. During Run 2, the AOD is further shortened to form MiniAOD, NanoAOD datasets, having lesser but more compact information of the event. Typical size of an AOD event is ~500 KB and that for MiniAOD

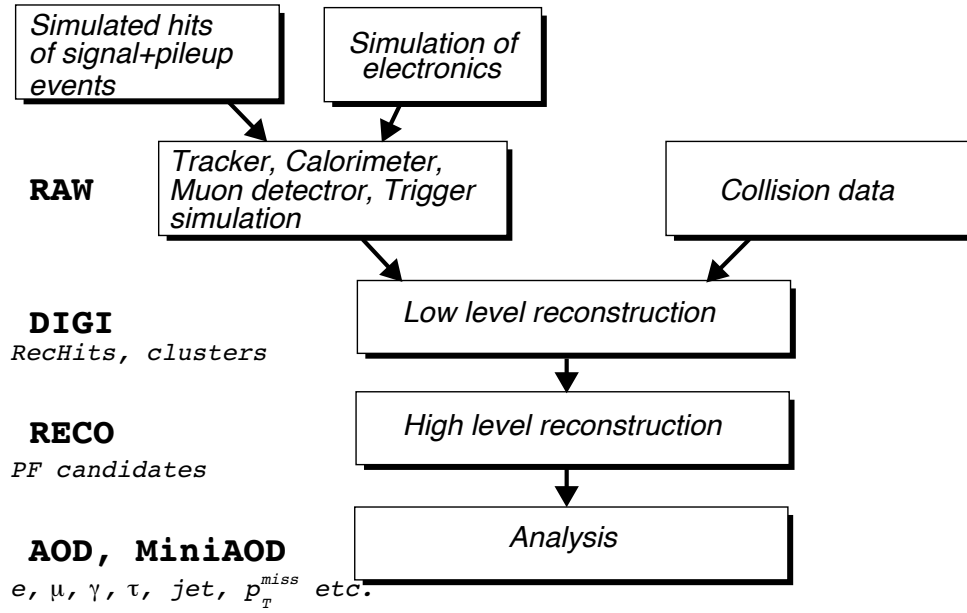


Figure 3.1: Steps of event reconstruction in the CMS analyses.

and NanoAOD are ~ 100 KB and ~ 1 KB respectively. Collision datasets immediately reconstructed with default detector calibrations for analysis are referred to as PromptReco (prompt) data. These datasets are reprocessed with updated CMSSW versions and detector calibrations at later time, called ReReco (re-reco).

3.3 The PF algorithm

The PF algorithm [101] combines information from different sub-detectors to first provide particle identification and then put forward a global picture, eventually resulting in a better measurement of energy, momentum and position of the concerned particles. This is possible because of the fine segmentation and excellent performance of the silicon tracker and electromagnetic calorimeter. These sub-detectors are crucial for removing overlap between energy deposits from charged and neutral hadrons with photons, as discussed later on.

The *global reconstruction* of the event starts with locally reconstructed charged tracks in the tracker, matched to muon chamber hits using an iterative process. Once good quality

tracks are formed they are removed from the event for further reconstruction. After muons, tracker hits are matched to calorimeter clusters to identify electrons and charged hadrons. When all tracker hits are associated to charged particles, in subsequent steps passage of photons and neutral hadrons are reconstructed using remainder clusters in ECAL and HCAL. These objects are called PF particles, which are used as input to jet or missing transverse momentum (p_T^{miss}) clustering algorithms. The PF algorithm is also able to identify particles from pileup interactions in this approach and correct for the effects in identification. A simplified depiction of the typical signatures of various particles in different sub-systems of the CMS detector is shown in Fig. 3.2.

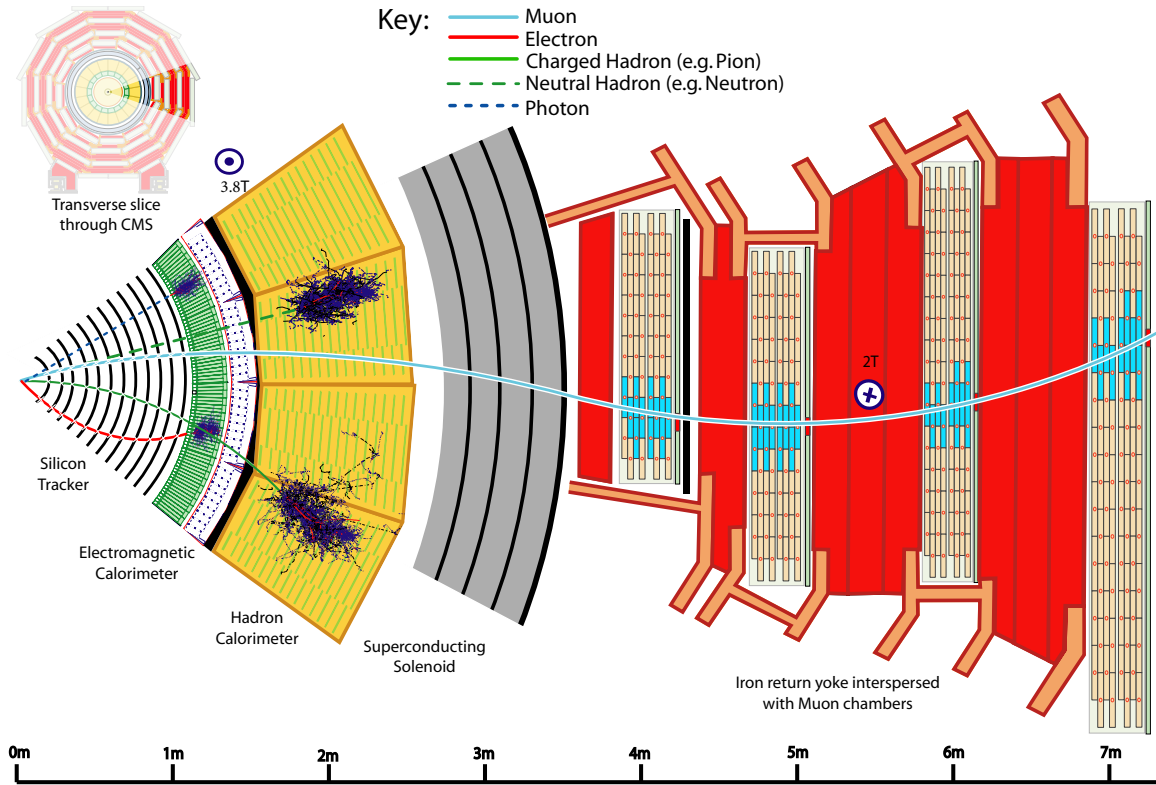


Figure 3.2: Slice view of CMS detector showing examples for muon, electron, photon, charged and neutral hadron [87]. In this depiction the muon and charged hadron are positively charged, the electron is negatively charged.

3.3.1 Track and vertex

In Run 2, there was on average 25–40 pileup interactions in a single bunch crossing leading to ~ 1000 charged tracks. To determine the event kinematics of the hard scattering, it is important to correctly associate the tracks to respective interaction vertices. However, reconstructing the tracks in such a busy environment is a challenge in itself. Falsely reconstructed tracks from combination of unrelated hits or spurious noise must be as low in number as possible. In addition, the reconstruction must be fast enough (~ 100 ms) to be integrated with the HLT algorithm.

Hit reconstruction

In the first step *zero-suppressed* signals (i.e., after noise subtraction) above a specified threshold in pixel and strip sensors are clustered into hits. The clusters are formed in both transverse and longitudinal directions in an identical way by matching adjacent pixels, side-by-side and corner-by-corner, estimating position and uncertainty in local coordinates. Two different algorithms are used in the pixel system for the hit reconstruction, one used for *seeding* and the other for final track fitting. *Seeds* are hit points with relatively large p_T at the innermost pixel layer from which the track reconstruction starts.

Track reconstruction

A cone is projected from the seed position to the next layer in radial direction to find matching hits. If no hit is found in next layer, the seed is discarded as noise or having too low energy for reconstruction. The procedure is then repeated until the full track through all the layers is found. This is the principle of the Kalman Filter algorithm [102] used in the Combinatorial Track Fitter (CTF) of CMS. It takes into account the effect of magnetic field as well and the bending radius. Direction from the resulting trajectory is used to associate a transverse momentum and charge to the track. These tracks are subsequently combined with matching calorimeter clusters and muon chamber hits to form higher level objects, like electrons, muons etc.

Vertex reconstruction

An interaction vertex is reconstructed by selecting tracks, followed by clustering and fitting them to find the actual vertex position. The tracks are selected using strict requirement on the reconstruction quality, such as the number of hits in pixel and strip detector layers, quality of the fit from CTF algorithm and significance of the transverse impact parameter relative to the beam spot centre. The track clustering is done using the deterministic annealing (DA) [103] algorithm and finally adaptive vertex fitter [104] is used to obtain the best estimate of the vertex position. The reconstructed vertices are referred to as primary vertices (PV) which are ordered by their respective sum of p_T^2 of all associated tracks. The interaction vertex is defined to have the highest sum of p_T^2 related to the hard scattering itself, whereas secondary vertices may be associated with decays of particles with longer lifetime. For example, daughter particle tracks of b-hadrons are displaced from the interaction vertex by a few mm, which can lead to reconstruction of a secondary vertex. The position resolution for reconstructed primary vertices in three directions were measured to be within 10-12 μm in Run 1 [105]. In Run 2 the resolutions have worsened due to higher instantaneous luminosity, e.g. in 2016 they were measured to be within 14-30 μm . With the upgraded pixel detector the resolutions were better in 2017 with values between 12-25 μm .

3.3.2 Muon

Muons are the very first particles to be reconstructed by the PF algorithm. The reconstruction is based on independent tracks in the inner tracker and the muon system, which are subsequently combined. Muons reconstructed solely from muon station hits are termed as *standalone* muons. The standalone muon is matched to tracks from the inner tracking system to form the *global* muon. The global reconstruction must pass requirements based on the fit quality, number of hits in sub-detectors, agreement in measured momenta etc. Alternative to this outside-to-inside approach, tracks from inner tracker are extrapolated to muon station hits, in case of poorly reconstructed segments in the muon detectors. These are termed as *tracker* muons, mainly useful to recover low

p_T muons which fail to penetrate the complete muon spectrometer. As mentioned in Chapter 2, muons are the only particle that traverse the full detector volume with negligible energy loss and thus have the best reconstruction. However, charged hadrons can be misidentified as muons because of *punch-through*, which are rejected using an isolation criteria considering additional charged tracks and calorimeter deposits nearby. Non-isolated global muons with poorly reconstructed inner track segment can be recovered if the standalone reconstruction is of high quality, i.e. large number of hits in the muon stations. In Run 2 muons are reconstructed and identified with greater than 96% efficiency, having negligible false reconstruction rate [106]. Muon station hits are used at L1 and the full reconstruction at HLT to select potentially interesting physics events.

3.3.3 Electron

Electrons are identified using tracks in the inner tracker and energy deposits in ECAL. Their track reconstruction quality is worse than muons since electrons interact with the tracker material and lose energy due to bremsstrahlung at a much faster rate. The ECAL energy deposits are measured from arrays of 3×3 or 5×5 crystals, called clusters, which give the best performance for identification of unconverted photons or electrons. Due to the magnetic field, the energy deposit may spread along the ϕ direction for electrons, thus several clusters are considered together, forming superclusters. The PF algorithm starts with an energy deposit (greater than 4 GeV) in a ECAL supercluster and uses the position of the cluster to infer the hit positions within inner tracker, assuming the particle to be either an electron or a positron [107]. Emission of bremsstrahlung photons and bending of the electron are taken into account while grouping the ECAL supercluster in the η - ϕ space. However, the ECAL based approach does not work well for electrons with low p_T or those originating from a jet. In such cases track seeding is used along with iterative tracking method. A Gaussian-sum filter (GSF) [108] is finally used to fit pre-selected tracks, which is more suited to electron reconstruction as it allows for substantial energy loss along the trajectory. For good reconstruction, momentum-to-energy ratio of the electron is close to unity and there should not be any matching cluster in HCAL for the corresponding track.

3.3.4 Photon and hadrons

After the PF muon and PF electron objects are reconstructed and removed from the collection of hits and clusters, the PF algorithm associates the remaining tracks from the inner tracking system to charged hadrons. Photon and neutral hadrons are identified respectively from ECAL and HCAL clusters without associated track.

The photon mainly loses energy due to pair-production in the ECAL material that turn into electromagnetic shower. The formation of the ECAL supercluster begins from a seed crystal: a crystal which has higher energy deposit than adjacent crystals, above a certain energy threshold. The clusters have a fixed width of 5 crystals around the seed in the η direction; in contrast, the supercluster size varies in the ϕ direction. The extended ϕ window in the barrel region can reach up to ± 17 crystals around the seed, controlled by a pre-defined energy threshold. In the ECAL endcap the supercluster spans a fixed grid of 5×5 crystals. This supercluster formation is identical for electrons and photons, as the PF algorithm does not distinguish between the particles while clustering. However, photons do not have associated tracks and their identification is mainly based on the shower shapes and lateral spread of the shower.

After identifying and removing PF photons from the list of particle signatures, only remaining particles are hadrons from parton fragmentation and hadronization. The PF algorithm does not distinguish which SM particles result in the hadron signatures, thus particles like $\pi^\pm$, $K^\pm$, protons constitute PF charged hadron collection; neutrons, K_L^0 etc. constitute the PF neutral hadron collection. For ECAL clusters non-associated with a track, the primary contributor is a photon inside the tracker coverage of $|\eta| < 2.5$. This is justified with the observation that within a jet 25% energy is carried by photons, whereas neutral hadrons carry 3%. Outside tracker, ECAL clusters linked with a HCAL cluster are assumed to be from the same hadron shower, charged or neutral.

3.3.5 Jet

Quarks and gluons cannot be directly observed as particles due to color confinement, a result of carrying the color charge of strong interactions. After emerging from the

hard interaction, they quickly form multiple semi-stable mesons and baryons. Indirect observation of quarks and gluons is thus possible in the form of a collection of such color neutral particles, referred to as hadronic “jets”.

Jets are reconstructed using clustering algorithms which ideally group all particles originating from the same parton into one jet. Two important conditions need to be satisfied while reconstructing, so that the jet clustering outcome does not arbitrarily change:

- **collinear safety:** one of the constituent particles of the jet is replaced with two particles, each having half the momentum of the original particle.
- **infrared safety:** additional soft radiation from a constituent quark.

In CMS experiment, jets are reconstructed using the anti- k_T [109] algorithm that satisfies both the criteria. In jet algorithms distance parameters are defined for an input PF candidate (i) with respect to a second PF candidate (j) and with the beam (B):

$$\begin{aligned} d_{ij} &= \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \frac{\Delta_{ij}^2}{R^2} \\ d_{iB} &= p_{Ti}^{2n} \end{aligned} \quad (3.2)$$

where p_{Ti} and p_{Tj} are transverse momenta of the i^{th} and j^{th} particles, n is an algorithm specific parameter, R is the size parameter and $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$. In case of anti- k_T , $n = -1$ and in general for CMS analyses $R = 0.4$; the jet collections so obtained are referred to as AK4. The clustering starts with the highest p_T candidate and iteratively looks for the nearest particle comparing the distances d_{ij} and d_{iB} . If $d_{ij} < d_{iB}$, two particles are combined into one and the individual particles are discarded from the list. If $d_{ij} > d_{iB}$, the clustering ends for this jet and it is removed from the particle collection. Subsequently, a strategy called charged hadron subtraction (CHS) is used to remove charged particle tracks originating from pileup vertices before reconstructing jets, these collections are called AK4PFCHS, utilized by most of the CMS analyses, including the ones discussed in this thesis.

Jet energy correction

In order to match the kinematics of the reconstructed jet as closely as possible to the initial parton, p_T and η dependent corrections are needed to be applied [110] to take into

account the detector effect which depends on the path of the particles:

- **L1 offset:** subtracts the average pileup energy from the jet by calculating the offset energy density (ρ) [111]. The correction to the jet energy is obtained as function of ρ , jet area A , jet η and p_T :

$$\Delta p_T^j = [\rho_0(\eta) + \rho\beta(\eta)(1 + \gamma(\eta)\log p_T)]A$$

where parameters $\rho_0(\eta)$, $\beta(\eta)$, $\gamma(\eta)$ are determined from simulation.

- **L2 relative and L3 absolute:** The jet response, defined as:

$$R(< p_{T, reco} >, \eta) = \frac{< p_{T, reco} >}{< p_{T, gen} >} [p_{T, gen}, \eta]$$

compares generated p_T of the parton to the reconstructed jet p_T . Ideally this ratio should be nearly one and stable with respect to p_T and η of the jet. However, due to non-linearity of the calorimeter, the jet response needs to be corrected in MC, using scale factors derived from the truth information.

- **L2L3 residual:** small residual differences between data and MC is accounted for by applying a correction factor to data alone.

b-tagging

Using certain characteristics of the reconstructed jet, the initial flavor of the quark originating the jet can be determined to a reasonable extent. This is especially applicable for heavy flavor quarks like b and c. The b quarks hadronize to form b-hadrons having a relatively longer lifetime of about 1.6 ps, so the decay (secondary) vertex is typically displaced from the interaction vertex by a few mm. Several *b-tagging* algorithms have been developed to take into account the secondary vertex feature of the b decay as well as track-based lifetime information to identify or *tag* the jets originating from a b quark. The most popular one being the combined secondary vertex (CSV) tagging which provides each jet with a likelihood score that determines the probability of correctly identifying a b-jet. The widely used versions of CSV algorithm in Run 2 analyses are CSV_{v2} using

Multi-Variate (MVA) based techniques and *DeepCSV* using deep neural network algorithms. The working points for the tagging is chosen on the basis of the mis-tag rate: the probability of identifying a light flavor parton (u, d, s, g) as a b quark. The tight working point has a mis-tag rate of 0.1%, whereas medium and loose working points have mis-tag rates of 1% and 10% respectively. A given mis-tag rate has a fixed b-tag efficiency value, qualitatively. The b-tagging efficiencies for these algorithms are shown in Fig. 3.3. As it can be seen, these algorithms also have certain efficiency to tag charm jets, however, recently dedicated charm taggers have also been developed for analysis purpose, such as the *cMVA*.

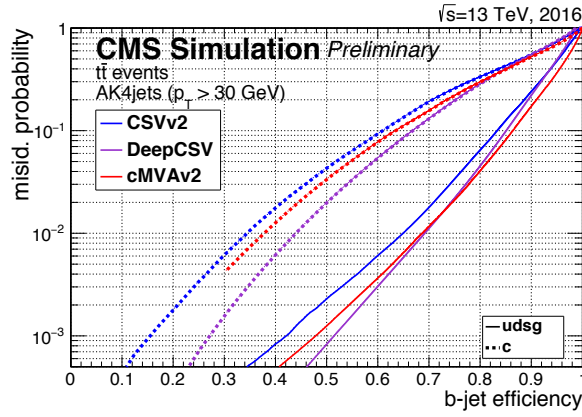


Figure 3.3: Performance of the b-tagging algorithms demonstrating the mis-tag probabilities for non-b jets as a function of the efficiency to correctly identify b jets [112]. In case of jets with $p_T > 30$ GeV, the b-tagging efficiencies range between 40–50% for the tight w.p., 65–70% for the medium and $\sim 85\%$ for the loose.

3.3.6 p_T^{miss}

Weakly interacting particles such as neutrinos do not leave any signature in the detector material. Their presence can be indirectly inferred from the momentum imbalance of the reconstructed event. The missing transverse momentum due to invisible particles are defined in terms of the energy/momentum flow in the plane transverse to the beam

direction:

$$p_T^{miss} = |\vec{p}_T^{miss}| = \left| - \sum_i \vec{p}_T^i \right| \quad (3.3)$$

where $\vec{p}_T^i$ is the transverse momenta of the i^{th} visible particle present in the event. The amount of p_T^{miss} can be clear signature of many BSM scenarios, thus the clustering and momentum balance is very important. In CMS, the “raw” p_T^{miss} value is determined by the PF algorithm by vectorial sum of all the PF particle momenta. If the jets used in the reconstruction are corrected with JEC, the same can also be propagated to the p_T^{miss} , which is referred to as the “Type 1” correction. Alternative approach to cluster the p_T^{miss} exist, such as the PUPPI algorithm [113]. A detailed study of the measurement capacity of p_T^{miss} in CMS experiment has been carried out as discussed in next chapter.

Chapter 4

Studies with p_T^{miss} and jets

An important aspect of collision data analysis is good understanding of the detector. The CMS detector has many components and the response of each evolves with time due to radiation damage, ageing, changing luminosities as well as actual hardware upgrades of specific sub-systems in terms of electronics etc. Continuous efforts are made to maintain and upgrade the performance in terms of reconstruction and particle identification. This chapter describes two such studies with Run 2 data collected at different periods of time.

4.1 Measurement of early trigger rates in HCAL

Jet pre-firing in CMS L1 trigger is a crucial source of possible mis-identification of interesting event at the LHC p-p collision. Pre-firing or early trigger occurs when the trigger fires in the bunch crossing before the one in which the event of interest has occurred. Furthermore, CMS system allows to accept only one L1 trigger decision issued in any three consecutive bunch crossings. With high instantaneous luminosity of the LHC machine during operations in Run 2, reaching up to twice the nominal value of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ the probability of early trigger is expected to be few % due to pileup jets. This can lead to important physics events to be potentially rejected and hence causing reduced L1 trigger efficiency for the signal events. A study is performed to measure the rates of trigger pre-firing in the HCAL for data collected in 2016 as a function of L1 jet transverse energy (E_T).

The pre-firing rate (R) is defined as the ratio of the events which have the leading jet in the early bunch crossing ($BX = -1$) passing a transverse energy threshold $E_{th}^{BX=-1}$ and to that of the leading jet in the nominal bunch crossing ($BX = 0$) passing a different threshold $E_{th}^{BX=0}$. If the leading jet E_T in $BX = 0$ and $BX = -1$ are denoted by $E_T^{BX=0}$ and $E_T^{BX=-1}$ respectively, then:

$$R = \frac{N_{E_T^{BX=0} > E_{th}^{BX=0}, E_T^{BX=-1} > E_{th}^{BX=-1}}}{N_{E_T^{BX=0} > E_{th}^{BX=0}}}. \quad (4.1)$$

The pre-firing rate as defined above is calculated for the barrel and endcap regions separately.

Even though pileup is the main reason for early triggers, other possibility can be beam halo particles, resulting from the secondary interactions of the beam remnant with residual gas molecules inside the detector. These particles are mainly muons, travelling parallel to the beam direction since they do not emerge from an interaction vertex. The halo muons can be identified using reconstructed tracks parallel to the beam line, traversing both the muon endcaps. An event filter is available within the CMSSW framework to reject events with such kind of reconstructed muons. Effects of the beam halo filter is studied to see if it can improve the pre-triggering rate.

4.1.1 Event selection

Events of interest are chosen by referring to the LHC bunch fill structure. As mentioned in Chapter 2, protons at the LHC collide in bunches; this is because the acceleration mechanism traps the particles in *buckets* of RF voltage frequency, which is 400 MHz. A proton exactly synchronized with the RF oscillation is accelerated by the cavity, whereas protons arriving with slightly different timing and energy are either accelerated or decelerated to form the bunch. Not all buckets are filled with particles, the LHC design bunch spacing being 25 ns or 10 buckets. Further, the bunch fill structure is determined by the LHC operational team by carefully considering the cross over and collision points. A significant empty space is also kept for the purpose of beam dump when the beam intensity goes below a certain threshold. The CMS Run Summary page provides detailed information of the bunch pattern, e.g. Fig. 4.1 from which particular bunch crossings with empty

bunches at front are chosen for this analysis.

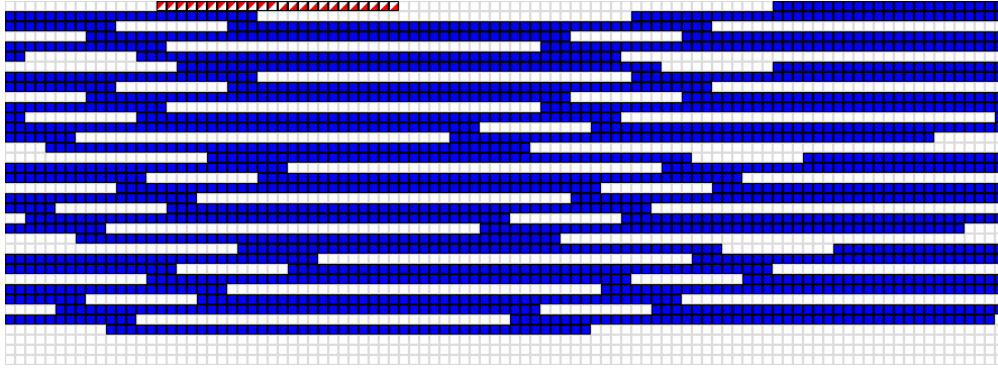


Figure 4.1: Bunch pattern of fill number 5078. By design the LHC can have at most 3564 colliding bunches, this particular fill having 2064 denoted by the blue squares. The red squares are also part of the beam but not colliding. The empty squares are not configured for the beam.

The data collected are tagged with a run number and lumi section. The run numbers depend on the DAQ operation which change with status of the beam, detector sub-system conditions etc. A lumi section is defined by a duration of ~ 23.3 seconds of beam revolution inside the LHC. The run and lumi information is collected in a JavaScript Object Notation (JSON) format used in analyses.

4.1.2 Data samples

The analysis of data collected in 2016 is performed using CMSSW version 8_0_7. Datasets triggered with H_T or sum of all jet E_T with RAW reconstruction is used for determining the pre-triggering rate at different *eras* or data taking periods of 2016, shown in Table 4.1. The CSC beam halo information is extracted from reconstructed events in corresponding MINIAOD datasets.

The event information is analyzed using official L1 tools to reconstruct jets including the corresponding bunch crossing information, not present in the usual data format. The following JSON file containing the information about lumi sections was used for selecting certified good events where the CMS magnet was on and all sub-systems recorded data:

Cert_271036-282092_13TeV_PromptReco_Collisions16_JSON.txt.

Table 4.1: Data samples used to determine pre-firing rates and the dates of collection. The instantaneous luminosity is increased in run periods as data collection progresses. For the four fills they are 0.3, 1.0, 1.2 and 1.3 times $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ in the order shown.

RAW data sample	MINIAOD data sample	Fill number	Date
/JetHT/Run2016B-v2/RAW	/JetHT/Run2016B-23Sep2016-v3/MINIAOD	4947	19 th May, 2016
/JetHT/Run2016D-v2/RAW	/JetHT/Run2016D-23Sep2016-v1/MINIAOD	5078	7 th July, 2016
/JetHT/Run2016G-v1/RAW	/JetHT/Run2016G-23Sep2016-v1/MINIAOD	5211	17 th August, 2016
/JetHT/Run2016H-v1/RAW	/JetHT/Run2016H-PromptReco-v2/MINIAOD	5355	2 nd October, 2016

Figure 4.2 shows the jet E_T distributions of leading jet from nominal and early bunch crossing in case of fill 5355.

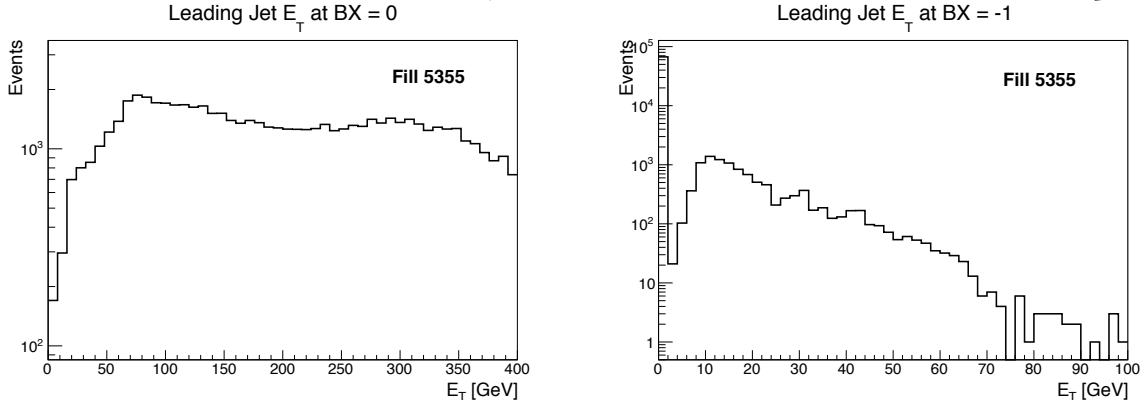


Figure 4.2: Leading jet E_T distributions in bunch fill 5355, $BX = 0$ at left, $BX = -1$ at right. It can be seen that many events have jets with $E_T > 50 \text{ GeV}$ at $BX = -1$.

4.1.3 Results

The following figures show the pre-firing rates as a function of $E_{th}^{BX=0}$ for different $E_{th}^{BX=-1}$ for the two η regions in HCAL: barrel and endcap, for the bunch fill 4947 in Fig. 4.3, fill 5078 in Fig. 4.4, fill 5211 in Fig. 4.5 and fill 5355 in Fig. 4.6. It is observed that the barrel pre-firing rates are much smaller and overall the rates are less than 1% in 2016.

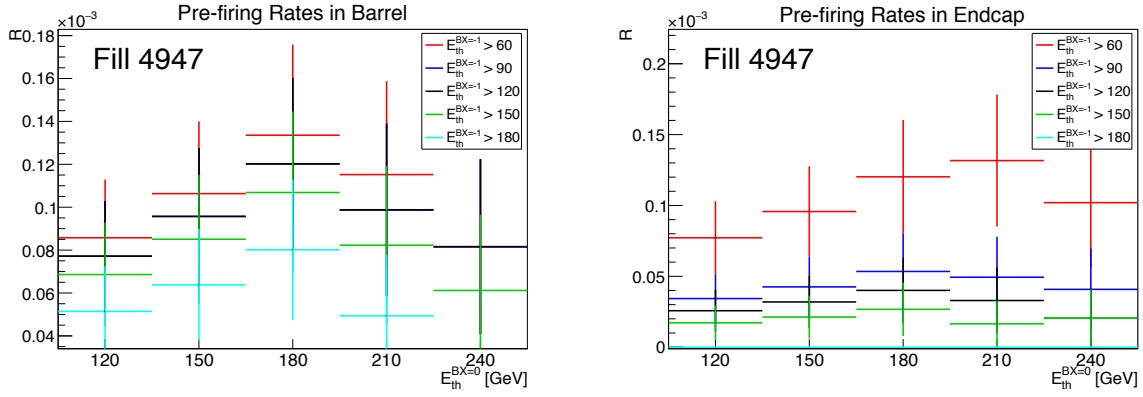


Figure 4.3: Pre-firing rates in bunch fill 4947, left plot for barrel, right for endcap.

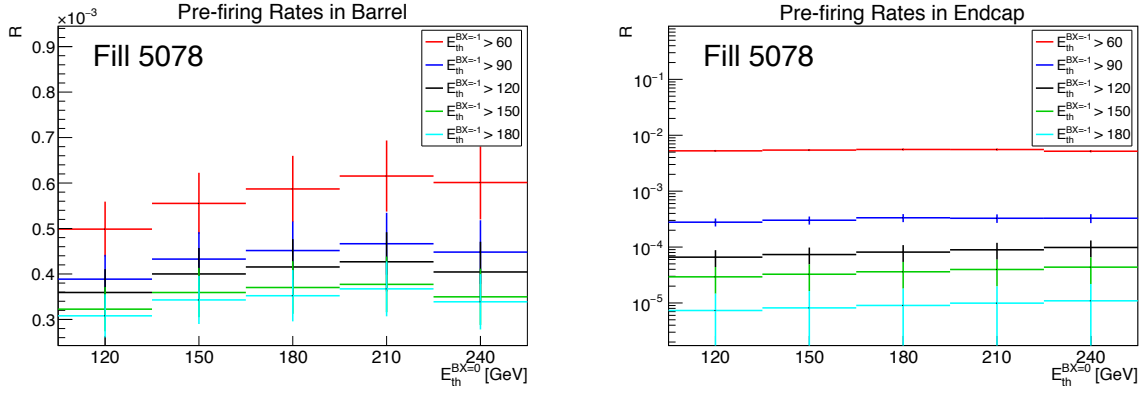


Figure 4.4: Pre-firing rates in bunch fill 5078, left plot for barrel, right for endcap.

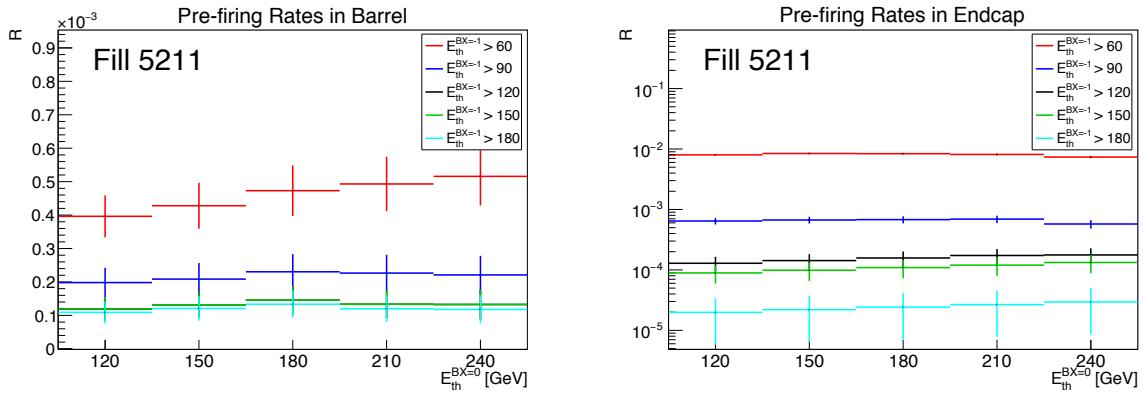


Figure 4.5: Pre-firing rates in bunch fill 5211, left plot for barrel, right for endcap.

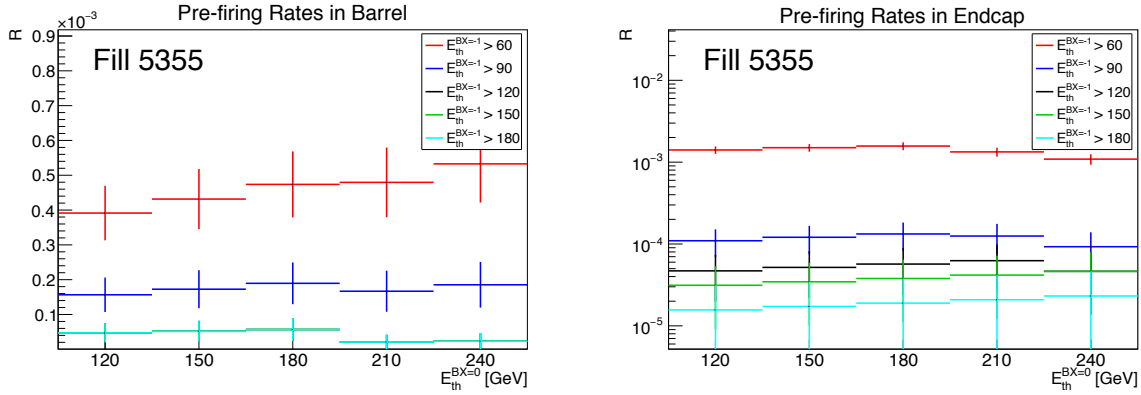


Figure 4.6: Pre-firing rates in bunch fill 5355, left plot for barrel, right for endcap.

4.1.4 Effect of beam halo filter

The halo filter used in this study utilizes information from the CSC sub-detector which covers a region $1.2 < |\eta| < 2.4$. Occurrence of pre-firing jets are studied for events passing the CSC filter in Fig. 4.7 and for all the events, irrespective of any filter in Fig. 4.8, in this region. No change is observed in the number of early jets due to the beam halo filter in

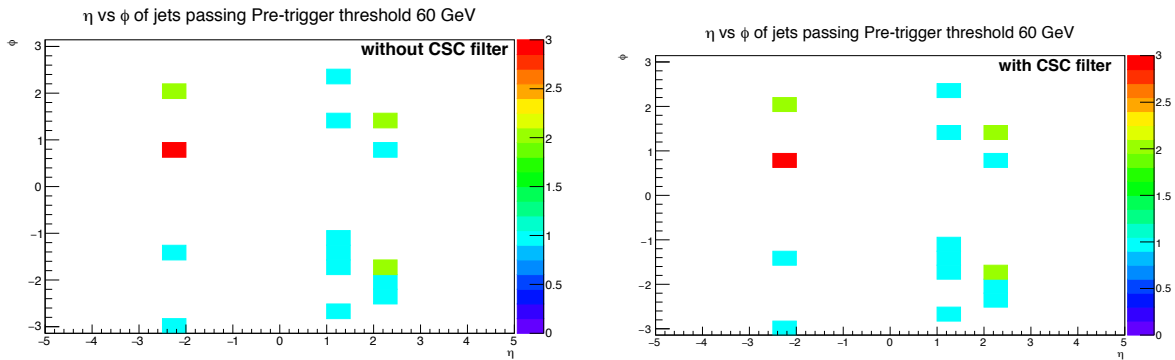


Figure 4.7: Pre-firing jets without (left) and with (right) beam halo filter.

the CSC covered region. The number of in-time jets reduces slightly because of the filter, but the change in pre-firing rate is negligible. This may be due to the fact that halo events have mostly small E_T jets, so it rarely passes any of the thresholds.

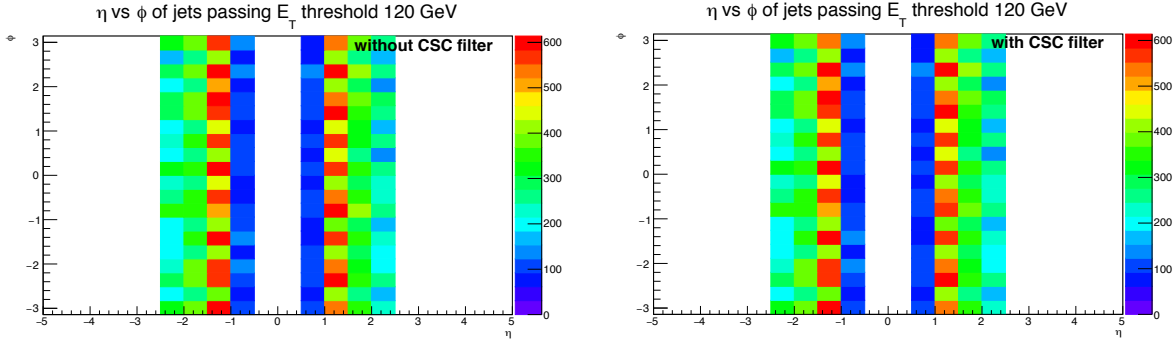


Figure 4.8: Nominal BX jets without (left) and with (right) beam halo filter.

4.2 Performance of algorithms for p_T^{miss} measurements

The missing transverse momentum is an important measurable quantity for many processes in SM physics and beyond, viz. leptonic decay of W or production of SUSY particles in certain scenarios. It is defined in terms of measured momentum flow in the transverse plane to the beam direction as given by Eq. (3.3). In the experiment there can be other sources of momentum mismeasurement due to detector resolution and sensitivity. The p_T^{miss} is also sensitive to the presence of pileup interactions, both in-time and out-of-time.

To reduce the pileup dependence, algorithms such as PUPPI (Pileup Per Particle Identification) [113] uses event pileup properties and tracking information to construct a shape variable α_i for each particle i which is used to compute the likeness of a PF candidate to be from pileup. The particle is then assigned a corresponding weight w_i which is used to reweight the four-vector of the particle. After this the clustering of p_T^{miss} is done in the usual way, referred to PUPPI p_T^{miss} in contrast to PF p_T^{miss} :

$$\text{PUPPI } p_T^{miss} = \left| - \sum_i w_i \vec{p}_T^i \right|. \quad (4.2)$$

The performance of p_T^{miss} reconstruction algorithm can be measured in events where there is no genuine source of missing energy and the final state particles are precisely calibrated. Hence, excellent measurement capability of photon energy in CMS detector can be utilized to measure the instrumental p_T^{miss} in the events of inclusive single photon and

jets. If the photon transverse momentum is $\vec{q}_T$, and transverse momentum of the hadronic recoil, defined as the vectorial sum of the transverse momenta of all particles except the photon, is $\vec{u}_T$, then from momentum conservation, for such events with measurable p_T^{miss} , as illustrated in Fig. 4.9:

$$\vec{q}_T + \vec{u}_T + \vec{p}_T^{miss} = 0. \quad (4.3)$$

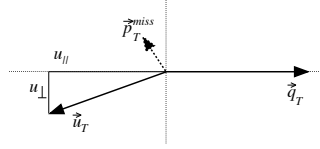


Figure 4.9: Illustration of γ +jets event kinematics.

The photon provides a precise momentum scale of the event and a unique event axis along the direction of $\vec{q}_T$. The projections of the hadronic recoil parallel ($u_{\parallel}$) or perpendicular ($u_{\perp}$) to this axis are measures of the momentum balance of the event.

Decomposing $u_{\parallel}$ and $u_{\perp}$ in terms of the x and y components of $\vec{p}_T^{miss}$ and $\vec{q}_T$:

$$\begin{aligned} u_{\parallel} &= (u_x q_x + u_y q_y) / q_T \\ u_{\perp} &= (u_x q_y - u_y q_x) / q_T \end{aligned} \quad (4.4)$$

where $u_x = -p_x^{miss} - q_x$ and $u_y = -p_y^{miss} - q_y$ are the components of the hadronic recoil in the transverse plane. In the ideal case where there is no p_T^{miss} in the final state, $|u_{\parallel}| = |q_T|$ and $u_{\perp} = 0$.

4.2.1 Corrections and uncertainties

The jet energy scale (JES) corrections derived using γ +jets, Z+jets and QCD multi-jet events are applied to the jets reconstructed by the PF algorithm. The corrections are propagated to the p_T^{miss} and it is referred to as Type 1 corrected p_T^{miss} :

$$\text{Type 1 } \vec{p}_T^{miss} = \vec{p}_T^{miss} - \sum_{jet} (\vec{p}_{T,jet}^{corr} - \vec{p}_{T,jet}). \quad (4.5)$$

Associated uncertainties due to the JES measurement is propagated to the above relation. In addition, the p_T^{miss} also includes uncertainties due to particles that are not clustered into jets.

An asymmetry in the ϕ distribution of p_T^{miss} is observed due to shifts in its x and y components in the transverse plane. This effect is mainly due to pileup and imperfect detector alignment, specially in the tracker sub-system. Corrections can be derived for each component as a function of the number of interaction vertices in the event (n_{PV}), for example. For p_T^{miss} validation studies in 2018 data, the following linear parametrization has been used:

$$\begin{aligned} p_X^{miss} &= a_1 * n_{PV} + b_1 \\ p_Y^{miss} &= a_2 * n_{PV} + b_2 \end{aligned} \tag{4.6}$$

where constants a_1, b_1, a_2, b_2 depend on the run period. With increasing instantaneous luminosity n_{PV} increases and hence the uncertainty for identification of the primary interaction vertices increases, effectively leading to non-zero p_T^{miss} .

4.2.2 Data and simulated samples

The full analyzed datasets for 2017 and 2018 correspond to integrated luminosities 41.9 and 57.9 fb⁻¹ respectively. The data were collected when the CMS magnet was on, during p-p collisions at $\sqrt{s} = 13$ TeV with bunch crossings every 25ns. The 31Mar2018 re-reco versions of SinglePhoton datasets triggered with the presence of photons in 2017 are analyzed (run periods B to F). For 2018 run periods A to C, 17Sep2018 re-reco and for run period D prompt reconstructed version of datasets triggered with electron or photon have been analyzed. They are listed in Table [4.2](#).

Following are the JSON files used to select the good data events:

- 2017: Cert_294927-306462_13TeV_E0Y2017ReReco_Collisions17_JSON.txt
- 2018: Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt

Table 4.2: Data samples used in the determination of p_T^{miss} reconstruction performance. Also shown are the number of collision events and integrated luminosities for each run period.

Sample	# events	L_{tot} (fb ⁻¹)
/SinglePhoton/Run2017B-31Mar2018-v1/MINIAOD	15950935	4.8
/SinglePhoton/Run2017C-31Mar2018-v1/MINIAOD	42182948	9.8
/SinglePhoton/Run2017D-31Mar2018-v1/MINIAOD	9753462	4.3
/SinglePhoton/Run2017E-31Mar2018-v1/MINIAOD	19011446	9.4
/SinglePhoton/Run2017F-31Mar2018-v1/MINIAOD	29783015	13.6
/EGamma/Run2018A-17Sep2018-v2/MINIAOD	327843843	13.2
/EGamma/Run2018B-17Sep2018-v1/MINIAOD	153822427	6.8
/EGamma/Run2018C-17Sep2018-v1/MINIAOD	147827904	6.6
/EGamma/Run2018D-PromptReco-v2/MINIAOD	738976078	31.3

CMSSW versions used for reconstruction and analysis are 9_4_13 in case of 2017 and 10_2_15 in case of 2018. MC events are generated using MADGRAPH in LO for the signal γ +jets samples, each event is then weighted to NLO cross section as a function of the p_T and η of the leading photon. LO samples are used for QCD multi-jet and W +jets processes, whereas NLO samples are analyzed for $t\bar{t}\gamma$, $t\gamma$, $W\gamma$ and $Z\gamma$ processes. MC samples generated for 2017 and 2018 analyses utilize PYTHIA8 tune CP5 [114] for underlying event generation. Due to $W\gamma$ and $Z\gamma$ samples not being available with this tune, older samples generated with CUETP8M1 tune [115] are used. In all cases GEANT4 is used for detector simulation. Cross sections for all processes are normalized using NLO or NNLO values as available. Table 4.3 shows the detailed list of MC samples used.

Table 4.4 shows the set of prescaled photon triggers applied to select both data and MC events in both years. The number next to the object name specifies the p_T threshold. The triggering photons are also required to pass medium isolation cuts and requirements on the R9 and H/E variables, as discussed in next section.

Table 4.3: List of simulated samples considered for the γ +jets final state, along with the cross sections. Inclusive cross sections are listed for all samples, except for $W\gamma$ and $Z\gamma$ where leptonic decays of the vector bosons are considered: $W \rightarrow \ell\nu$, $Z \rightarrow \ell\ell, \nu\nu$.

Process	Sample	$\sigma \times BR$ [pb]
γ +jets in H_T bins	GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8	19000
	GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	8300
	GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	2200
	GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	260
	GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	84
QCD multijet in H_T bins	QCD_HT100to200_TuneCP5_13TeV-madgraphMLM-pythia8	27849880
	QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8	1716997
	QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8	351302
	QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8	31630
	QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8	6802
	QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8	1206
	QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8	98.71
	QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8	20.2
W+jets in H_T bins	WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8	15965.99
	WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	1632.15
	WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	478.179
	WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	67.298
	WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	14.953
	WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	6.138
	WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.253
	WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.00958
top+ γ	TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	4.322
	TGJets_TuneCP5_13TeV-amcatnlo-madspin-pythia8	2.967
$W\gamma, Z\gamma$	WGToLNuG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	489.0
	ZGTo2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	117.864
	ZGTo2NuG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	28.05

Table 4.4: High-level triggers used to select events with photons. The trigger photon is required to have hadronic over electromagnetic energy fraction (H/E) less than 0.10 and the 3×3 array around the supercluster seed has to contain at least 90% of the total supercluster energy. In addition, the photon is required to pass the medium working point for the isolation. The v* refers to different versions of the HLT path, which are updated with detector condition and calibration.

HLT_Photon50_R9Id90_HE10_IsoM_v*
HLT_Photon75_R9Id90_HE10_IsoM_v*
HLT_Photon90_R9Id90_HE10_IsoM_v*
HLT_Photon120_R9Id90_HE10_IsoM_v*
HLT_Photon165_R9Id90_HE10_IsoM_v*

4.2.3 Objects

Photons

The best reconstruction of unconverted photons (i.e. not decaying to e^+e^- pair) in the CMS detector is obtained starting from superclusters, as detailed in Chapter 3. The selected photon object must pass the following requirements:

- Photon transverse momentum $q_T > 50$ GeV and $|\eta| < 1.44$ in order to select highly energetic photons in the barrel region; endcap photons are not considered in order to suppress noise due to spikes (spurious electronic pulse) and beam-halo.
- The photon must be isolated, i.e. away from other charged particles. The isolation (I) is defined to be the sum of particle momenta within a cone in η - ϕ space around the photon. The effect of pileup on I is corrected using effective area (EA) method:

$$I(\text{corrected}) = \max(I - \rho * EA, 0).$$

Here ρ is the median background energy density per unit area calculated in event-by-event basis, depending on the jet activity [111]. The effective area is the ratio of slopes obtained by linearly fitting isolation and ρ as a function of number of vertices.

The photon is required to pass the tight identification criteria [116] having 70.24% efficiency of selecting photons and 90.97% background (jets faking as photons) rejection efficiency, as detailed below.

- **H/E:** The ratio of hadronic and electromagnetic energy deposits has to be less than 0.02148.
- $\sigma_{i\eta i\eta}$: The ECAL crystals are indexed using numbers $(i\eta, i\phi)$ which label the (η, ϕ) co-ordinates. Shower shape variable $\sigma_{i\eta i\eta}$ measures the shower width in $i\eta$ -space and is required to be less than 0.00996.
- **PF charged hadron isolation:** Sum of p_T of all charged hadrons within a hollow cone of $0.02 < \Delta R < 0.3$ around the supercluster is required to be less than 0.65.
- **PF neutral isolation:** Sum of p_T of all neutral hadrons within a cone of $\Delta R = 0.3$ around the supercluster, is required to be less than:

$$0.317 + 0.01512 \times q_T + 2.259 \times 10^{-5} \times q_T^2.$$

- **PF photon isolation:** Sum of p_T of all photons within a hollow cone of $0.015 < \Delta R < 0.3$ around the supercluster, is required to be less than:

$$2.044 + 0.004017 \times q_T.$$

Table 4.5 shows the values of effective area considered for photon isolation.

Table 4.5: Effective areas for photon isolation [116].

$ \eta $ range	charged hadrons	neutral hadrons	photons
< 1.0	0.0112	0.0668	0.1113
1.0–1.479	0.0108	0.1054	0.0953

In addition to the tight photon selection, the following requirements are to be satisfied:

- **Pixel track seed veto:** The pixel track seed veto rejects any photon like object for which a pixel track seed consisting of at least two hits in the pixel detector is found; effective to reject electrons.

- **Extra ID for spike cleaning:** The R_9 variable is defined as the ratio of energy deposited in a 3×3 crystal region centered around the most energetic crystal in the supercluster divided by the total energy of the supercluster. This variable can effectively suppress background due to ECAL spikes and converted photons which have wider energy spread in the transverse direction. To match the trigger conditions, the leading photon is required to have $0.9 < R_9 < 1.0$.

Jets and p_T^{miss}

The analysis uses PF jets reconstructed with anti- k_T algorithm using a distance parameter equal to 0.4. Jets are selected using the tight identification criteria [117] having greater than 99.5% efficiency to select jets and greater than 98% noise rejection, shown in Table 4.6. Latest available JECs are applied to both 2017 and 2018 data and MC samples. These JECs are also used for corresponding Type 1 correction of the p_T^{miss} .

Table 4.6: PF jet identification criteria used to select γ +jets events [117].

Variable	$ \eta < 2.4$
Neutral hadron energy fraction	< 0.90
Neutral EM energy fraction	< 0.90
Number of constituents	> 1
Charged hadron energy fraction	> 0
Charged multiplicity	> 0

In 2017, radiation damage in the ECAL readout caused excess of low p_T jets in the endcap cells around $|\eta| \sim 3$. This results in instrumental p_T^{miss} in the data causing large disagreement with the simulation in the tail region of p_T^{miss} distribution. In order to mitigate this effect, the recipe for 2017, termed as EENoiseFix, requires jets and unclustered particles having $p_T < 50$ GeV, and within $2.650 < |\eta| < 3.139$ to be excluded from p_T^{miss} reconstruction. The effect of applying this recipe has been checked with different versions of JEC released for 2017 re-reco data. The ultimate solution to this issue is the “Ultra Legacy” re-reco of full Run 2 luminosity, scheduled to arrive in late 2019, where the ECAL PF

RecHit thresholds itself will have been adjusted to reject the noise. A trial version of the Ultra Legacy re-reco of data from 2017 era F has also been checked, without applying the above recipe and observed to have similar performance (see Section 4.2.7).

No single dedicated recipe for p_T^{miss} exists for 2018 data as of yet, many studies are on-going still. However there are known issues in the re-reco data such as the difference in energy deposits in HCAL endcap between data and simulation that may affect p_T^{miss} distributions. In both 2017 and 2018 data, severe mis-calibration of ECAL endcap energy exists which is to be corrected in the Ultra Legacy re-reco. Results discussed in next sections are obtained with EENoiseFix recipe applied to 2017 data and MC.

Electrons and muons

To reduce contamination from the top quark and electroweak backgrounds, events are rejected if they have electrons and muons, chosen with loose identification criteria as given in Tables 4.7 and 4.8. The selection is mainly based on track quality, distance from interaction point (d_{xy} and d_z), isolation, matching between sub-detector hits and requirements on shower shape for electrons ($\sigma_{i\eta i\eta}$, $|\Delta\eta_{in}^{seed}|$, $|\Delta\phi_{in}|$ variables).

Table 4.7: Loose identification criteria for electrons [118].

Variable	$ \eta < 1.479$	$1.479 < \eta < 2.5$
full 5×5 $\sigma_{i\eta i\eta}$	< 0.0126	< 0.0457
$ \Delta\eta_{in}^{seed} $	< 0.00463	< 0.00814
$ \Delta\phi_{in} $	< 0.148	< 0.19
H/E	$< 0.05 + 1.16/E + 0.0324 \times \rho/E$	$< 0.05 + 2.54/E + 0.183 \times \rho/E$
Relative isolation ($\Delta R = 0.3$)	$< 0.198 + 0.506/p_T$	$< 0.203 + 0.963/p_T$
1/E-1/p	< 0.209	< 0.132
Number of missing hits	≤ 2	≤ 3
Pass conversion veto	yes	yes
$ d_{xy} $ (cm)	< 0.05	< 0.1
$ d_z $ (cm)	< 0.1	< 0.2

Table 4.8: Loose identification criteria for muons [119].

Variable	$ \eta < 2.4$
$ d_{xy} $ (cm)	< 0.05
$ d_z $ (cm)	< 0.1
Relative isolation ($\Delta R = 0.3$)	< 0.4
tracker/global muon	yes

4.2.4 Event selection

Events are selected if:

- Exactly one tight photon is present in the event which is also the leading photon
- At least one jet is present in the event
- No loose electrons and muons are present with $p_T > 10$ GeV

Furthermore, following p_T^{miss} filters are used to reject events with anomalous high p_T^{miss} :

- **Primary vertex filter:** Select events with at least one well reconstructed primary vertex, based on quality criteria such as degrees of freedom, χ^2 of the track fits.
- **Beam halo filter:** Halo events are caused due to muons originating from secondary interaction of particles with the gas molecules inside the beam. The halo muon trajectories extend along the beam and are rejected by finding parallel hits in the muon chamber.
- **HCAL noise filter:** Rejects sporadic noise from HCAL, originating from HPD sensors, readout box noise pulses etc.
- **HCAL isolated noise filter:** By comparing energy deposits in ECAL, HCAL as well as tracker, isolated noise activity are identified within HBHE using a topological algorithm.
- **ECAL trigger-primitive filter:** In some ECAL towers where precision readout is unavailable trigger primitive (TP) information is used to determine the energy loss.

If the TP value is saturated it indicates that the lost energy is significant enough to contribute to p_T^{miss} tails, such events are rejected.

- **ECAL endcap noise filter:** EE supercrystals giving anomalous pulses (mainly due to high voltage issue) are identified and events where this occurs are rejected.
- **ECAL bad calibration filter:** Some seed crystals of ECAL clusters are known to receive large energy correction (mainly in EE region). These crystals are identified using a minimum energy threshold and subsequently events where they appear are rejected.
- **Bad PF muon filter:** Low-quality muon reconstruction contributes to the p_T^{miss} calculation as mis-reconstructed PF muon or PF charged hadron. Such events are identified and rejected from the dataset.

4.2.5 Data/MC comparison

In this section the data and MC are compared while reconstructing p_T^{miss} with different algorithms. MC event yield is corrected using event-by-event weights, accounting for differences in the reconstructed vertices distribution with data, also using scale factors based on photon identification. The distributions for number of vertices (N_{vtx}) for 2017 and 2018 data are shown in Figs. 4.10 and 4.17 respectively. Kinematic variables related to the selected tight photon and leading jet in the events are shown in Figs. 4.11, 4.12 for 2017 and 4.18, 4.19 for 2018 data. The hadronic recoil momentum components are shown in Figs. 4.13 and 4.20 for 2017 and 2018 data respectively. The reconstruction algorithms considered are the uncorrected (raw) and Type 1 corrected PF and PUPPI p_T^{miss} , shown in Figs. 4.14, 4.15, 4.16 for 2017 and 4.21, 4.22, 4.23 for 2018. The components of hadronic recoil momentum and different p_T^{miss} distributions are shown together with the recommended statistical, jet energy scale, jet energy resolution and unclustered energy uncertainties in the ratio plot in colored envelope. Statistical uncertainty is shown for all distributions.

Data collected in 2017

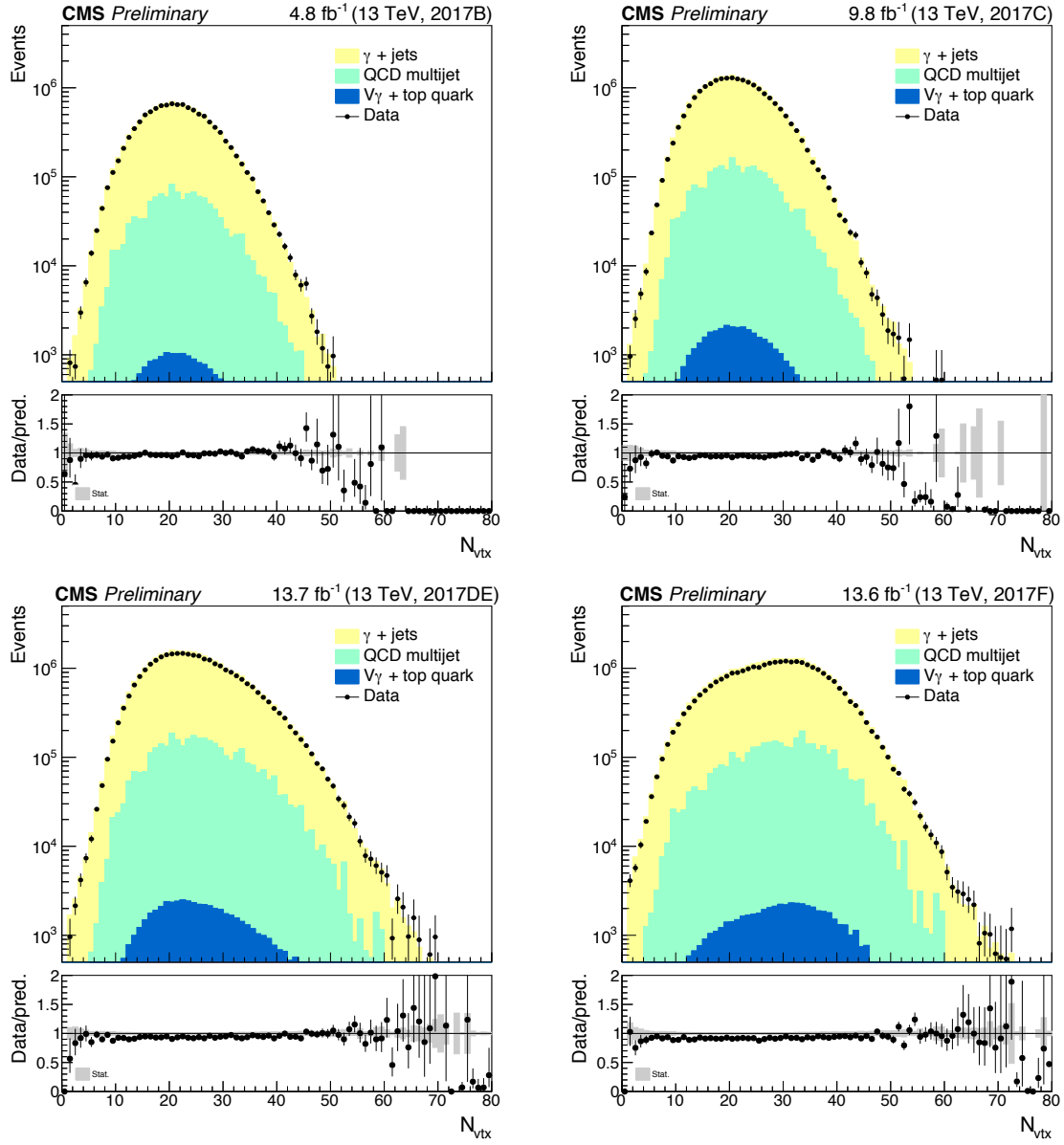
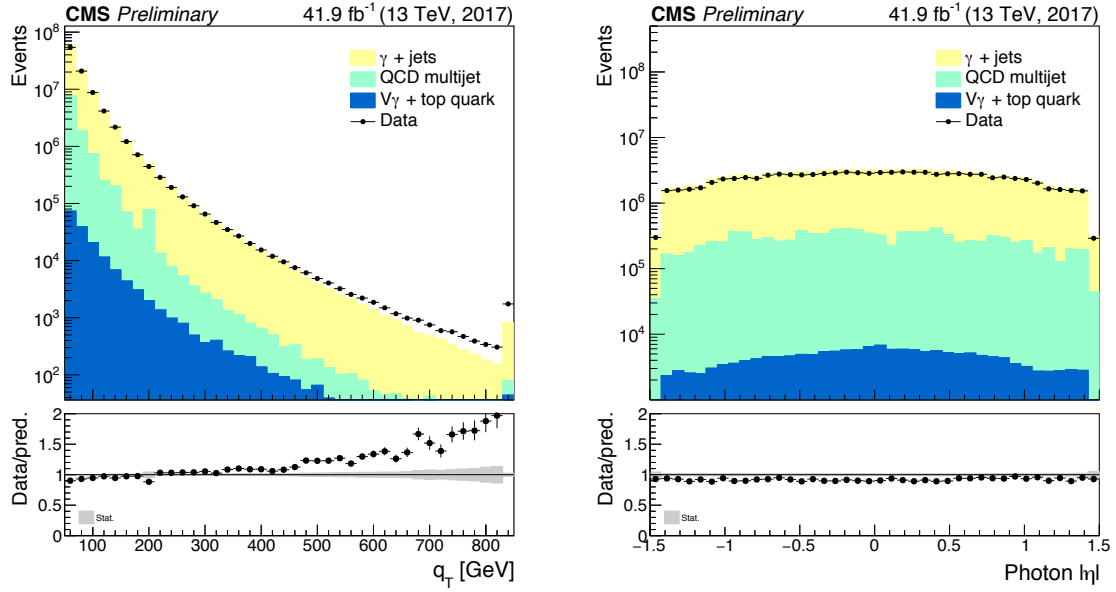
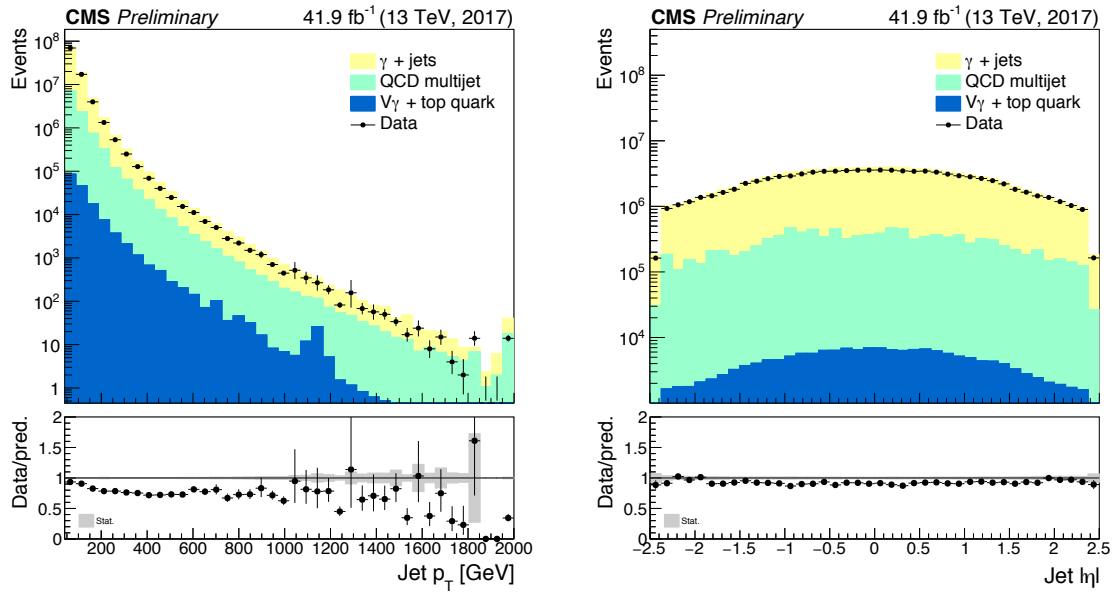


Figure 4.10: Number of vertices distributions for eras B (top left), C (top right), DE (bottom left) and F (bottom right) in the year 2017.

Figure 4.11: Distributions for p_T (left) and η (right) of the tight photon in 2017 data.Figure 4.12: Distributions for p_T (left) and η (right) of the leading jet in 2017 data.

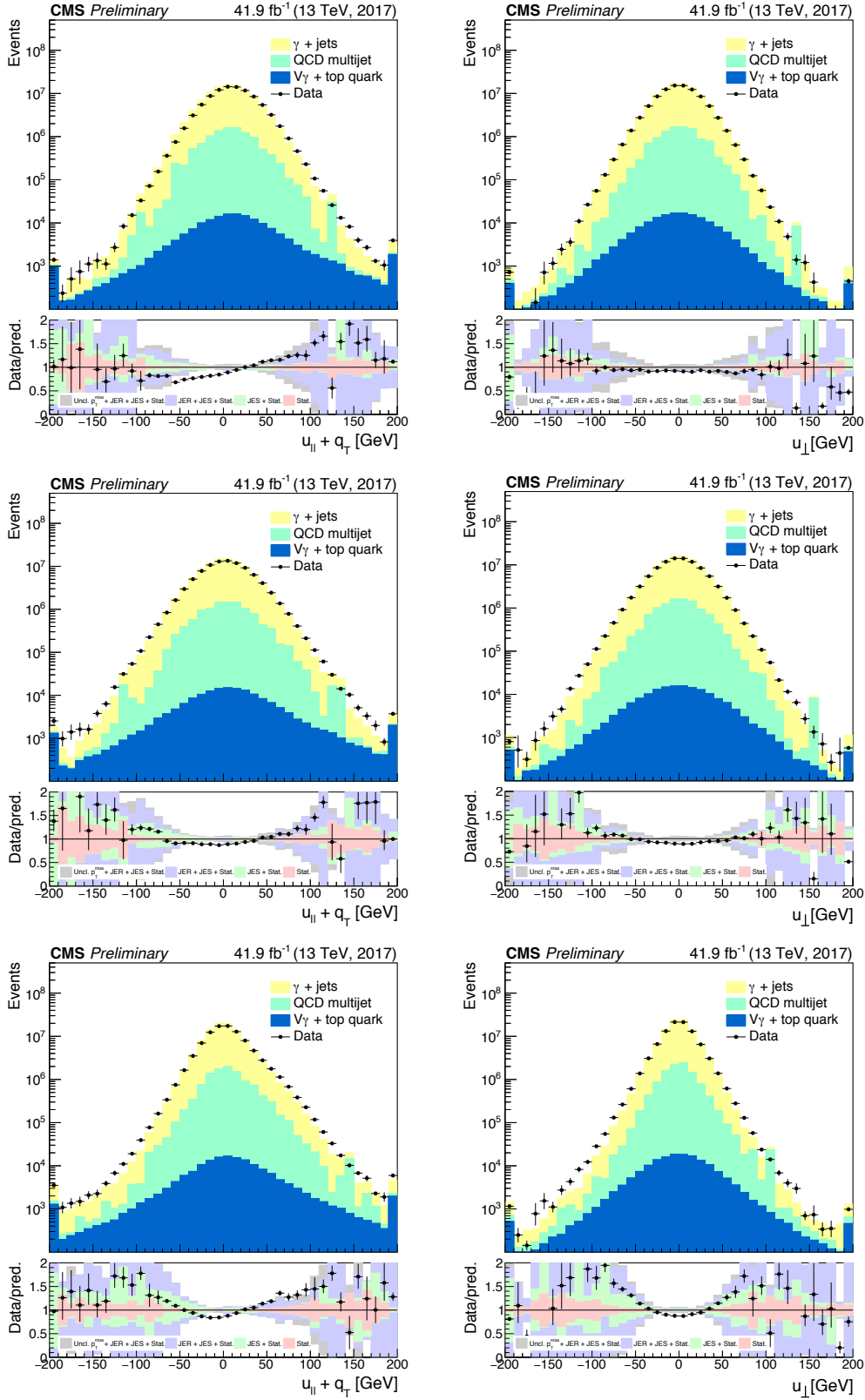


Figure 4.13: Distributions of the hadronic recoil components $u_{||}$ (left) and $u_{\perp}$ (right) in 2017 data for raw (top), Type 1 PF (middle), Type 1 PUPPI (bottom) reconstruction.

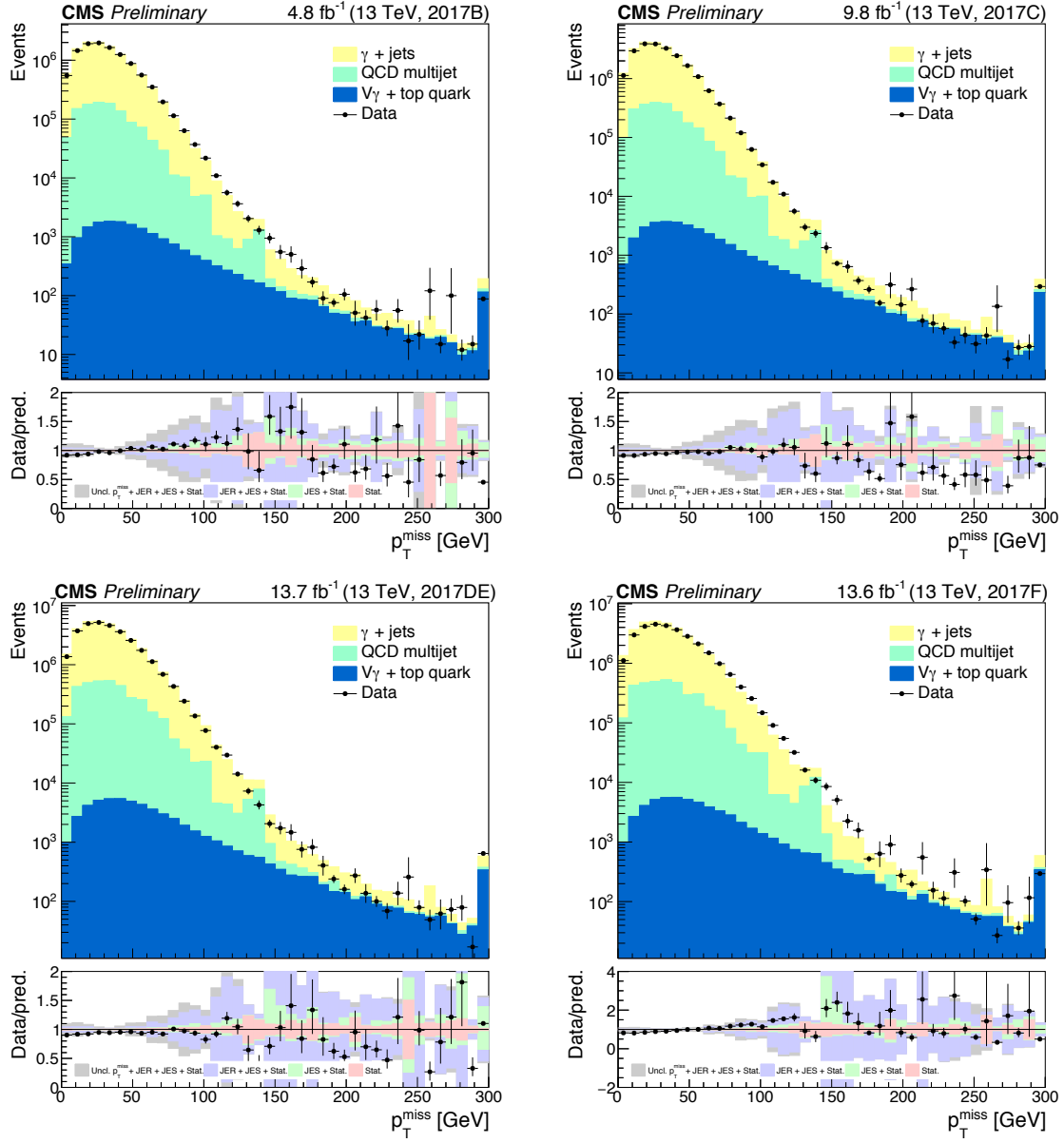


Figure 4.14: Raw p_T^{miss} distributions in the selected events for eras B (top left), C (top right), DE (bottom left) and F (bottom right) in 2017 data. Note large disagreement in the high p_T^{miss} region in era F, which is covered mostly by p_T^{miss} reconstruction uncertainties.

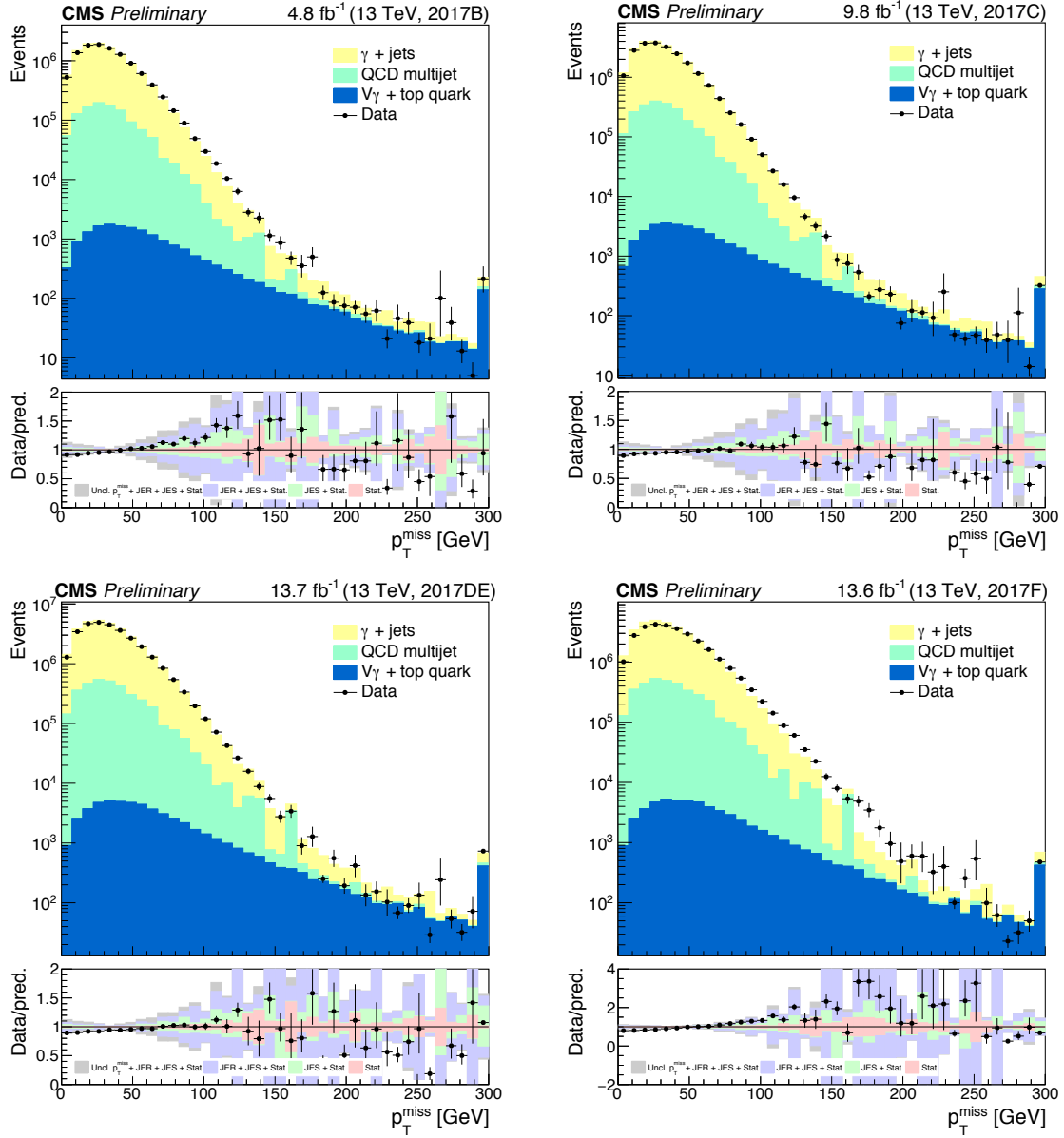


Figure 4.15: Type 1 PF p_T^{miss} distributions in the selected events for eras B (top left), C (top right), DE (bottom left) and F (bottom right) in 2017 data. With Type 1 correction agreement improves in the bulk region .

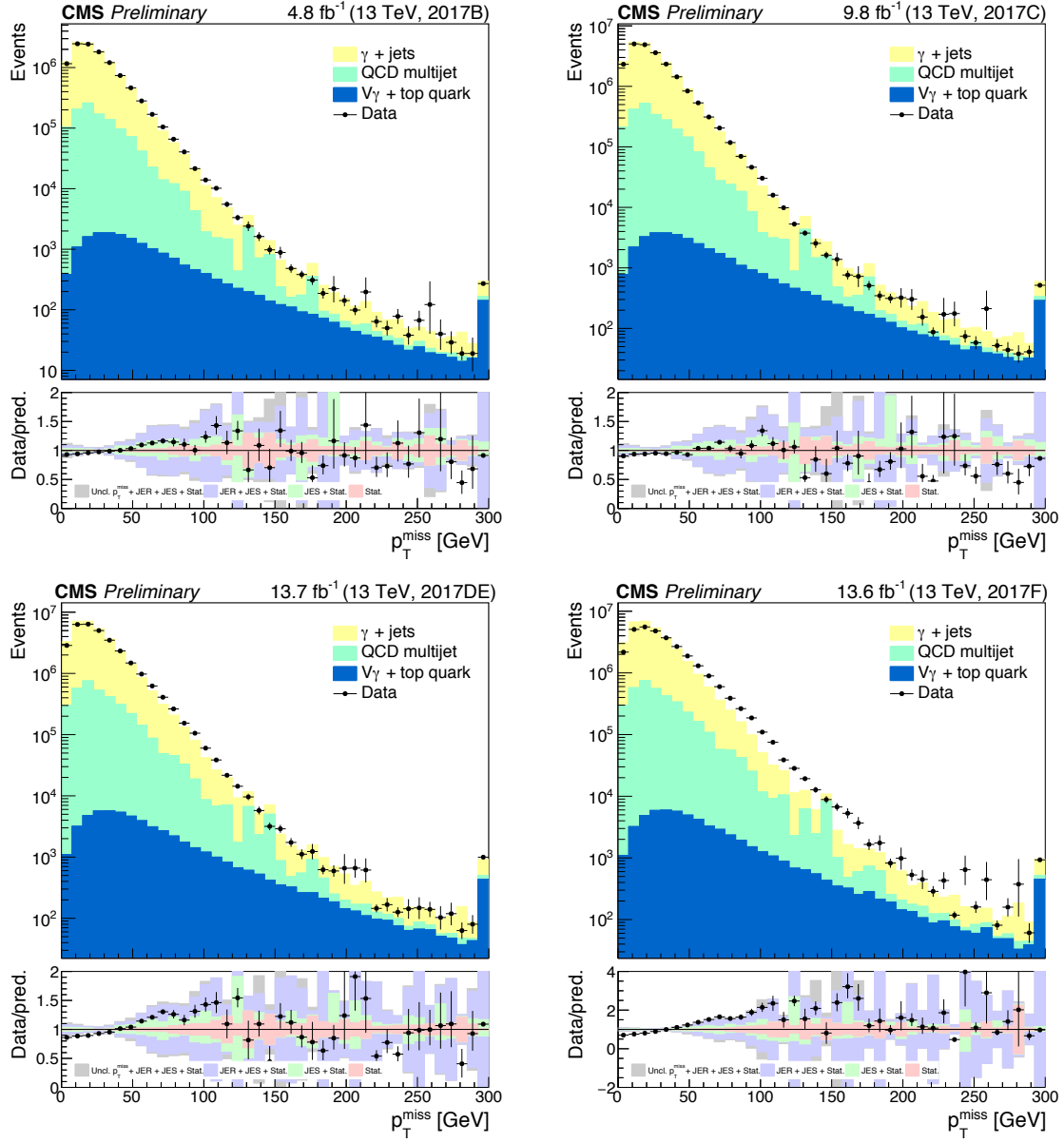


Figure 4.16: Type 1 PUPPI p_T^{miss} distributions in the selected events for eras B (top left), C (top right), DE (bottom left) and F (bottom right) in 2017 data. The agreement between data/MC is similar to Type 1 PF algorithm after the Eefix recipe is implemented.

Data collected in 2018

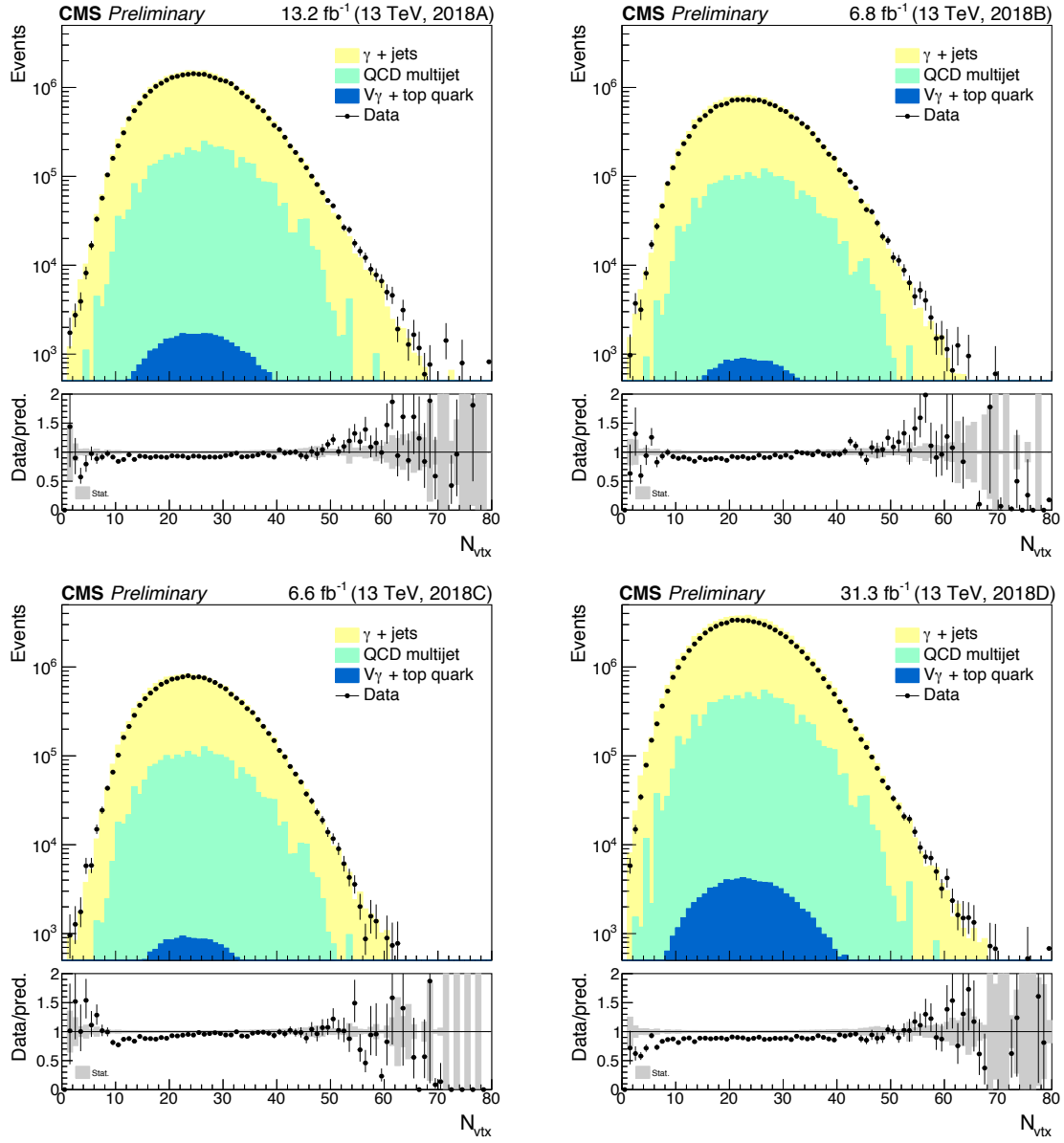


Figure 4.17: Number of vertices distributions for eras A (top left), B (top right), C (bottom left), and D (bottom right) in the year 2018.

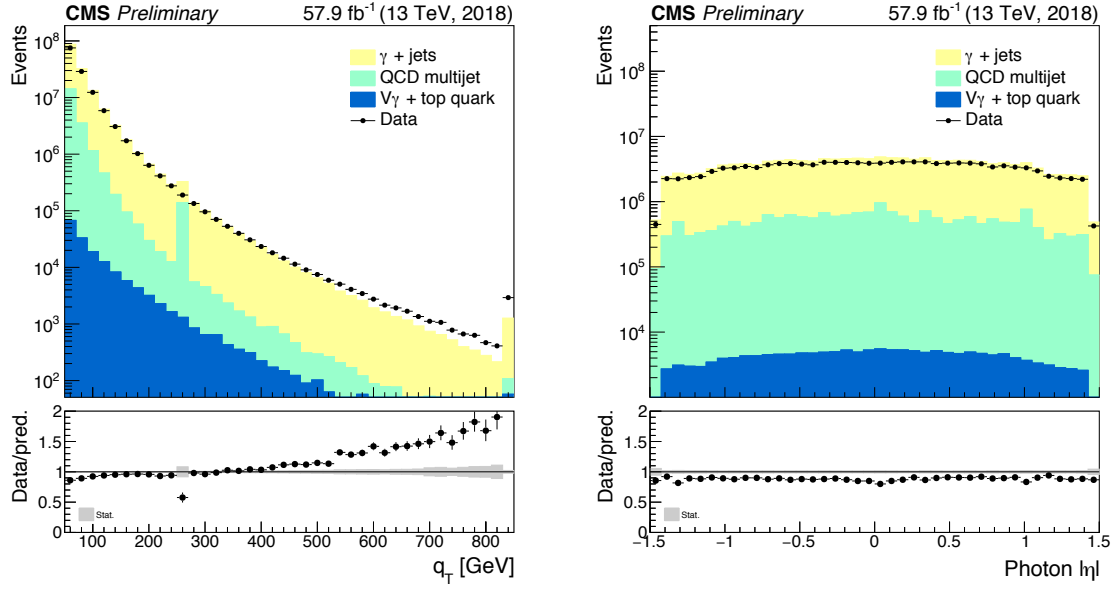


Figure 4.18: Distributions for p_T (left) and η (right) of the tight photon in 2018 data.

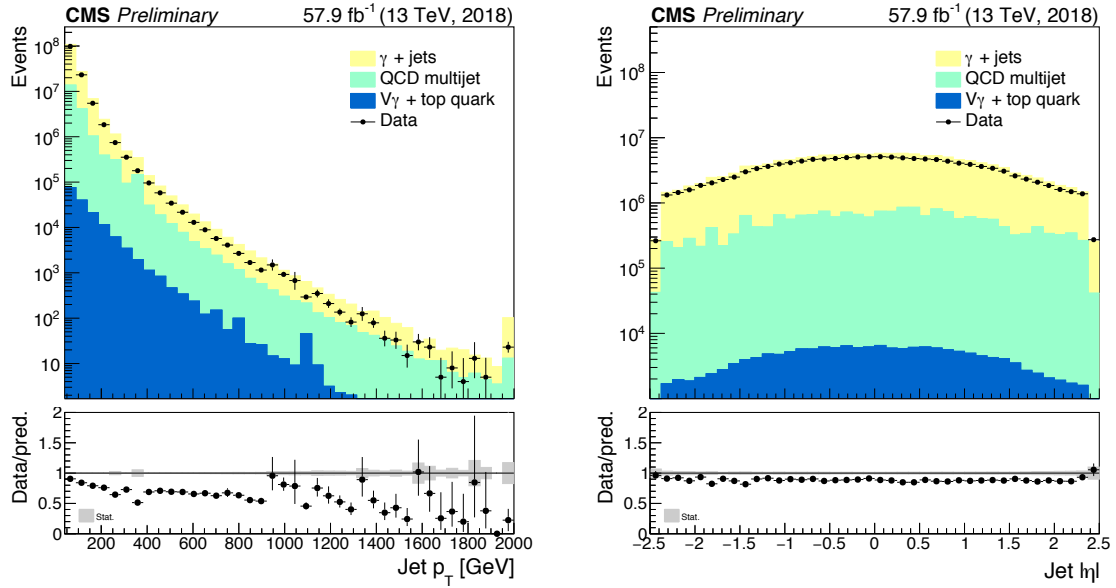


Figure 4.19: Distributions for p_T (left) and η (right) of the leading jet in 2018 data.

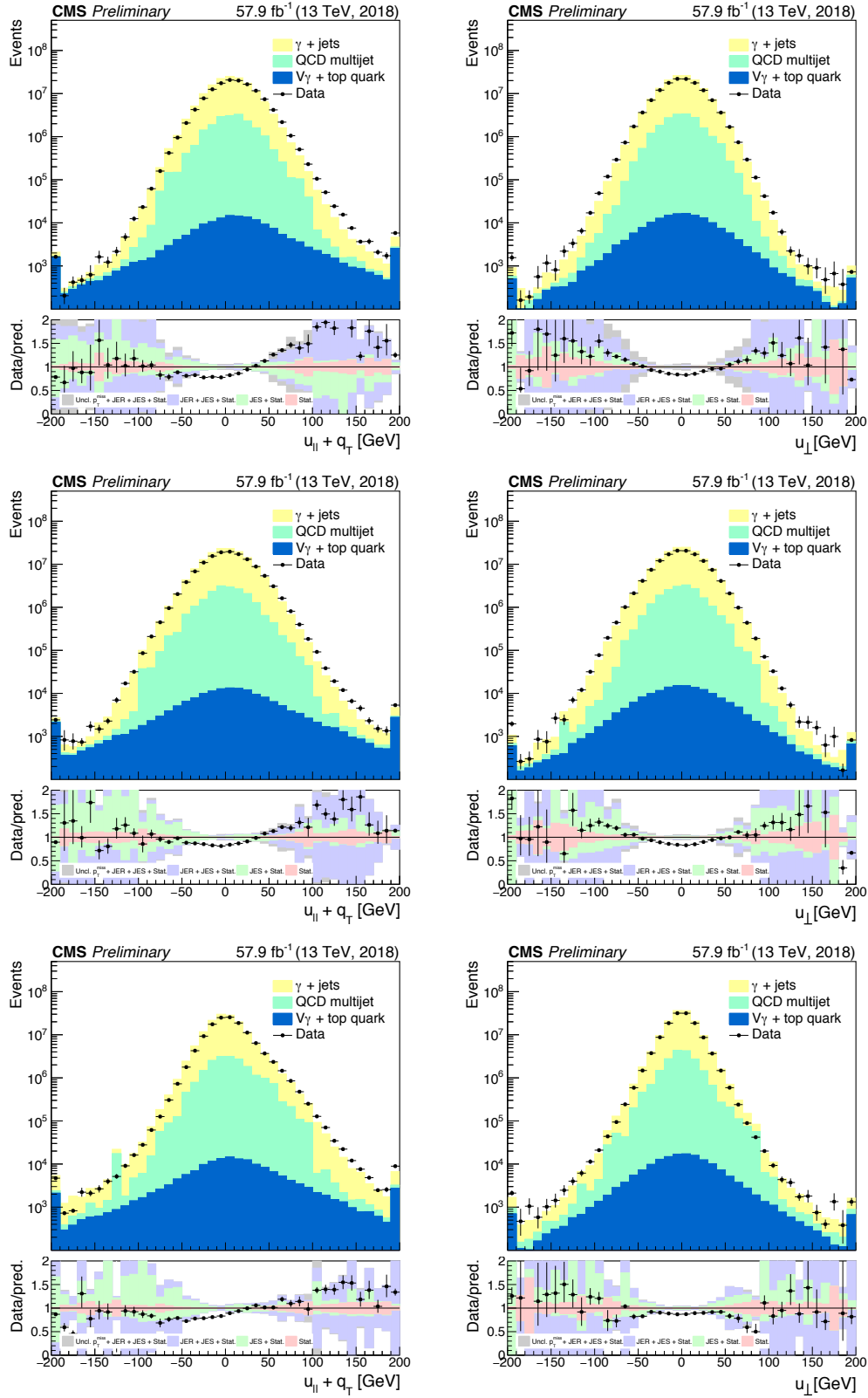


Figure 4.20: Distributions of the hadronic recoil components $u_{\parallel}$ (left) and $u_{\perp}$ (right) in 2018 data for raw (top), Type 1 PF (middle), Type 1 PUPPI (bottom) reconstruction.

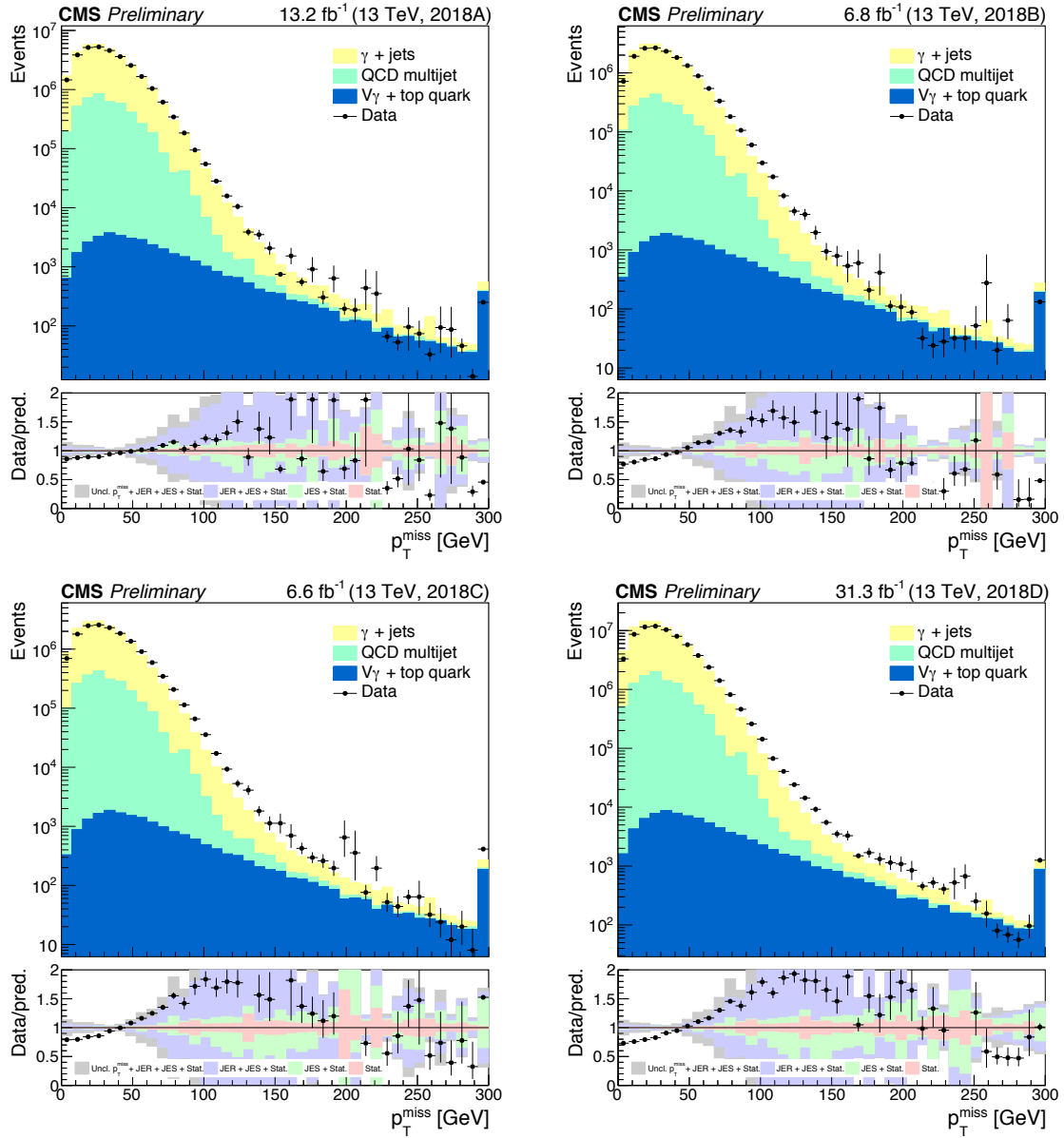


Figure 4.21: Raw p_T^{miss} distributions in the selected events for eras A (top left), B (top right), C (bottom left), and D (bottom right) in 2018 data. There is considerable disagreement in the bulk region of the p_T^{miss} distributions, specially in the later eras. Note increase in the p_T^{miss} reconstruction uncertainties in comparison to 2017.

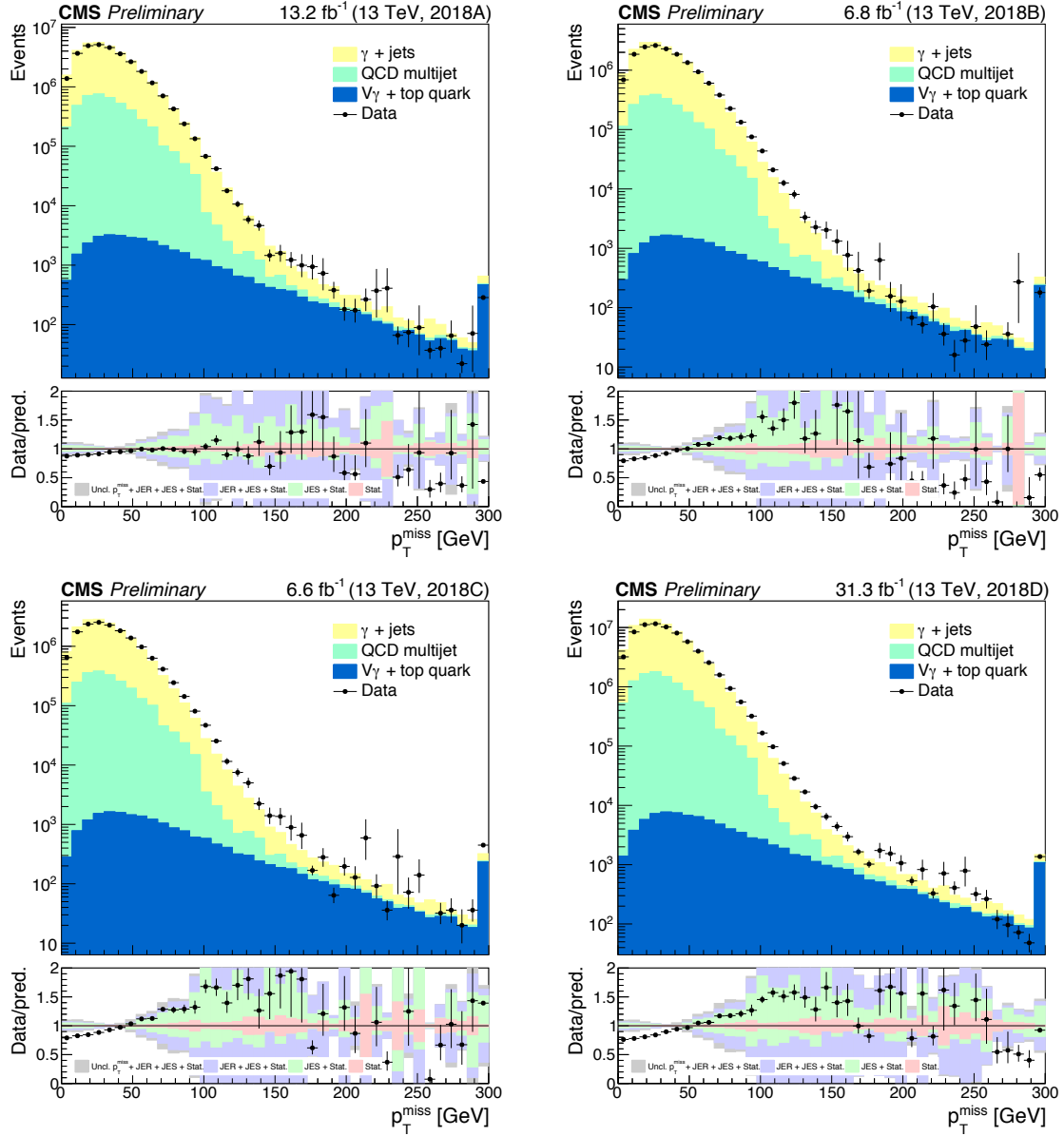


Figure 4.22: Type 1 PF p_T^{miss} distributions in the selected events for eras A (top left), B (top right), C (bottom left), and D (bottom right) in 2018 data. With Type 1 correction agreement improves, specially in the tail region.

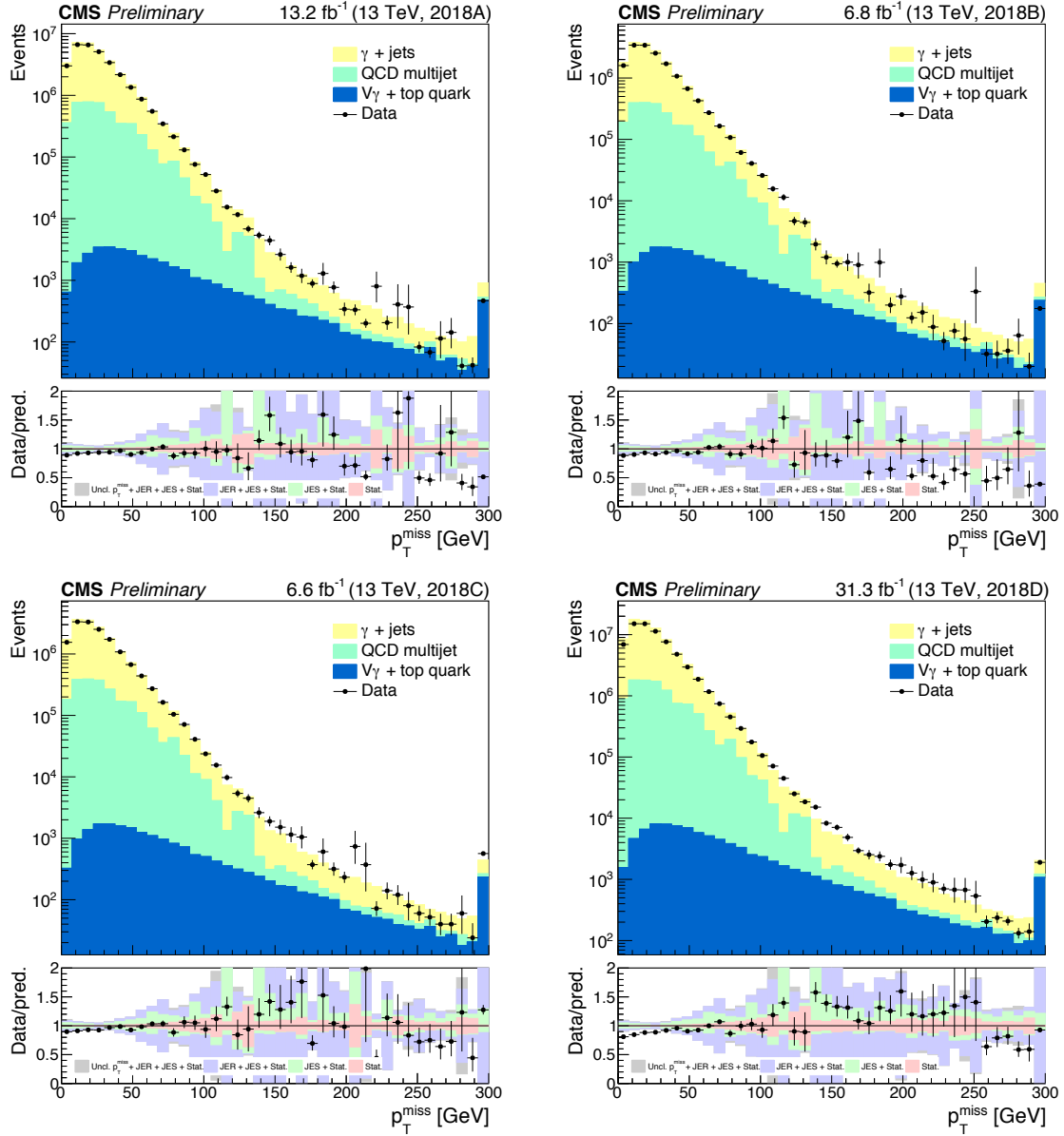


Figure 4.23: Type 1 PUPPI p_T^{miss} distributions in the selected events for eras A (top left), B (top right), C (bottom left), and D (bottom right) in 2018 data. This algorithm shows better data/MC agreement compared to PF p_T^{miss} .

Agreement between data and prediction is observed to be good in the N_{vtx} distributions for both years. Considerable change in the shape is observed over the different run periods in 2017 as the instantaneous luminosity increased from $1 \times 10^{-34} \text{ cm}^{-2}\text{s}^{-1}$ to $1.5 \times 10^{-34} \text{ cm}^{-2}\text{s}^{-1}$. The instantaneous luminosity varied around $2 \times 10^{-34} \text{ cm}^{-2}\text{s}^{-1}$ in 2018. Disagreement observed in the higher photon p_T is likely due to usage of NLO scale factors meant for 2016 γ +jets MC samples. Scale factors for current samples are not available as yet. The jet p_T distribution disagreement is expected to be within jet energy scale uncertainties, which is not shown in these plots.

4.2.6 Performance of reconstruction

The quantity $-\langle u_{\parallel} \rangle / q_T$ is denoted as the scale or response of the p_T^{miss} reconstruction algorithm. The distribution of $-\langle u_{\parallel} \rangle / q_T$ against q_T is referred to as the response curve, ideally which is a straight line with zero slope at $-\langle u_{\parallel} \rangle / q_T = 1$. The quantities $u_{\parallel} + q_T$ and $u_{\perp}$ are ideally equal to zero, but due to momentum imbalance these distributions have a finite width, estimated by the root mean squared (RMS) value. Widths $\sigma(u_{\parallel})$ and $\sigma(u_{\perp})$ represent the resolution of the p_T^{miss} reconstruction algorithms, parametrized as a function of the photon transverse momentum and the number of reconstructed vertices in the selected events. In each case, the distributions of $u_{\parallel} + q_T$ and $u_{\perp}$ are corrected for the response in order to determine the resolution. Figures 4.24, 4.25 and 4.26 compare the response of the different p_T^{miss} algorithms for years 2017 and 2018. The resolutions as function of photon transverse momentum are shown in Figs. 4.27, 4.28, 4.29 and as function of number of reconstructed vertices in Figs. 4.30, 4.31, 4.32. Error bars on the plots denote statistical uncertainty on the data.

Without the Type 1 correction the response is expected to be much lower than unity in case of raw p_T^{miss} . For Type 1 correction in PF reconstruction, the response curves saturate at values near 0.98 for year 2017, and it is slightly worse for year 2018. The response is better than 0.98 in the year 2017 and slightly worse for year 2018 for the Type 1 corrected PUPPI algorithm. The drop in response at lower q_T values in the PUPPI algorithm is likely due to aggressive cleaning of unclustered particles during reconstruction.

The resolutions for both components of the hadronic recoil momentum are observed to be

noticeably worse for era F in the year 2017 for all reconstruction algorithms considered. Average resolutions for the raw p_T^{miss} are comparable between two years. For the Type 1 PUPPI algorithm, resolutions are better for lower values of q_T for the parallel component and, overall, it is better in case of the perpendicular component compared to the raw and Type 1 PF. Resolutions as a function of N_{vtx} are significantly better in case of Type 1 PUPPI for both components compared to the raw and Type 1 PF algorithm, which is expected.

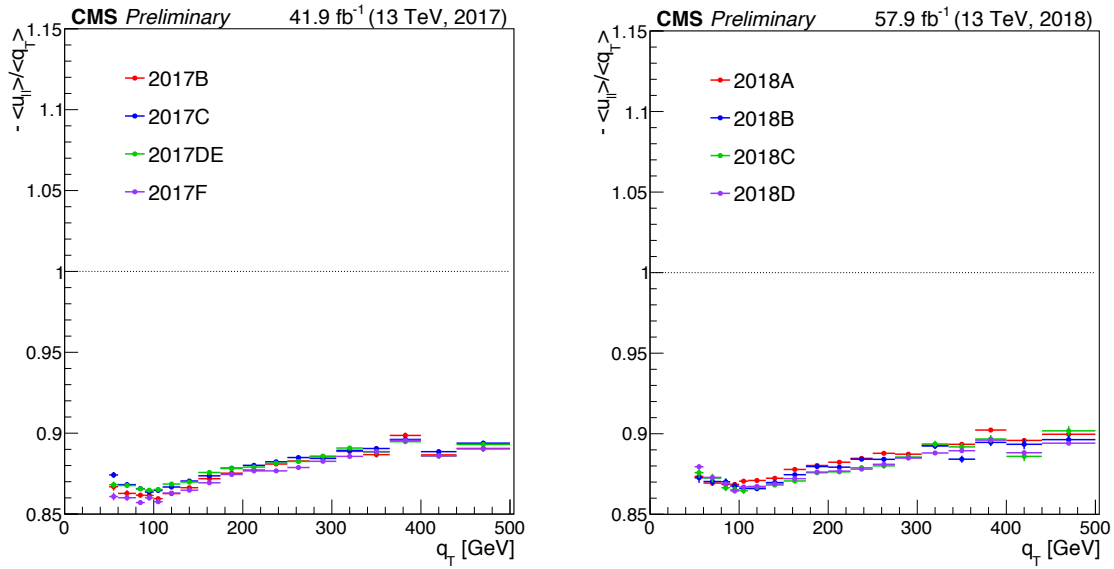


Figure 4.24: Response curves for raw p_T^{miss} in years 2017 (left) and 2018 (right). Without the Type 1 correction the response is expected to be much lower than 1.

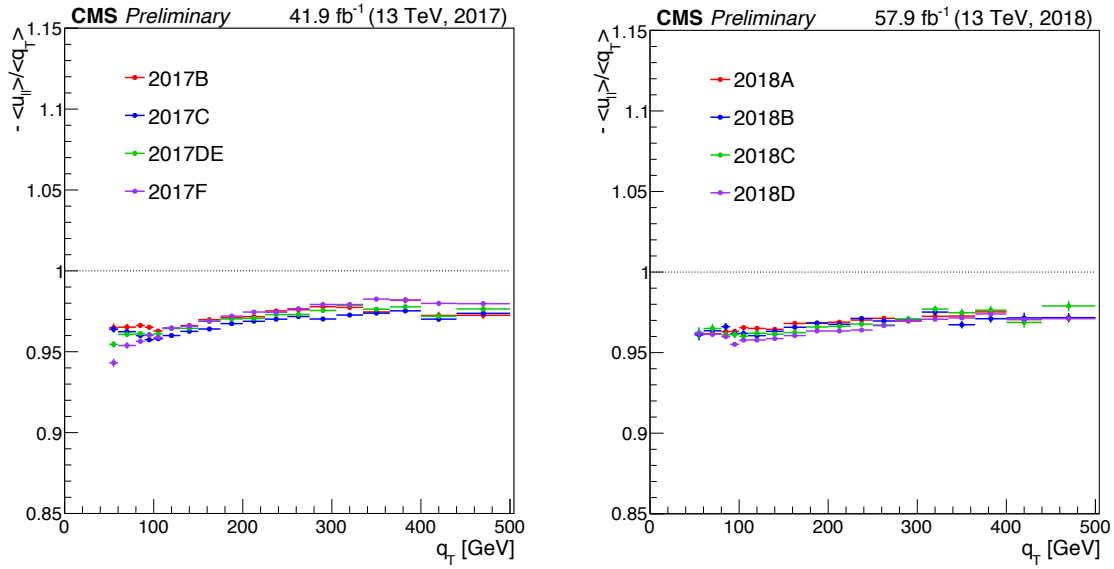


Figure 4.25: Response curves for Type 1 PF p_T^{miss} in years 2017 (left) and 2018 (right). The curves saturate at values near 0.98 for year 2017, and it is slightly worse for year 2018.

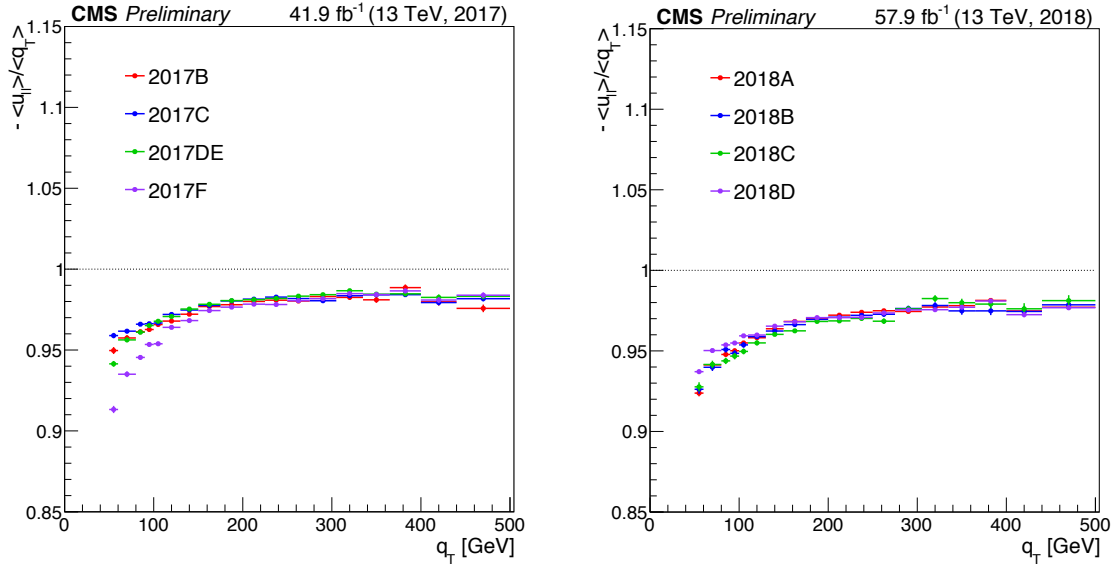


Figure 4.26: Response curves for Type 1 PUPPI p_T^{miss} in years 2017 (left) and 2018 (right). The response is better than 0.98 in the year 2017 and slightly worse for year 2018.

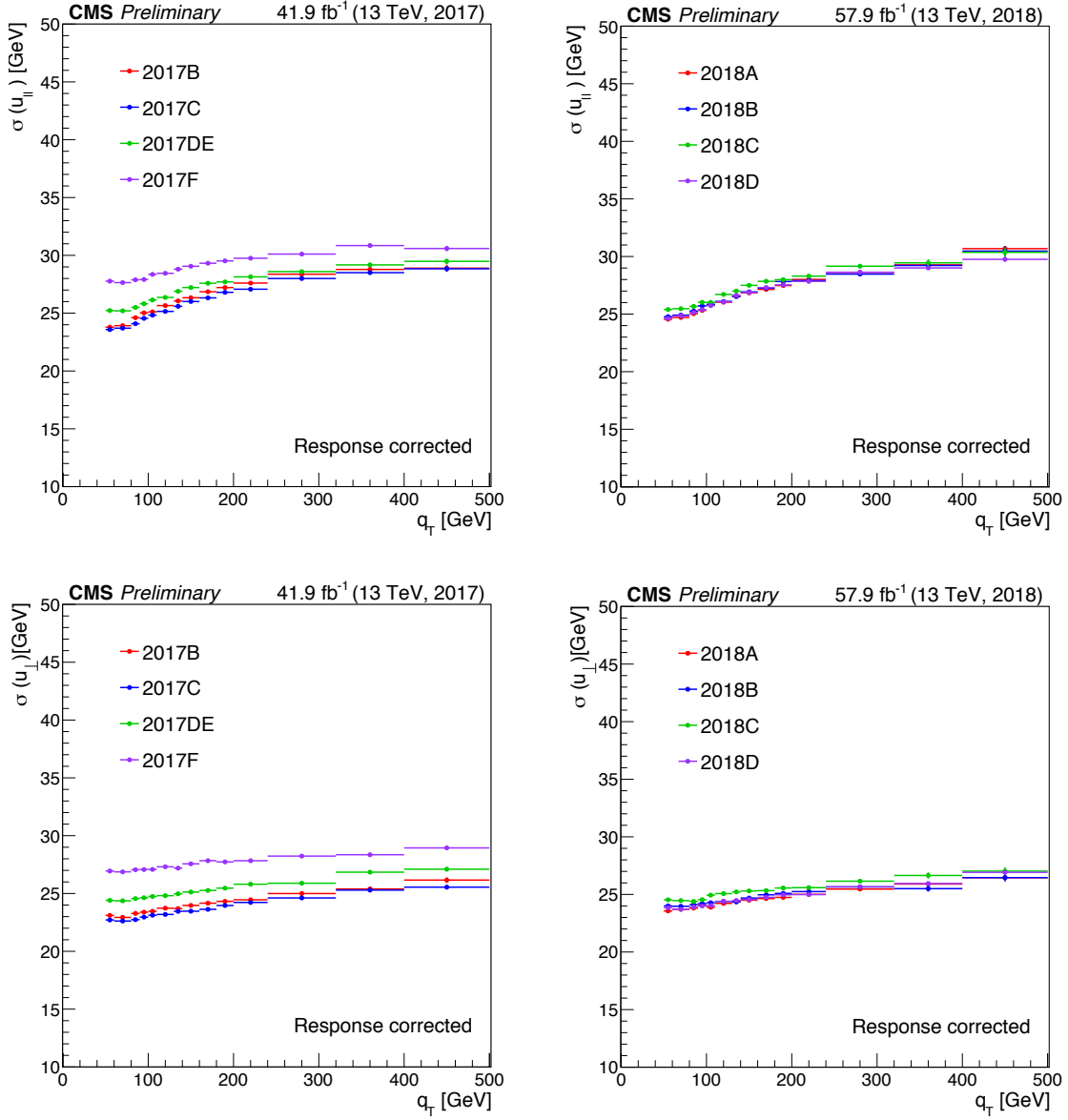


Figure 4.27: Resolution for the parallel (top) and perpendicular (bottom) components as a function of q_T for raw p_T^{miss} in years 2017 (left) and 2018 (right). The resolutions for both components are noticeably worse for era F in the year 2017. Average resolutions are comparable between two years.

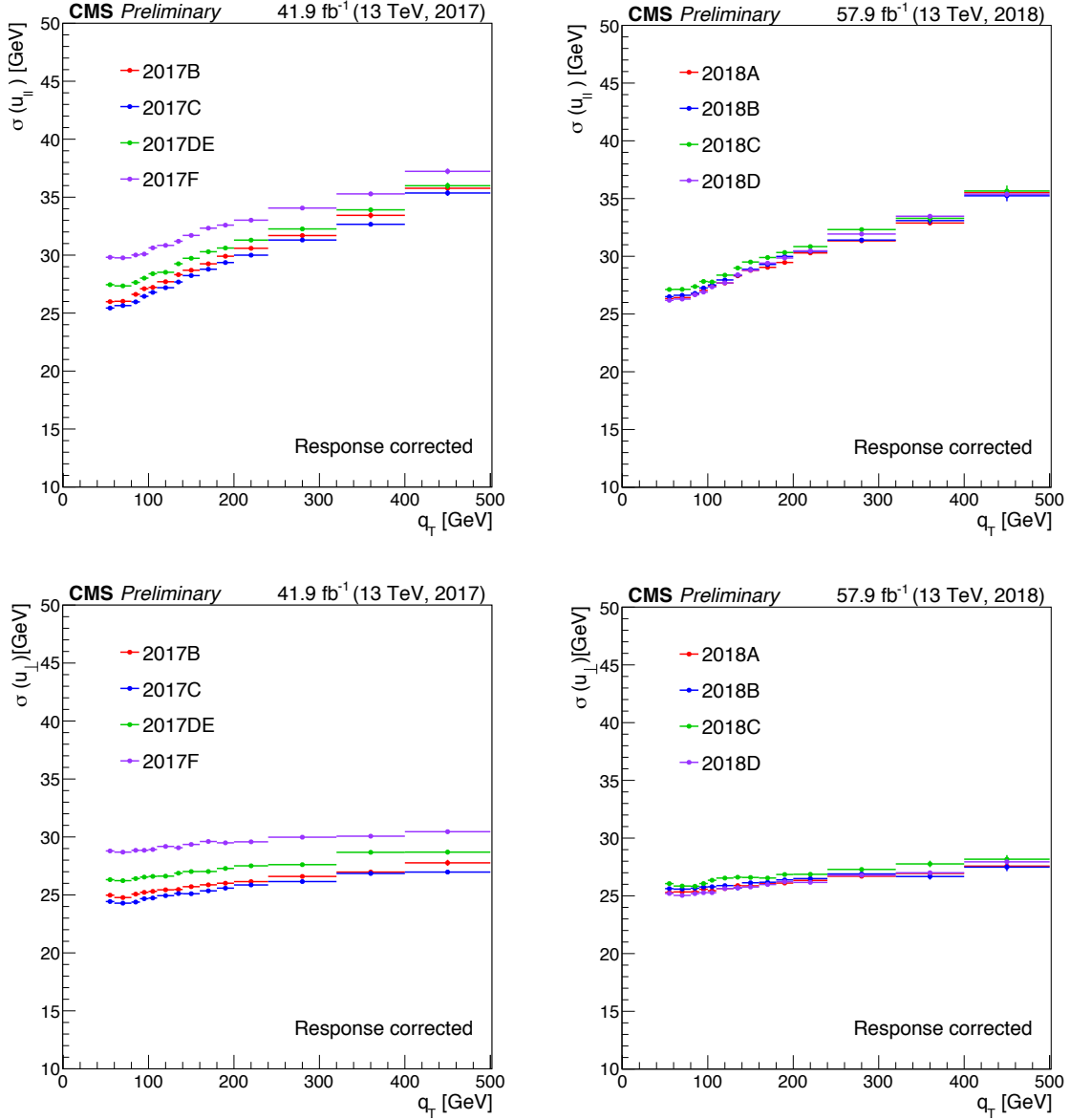


Figure 4.28: Resolution for the parallel (top) and perpendicular (bottom) components as a function of q_T for Type 1 PF p_T^{miss} in years 2017 (left) and 2018 (right). Again, resolutions are noticeably worse for era F in the year 2017. Average resolutions worsen compared to raw p_T^{miss} .

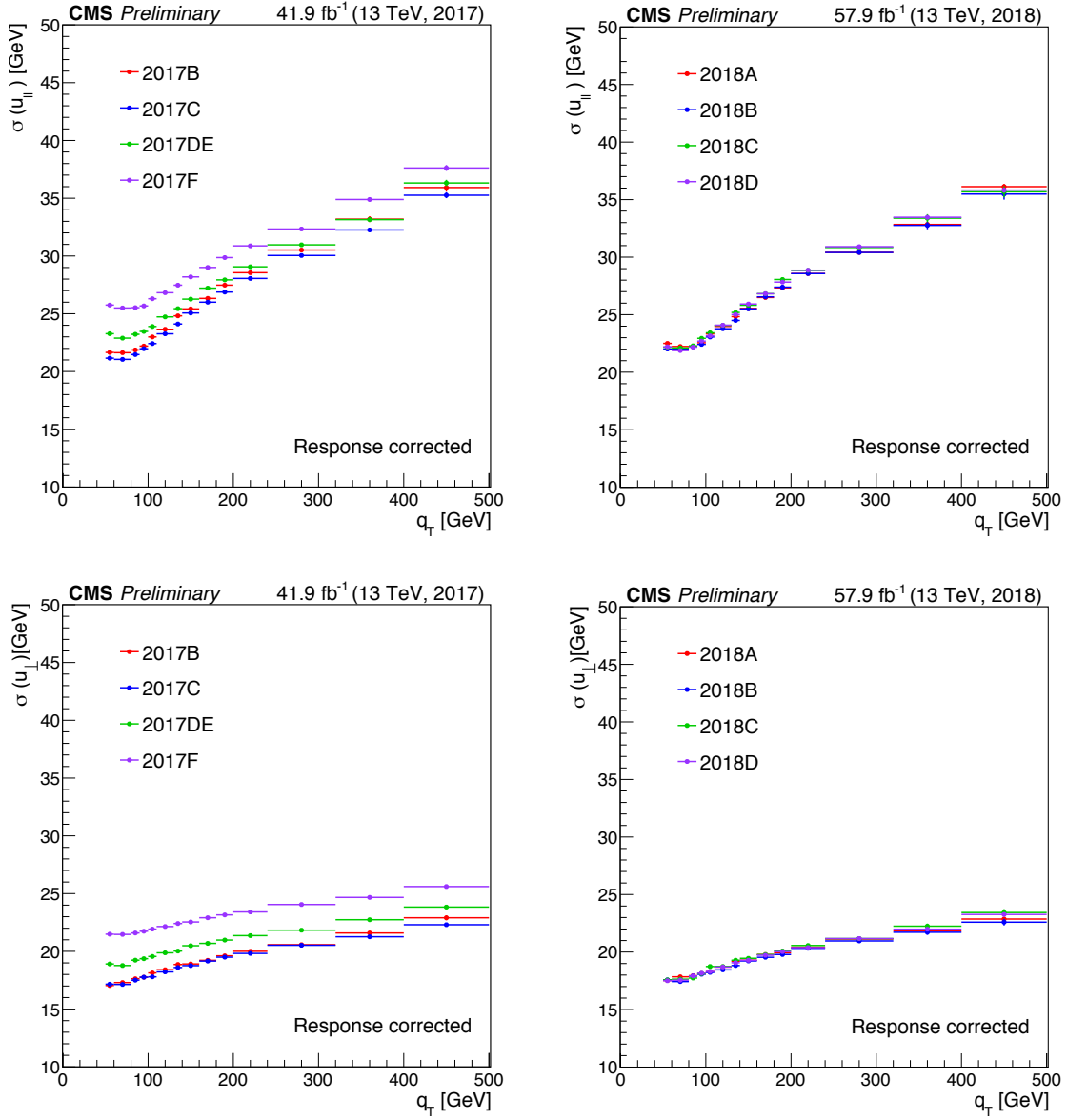


Figure 4.29: Resolution for the parallel (top) and perpendicular (bottom) components as a function of q_T for Type 1 PUPPI p_T^{miss} in years 2017 (left) and 2018 (right). Again, resolutions are noticeably worse for era F in the year 2017. Resolutions are better for lower values of q_T for the parallel component, overall better in case of the perpendicular component compared to the raw and Type 1 PF algorithm.

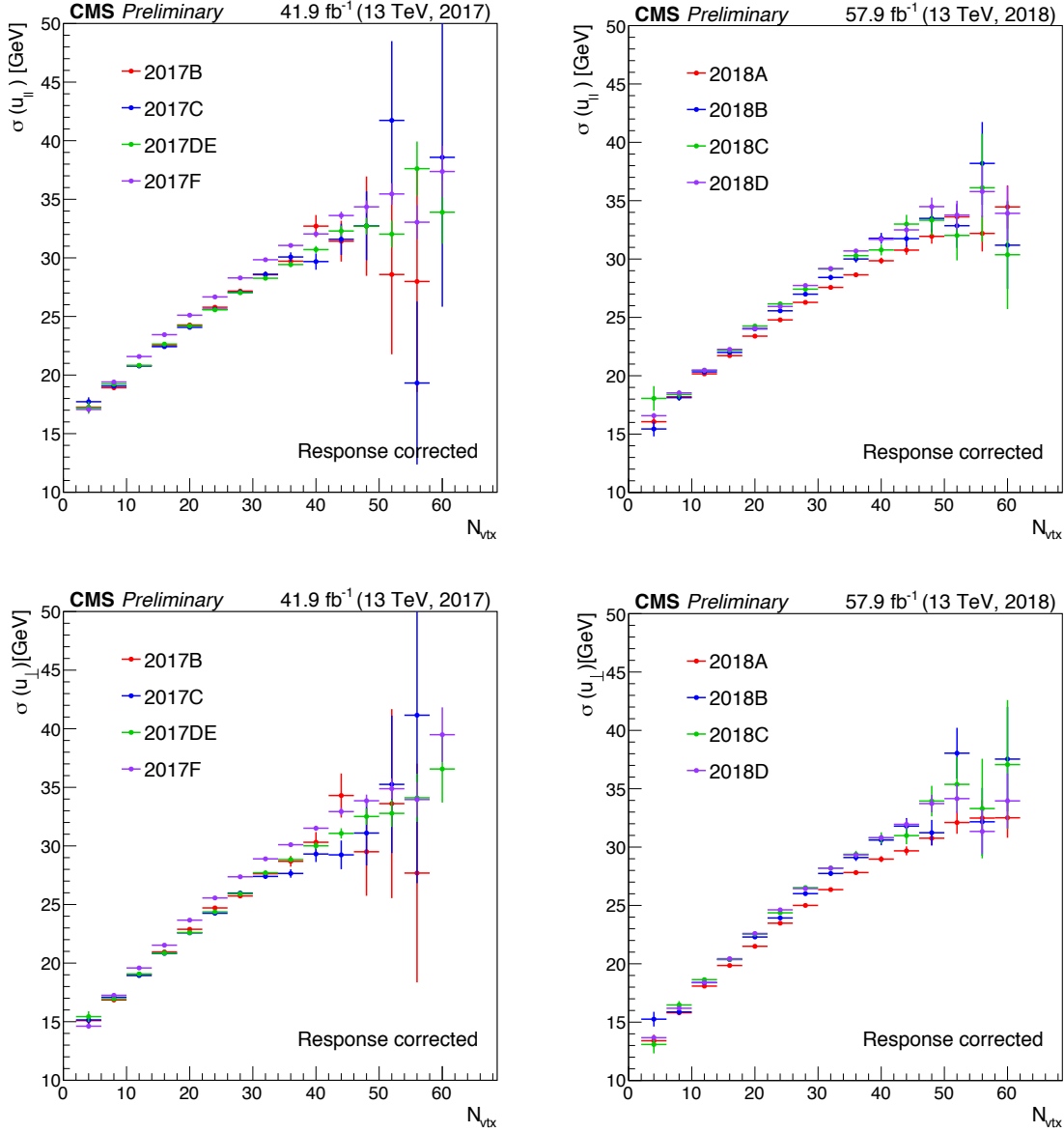


Figure 4.30: Resolution for the parallel (top) and perpendicular (bottom) components as a function of N_{vtx} for raw p_T^{miss} in years 2017 (left) and 2018 (right). Average resolutions are comparable between two years.

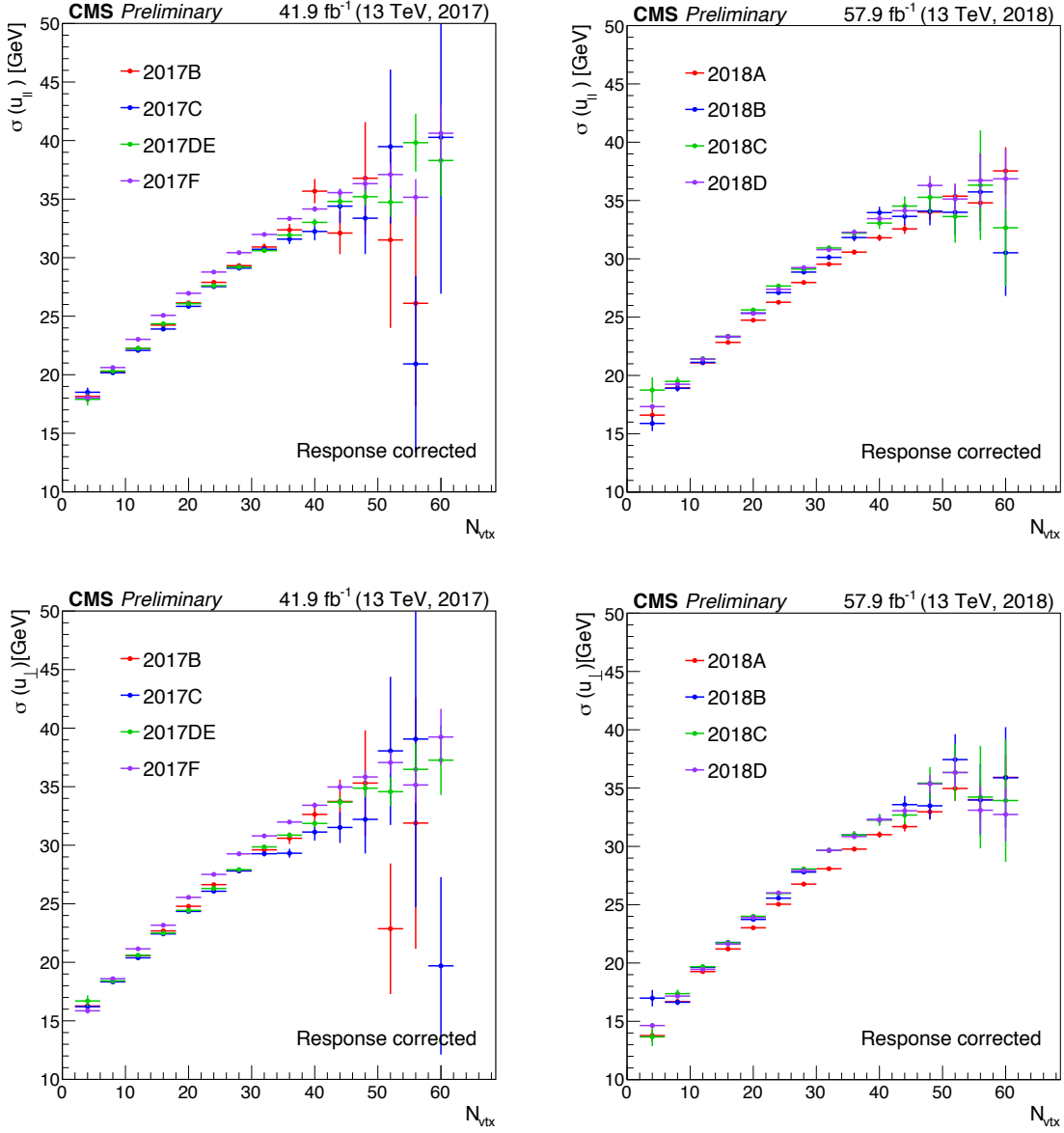


Figure 4.31: Resolution for the parallel (top) and perpendicular (bottom) components as a function of N_{vtx} for Type 1 PF p_T^{miss} in years 2017 (left) and 2018 (right). Average resolutions worsen compared to raw p_T^{miss} .

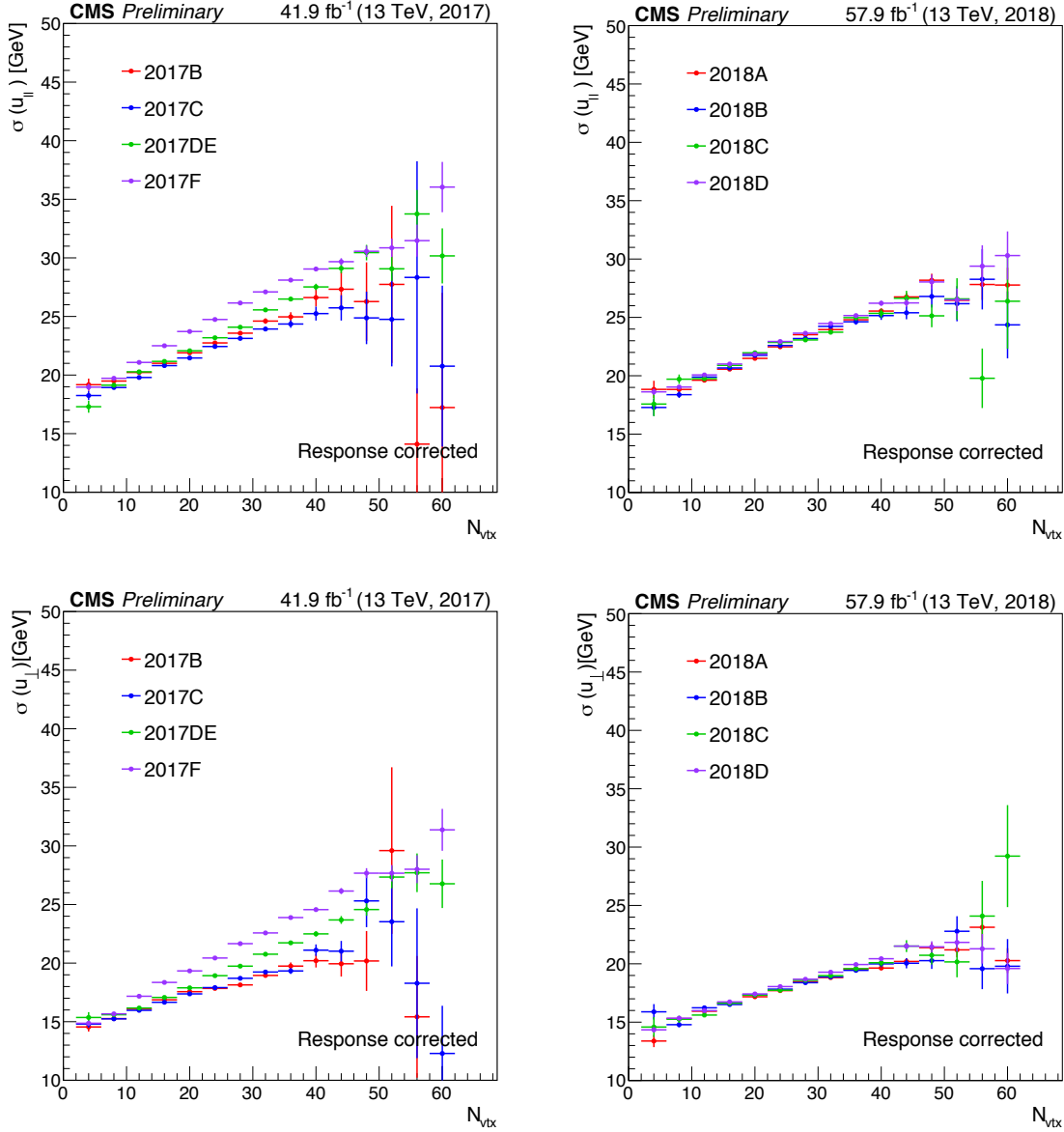


Figure 4.32: Resolution for the parallel (top) and perpendicular (bottom) components as a function of N_{vtx} for Type 1 PUPPI p_T^{miss} in years 2017 (left) and 2018 (right). Resolutions are significantly better for both components compared to the raw and Type 1 PF algorithm.

4.2.7 Studies with reconstruction recipes

Data collected in 2017

The following results show the effect of the EENoiseFix recipe on 2017 data/MC agreement and the performance of the reconstruction strategies. The initial recipe has been checked with 17Nov2017 re-reco of SinglePhoton datasets as opposed to 31Mar2018 used in the earlier sections. An earlier version of the JEC is also used here and only statistical uncertainty is shown in ratio plot. Comparing Figs. 4.33 and 4.34 the effect of applying the fix on Type 1 PF p_T^{miss} reconstruction is visible in the high p_T^{miss} region.

A re-reco (09May2018) for 2017, era F dataset alone is produced to check the agreement as an example for the Ultra Legacy reconstruction. In Fig. 4.35, the nominal and modified recipes in this re-reco are shown with the dedicated JEC. The agreement improves in the nominal reconstruction, compared to Fig. 4.33, modified recipe performs similar.

Figure 4.36 shows the comparison of the response curves of Type 1 PF p_T^{miss} algorithm with two different versions, v6 (early) and v32 (final), of JECs meant for 2017 analysis with and without the EENoiseFix recipe applied. The right plot shows the response for the Ultra Legacy candidate dataset. The difference between v6 and v32 response is attributed to a bug in a function used in reconstruction. The scale reduces in the modified recipe because genuine jets are also rejected with the high p_T threshold. The scale of the Ultra Legacy candidate dataset is comparable to the modified recipe of 2017 re-reco.

Data collected in 2018

In the year 2018, the pileup reweighting dependence of p_T^{miss} algorithms is checked in terms of data/MC agreement. Figure 4.37 compares the official pileup reweighting procedure to the dataset specific number of reconstructed vertices reweighting. Agreement improves with N_{vtx} reweighting for raw and Type 1 PF p_T^{miss} . Type 1 PUPPI p_T^{miss} reconstruction is observed to be mostly independent of the reweighting method, as expected. In order to improve the data/MC agreement in case of Type 1 PF p_T^{miss} , corrections are implemented as a function of interaction vertices, as given by Eq. 4.6. The resulting agreement is shown in Fig. 4.38 which can be compared to Fig. 4.22

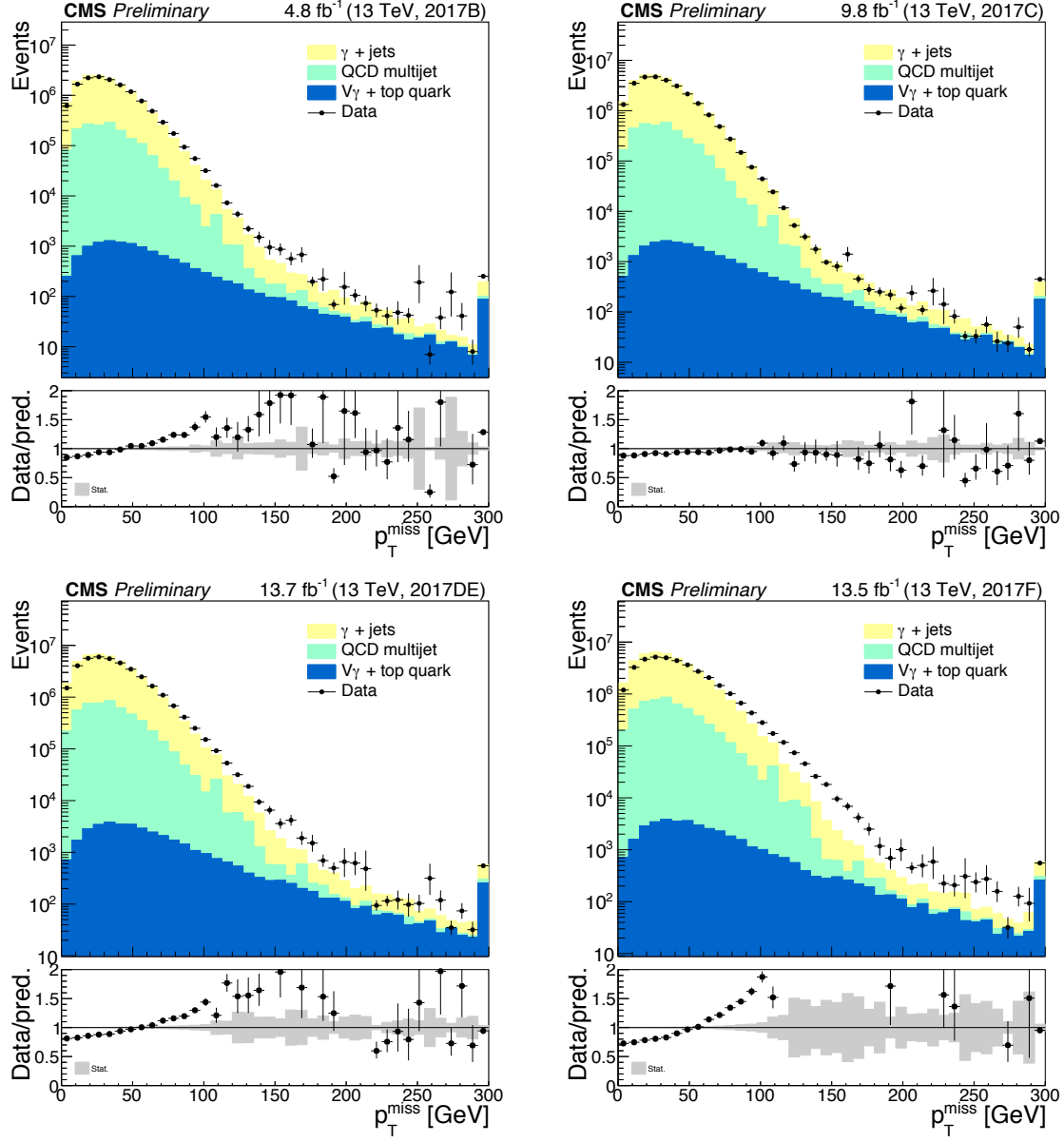


Figure 4.33: Type 1 PF p_T^{miss} distributions with nominal reconstruction for eras B (top left), C (top right), DE (bottom left) and F (bottom right) in 2017 data.

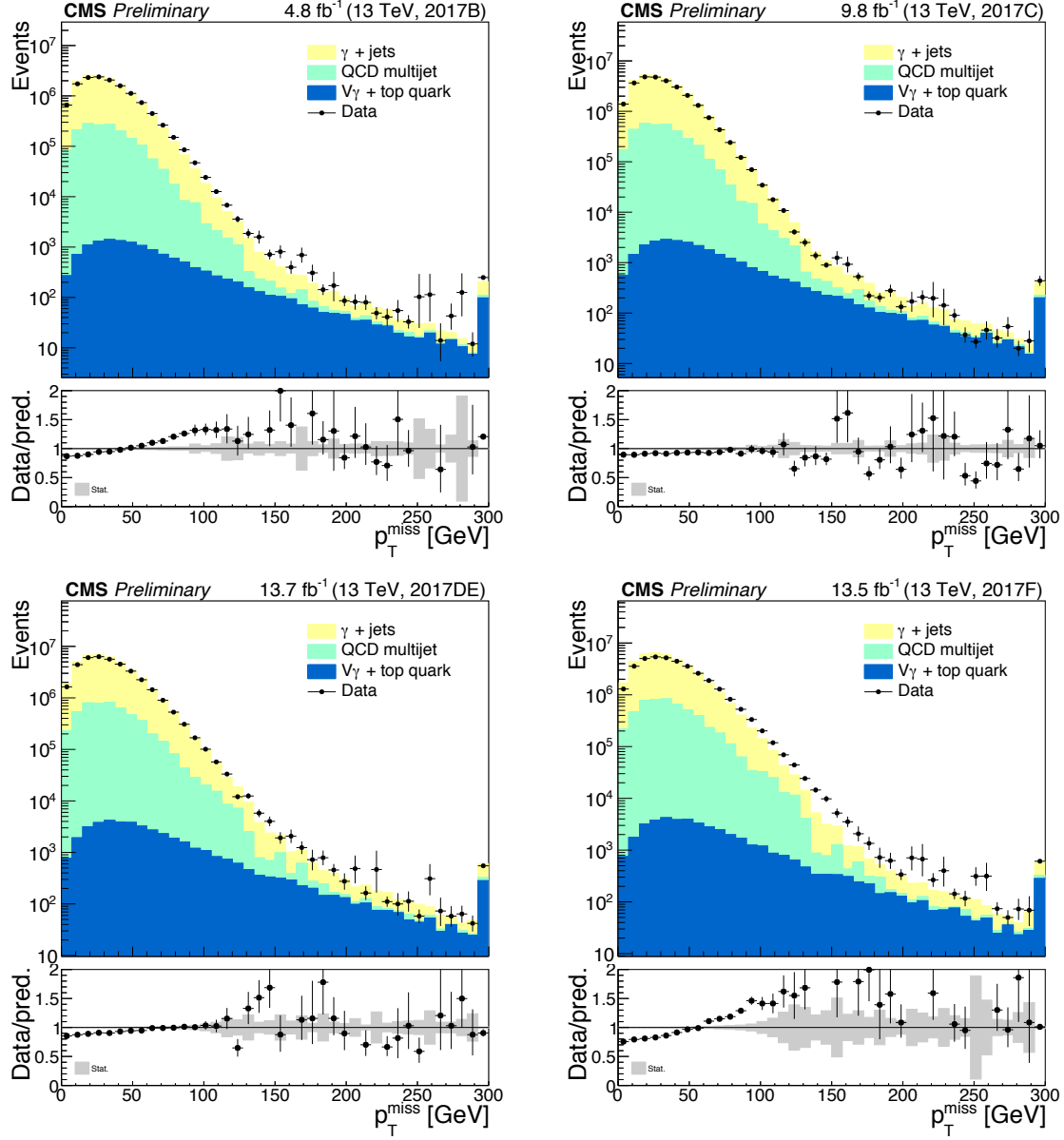


Figure 4.34: Type 1 PF p_T^{miss} distributions with EENoiseFix recipe for eras B (top left), C (top right), DE (bottom left) and F (bottom right) in 2017 data.

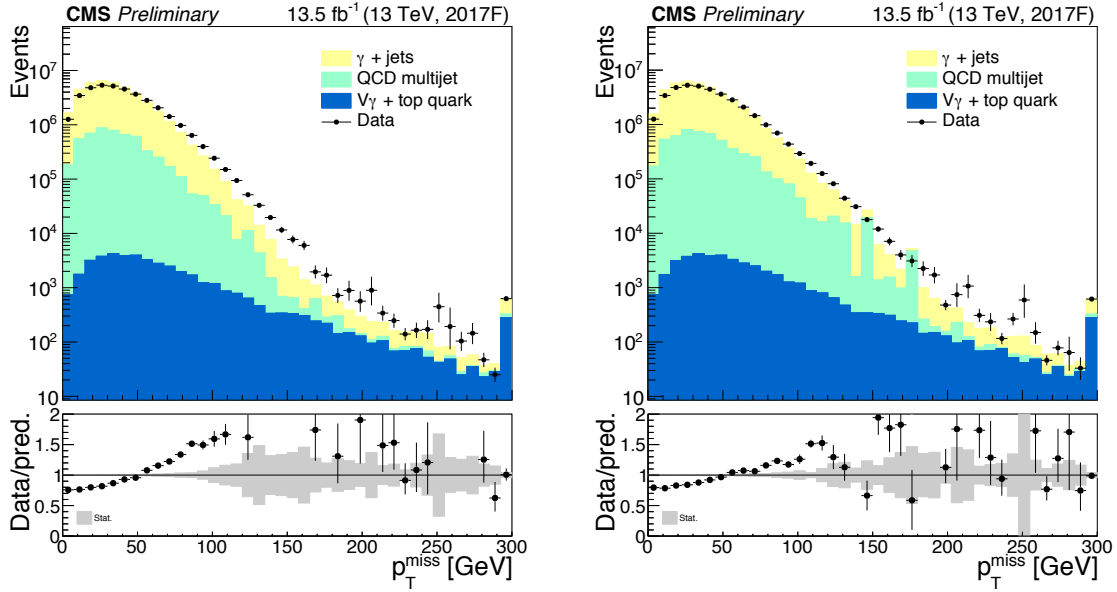


Figure 4.35: Type 1 PF p_T^{miss} distributions with Ultra Legacy candidate dataset, nominal recipe on left and with EENoiseFix on right.

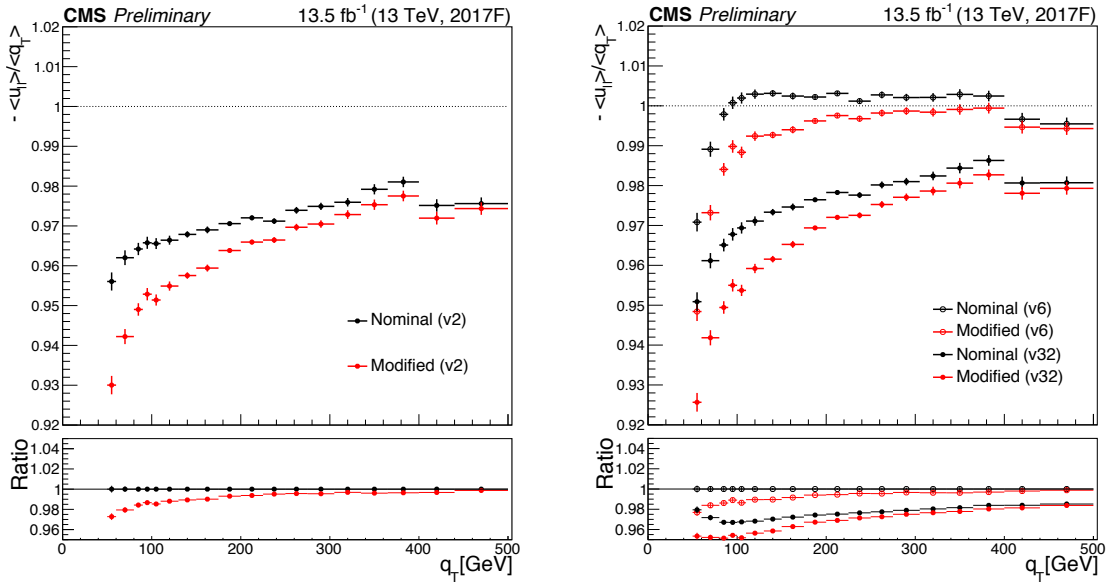


Figure 4.36: Response of the Type 1 PF p_T^{miss} algorithm with Ultra Legacy candidate dataset shown in left. Right plot corresponds to 2017 re-rco dataset for comparison.

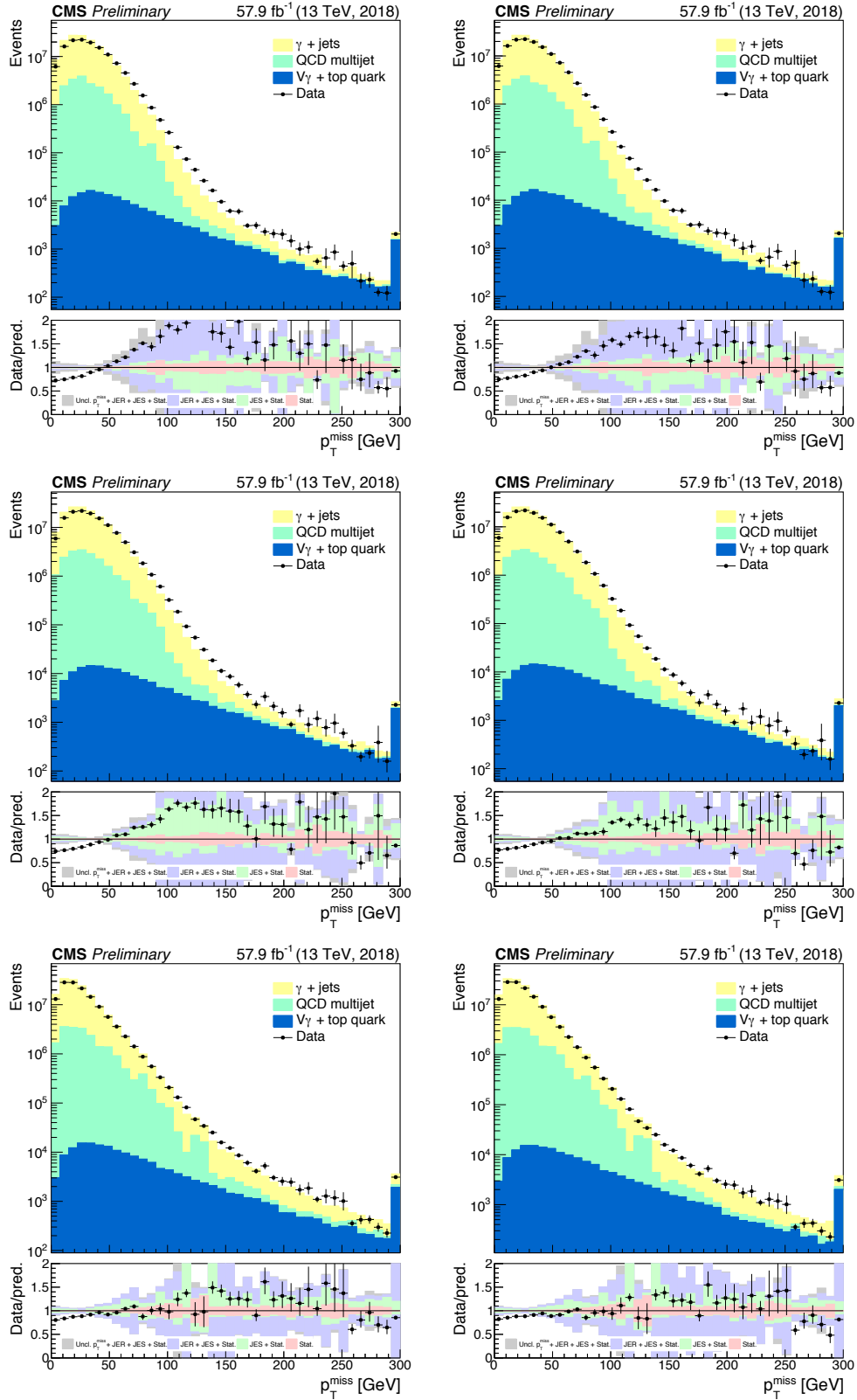


Figure 4.37: Raw (top), Type 1 PF (middle) and Type 1 PUPPI (bottom) p_T^{miss} distribution for full 2018 dataset with pileup (left) and N_{vtx} (right) reweighting.

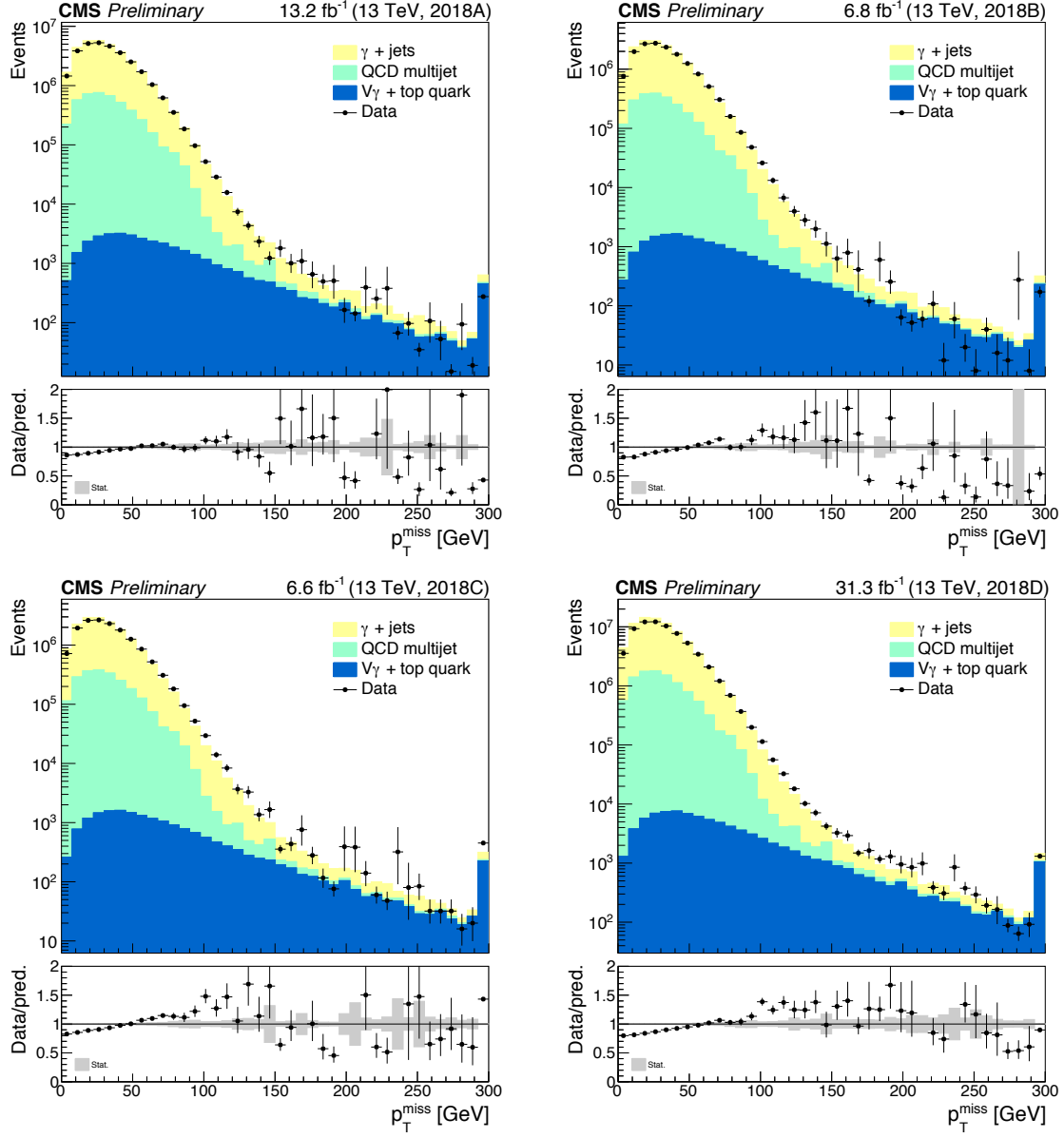


Figure 4.38: Type 1 PF p_T^{miss} distributions in the selected events for eras A (top left), B (top right), C (bottom left), and D (bottom right) in 2018 data with corrections in x and y components.

4.2.8 Conclusions

The performance of missing transverse momentum or p_T^{miss} reconstruction algorithms is studied in γ +jets events with full 2017 and 2018 datasets. This process does not have any intrinsic source of p_T^{miss} . However, several issues with the detector performance and calibration do give rise to substantial instrumental p_T^{miss} . Some of these have been addressed with dedicated recipes, while some are still being investigated. Validation of JEC recommendations using p_T^{miss} distribution, scale and resolution have been performed. The data agrees well with simulation within uncertainties, large fraction of which is due to jet energy scale dependence of p_T^{miss} . The scale and resolution measured in 2017 and 2018 data is observed to be worse compared to early 13 TeV data collected by CMS in 2016 [120]. The PUPPI algorithm shows promising performance for reconstructed p_T^{miss} in high pileup scenario, having significantly better resolutions and stability against pileup reweighting methods.

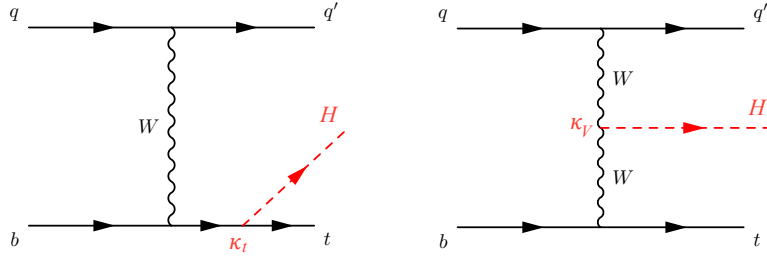
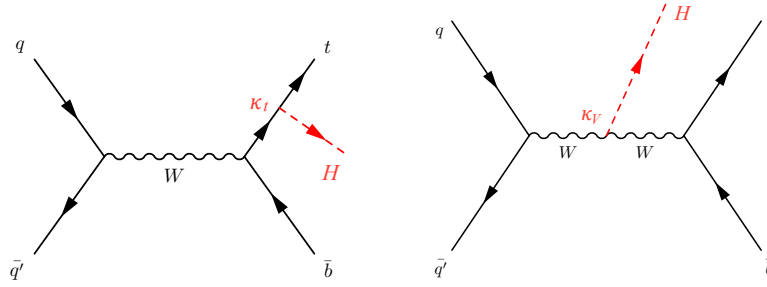
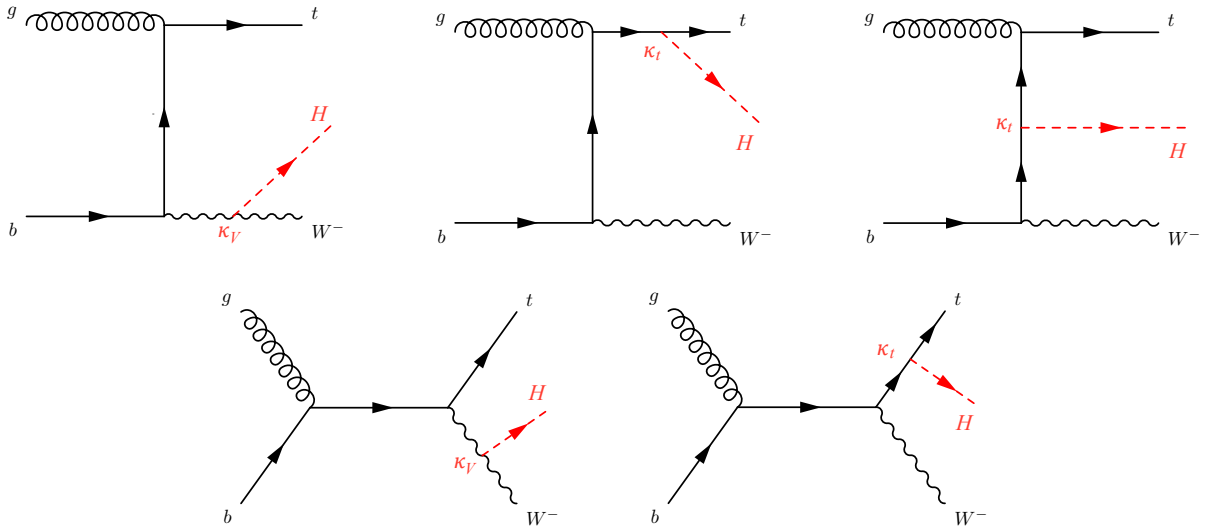
Starting next chapter, the main analysis work involving tH process is discussed. Chapter 5 gives an introduction to the different techniques involved, starting from selection of final states to statistical treatment.

Chapter 5

Associated production of the Higgs boson with single top quark

The main content of this thesis addresses the study of the tH production process, introduced in Section [1.4.3](#), carried out using Run 2 LHC data. Similar to the single top production at the leading order in QCD, the associated production of the Higgs boson with single top proceeds via three mechanisms: (i) t-channel where the virtual space-like W interacts with a b quark to produce the top, (ii) s-channel where the virtual W is time-like, and (iii) W -associated where a real on-shell W is emitted in the final state. The Feynman diagrams contributing to the tH production process are shown in Figs [5.1](#), [5.2](#) and [5.3](#) along with the Higgs couplings involved. The Higgs boson is radiated off either the top quark or the W boson, resulting in interfering diagrams with the same initial and final states. This feature of the tH production process makes it possible to study the relative sign between the top-Higgs and W -Higgs couplings [\[121\]](#). The studies of these couplings are performed using the κ -framework introduced in Section [1.4.1](#), κ_t and κ_V denoting the modifiers to the SM top-Yukawa and Higgs to vector boson couplings respectively.

The classification on the basis of the virtuality of the W boson is useful for event generation, however it is valid only at LO and in the five-flavor scheme (5FS) where the bottom quark is considered as one of the constituents of proton and modelled using PDFs. At the LHC energies the rate of s-channel production or tHb process is negligible, since initial state requires a quark and an anti-quark. The t-channel production dominates, hence-

Figure 5.1: Feynman diagrams contributing to the t-channel tH production.Figure 5.2: Feynman diagrams contributing to the s-channel tH production.Figure 5.3: Feynman diagrams contributing to the W -associated tH production.

forth referred to as tHq process. The W -associated or tHW production mode has smaller contribution at the LHC. In addition, the LO diagram of tHW interferes with $t\bar{t}H$ production if four-flavor scheme (4FS) is considered, where bottom quarks in the initial state are produced from gluon-splitting as shown in Fig. 5.4.

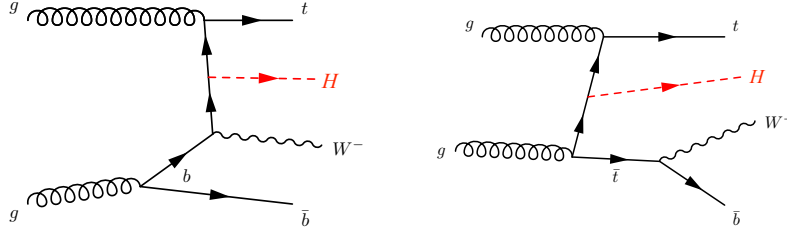


Figure 5.4: Interference of tHW production in the 4FS scheme (left) with the $t\bar{t}H$ process (right).

Cross sections of these processes in pp collisions at $\sqrt{s}=13$ TeV are evaluated to be [45, 122–124]:

$$\begin{aligned}\sigma(tHq)_{SM} &= 74.25 \text{ fb }^{+6.5\%}_{-14.9\%} (\text{scale} + \text{FS}) \pm 3.7\% (\text{PDF} + \alpha_s) \\ \sigma(tHW)_{SM} &= 15.17 \text{ fb }^{+4.9\%}_{-6.7\%} (\text{scale}) \pm 6.3\% (\text{PDF} + \alpha_s) \\ \sigma(tHb)_{SM} &= 2.879 \text{ fb }^{+2.4\%}_{-1.8\%} (\text{scale}) \pm 2.2\% (\text{PDF} + \alpha_s)\end{aligned}\tag{5.1}$$

considering $m_H = 125$ GeV, with NLO QCD accuracy at 5FS where the separation between the channels is exact. The channels start to interfere at NNLO in 5FS and at NLO in 4FS, thus contributing to the FS uncertainty of tHq cross section calculation. In the 4FS, the tHq cross section is approximately 71 fb [123]. For the analyses presented in this thesis, tHq and tHW events are generated in 4FS and 5FS respectively.

5.1 Measurements of Higgs properties

Studies detailed in this thesis consider the tHq and tHW production modes, which are referred together as tH in all subsequent discussions. These final states have the potential to determine several phenomenological characteristics of the Higgs boson, as noted in literature [121, 123, 125–130]. This thesis describes studies of the two main aspects: relative

sign between κ_t , κ_V and the CP nature of top-Yukawa coupling in the context of the Higgs characterization model [123, 131]. In addition, tHq final state can also be used to study FCNC interactions in top quark decays. Processes such as $t \rightarrow Hu$ or Hc are strongly suppressed by Glashow–Iliopoulos–Maiani (GIM) mechanism [132], however some BSM scenarios predict large enhancement in the rate of these processes [133]. These studies are beyond the scope of this thesis which focuses mainly on the Higgs boson properties.

5.1.1 Relative sign of top-Yukawa coupling

According to SM, tHq and tHW production cross sections depend strongly on the destructive interference between the contributions from the Higgs boson radiated off the top and off the W boson. However, anomalous couplings where $\kappa_t \neq 1$ or $\kappa_V \neq 1$ can significantly modify the cross sections. The parametrization of cross sections in terms of the coupling modifiers are given as [134]:

$$\begin{aligned}\sigma(tHq) &= (2.63\kappa_t^2 + 3.58\kappa_V^2 - 5.21\kappa_t\kappa_V).\sigma(tHq)_{SM} \\ \sigma(tHW) &= (2.91\kappa_t^2 + 2.31\kappa_V^2 - 4.22\kappa_t\kappa_V).\sigma(tHW)_{SM}.\end{aligned}\tag{5.2}$$

Thus the tHq (tHW) cross section will be enhanced by a factor of 11.42 (9.44) in the case $\kappa_t = -\kappa_V = -1$. Since κ_V is more precisely determined at the LHC to be SM-like, by convention the negative sign is allowed for κ_t and referred to as the ITC scenario. The ITC scenario is valid up to a cut-off scale $\Lambda \sim 9.3$ TeV without violation of perturbative unitarity [125]. From Eq. (5.2) it is clear that for a fixed value of κ_V the cross section and kinematics of the tH process depends only on the ratio of the coupling modifiers κ_t/κ_V , shown pictorially in Fig. 5.5 along with the $t\bar{t}H$ cross section at 13 TeV.

5.1.2 CP properties of the Higgs

In the SM, the Higgs boson is a pure scalar particle, invariant under CP transformations. It has already been established from Higgs boson decay to vector boson pair that the spin-parity quantum numbers of the Higgs boson are consistent with the SM [136]. However, these measurements only probe Higgs couplings to vector bosons. In case the observed

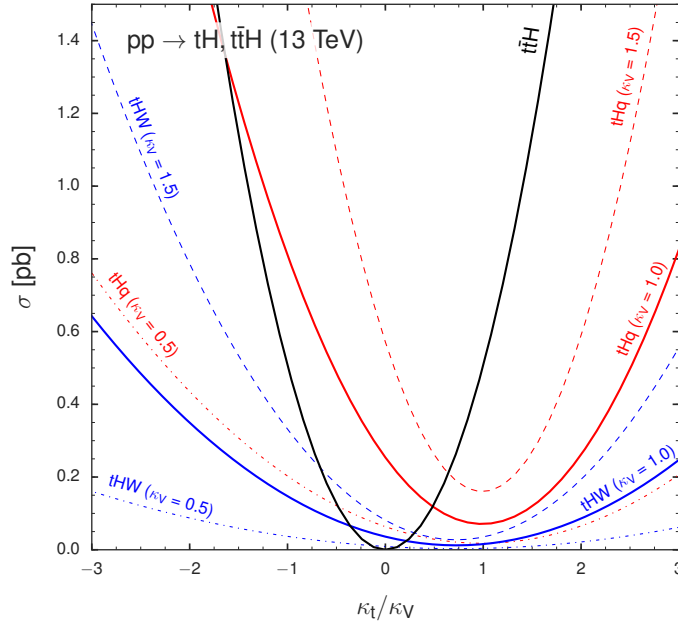


Figure 5.5: Variation of tH and $t\bar{t}H$ inclusive cross sections as a function of κ_t/κ_V [135]. Note $t\bar{t}H$ cross section becomes zero for $\kappa_t=0.0$ but tH cross section is always non-zero.

resonance is an admixture of CP eigenstates where only the CP-even component couples to the vector bosons, CP-violation in Yukawa interactions might not have been detected. The effective Lagrangian below the EWSB scale for a generic spin-0 particle X_0 with CP-even scalar component H and CP-odd pseudoscalar component A can be written as [123]:

$$\mathcal{L}_0^t = -\bar{\psi}_t (\cos \alpha \kappa_{Htt} g_{Htt} + i \sin \alpha \kappa_{Att} g_{Att} \gamma_5) \psi_t X_0 \quad (5.3)$$

where α is the CP-mixing phase between the states H and A ; g_{Htt} and g_{Att} are the coupling strengths equal to $m_t/v = y_t/\sqrt{2}$. Thus for $\alpha = 0^\circ$, X_0 is completely CP-even and for $\alpha = 90^\circ$, completely CP-odd. Any in-between values of α thus correspond to CP-violation in the top-Yukawa sector in this formulation. The cross sections and kinematics of the tH production processes are modified as a function of the mixing angle α . Furthermore, the ITC scenario is identified to be identical to the CP-mixed state with $\alpha = 180^\circ$. Figure 5.6 shows the variation of the tH and $t\bar{t}H$ cross sections as a function of α .

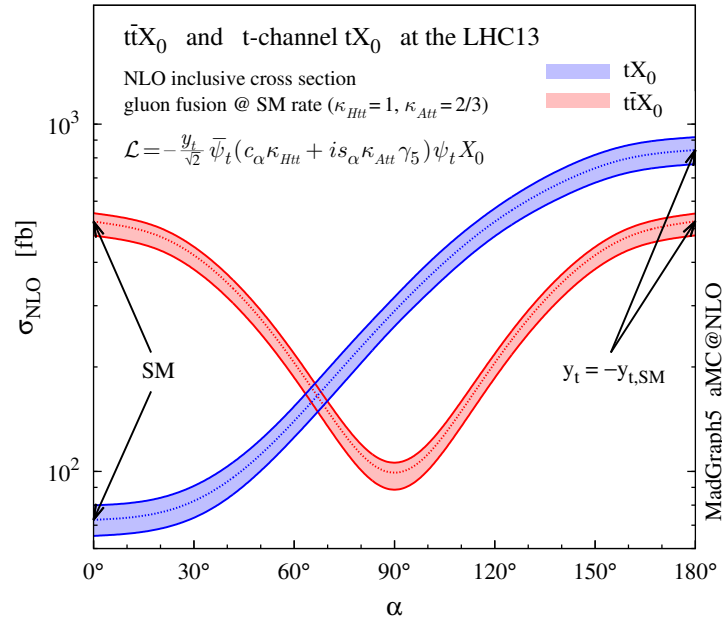


Figure 5.6: Variation of tH and $t\bar{t}H$ inclusive cross sections as a function of α [123].

5.2 Analysis strategy

Leptons, specially electrons and muons, can be identified and measured precisely in the CMS detector. In addition, final states with isolated leptons have much less background than hadronic ones in the LHC environment. The work in this thesis describes the leptonic final state analysis of tH process targeting Higgs boson decays to WW^* to leptons and semi-leptonic decays of the top quark. There are also minor contributions from $H \rightarrow ZZ^*$ and $\tau\tau$ channels, followed by their leptonic decay modes. The tH production at LHC is dominated by tHq process which can be preferentially selected by utilizing the presence of a light quark in the forward region. The analysis is optimized for a search of the negative value of κ_t relative to κ_V (ITC), in the tHq production mode. In addition to leptons, the targetted final state thus also contains one b-jet from $t \rightarrow b\ell\nu$ and a light forward jet.

With the full Run 2 data collected by CMS corresponding to an integrated luminosity of $\sim 137 \text{ fb}^{-1}$, ~ 100 tHq events are expected to be produced in the $H \rightarrow WW^*$, $H \rightarrow ZZ^*$, $H \rightarrow \tau\tau$ and $t \rightarrow b\ell\nu$ final state according to SM. This is to be folded in with the

acceptance of multi-lepton events which is about 10%. Hence only about ~ 10 SM-like tHq events are likely to be selected at the analysis level.

The sensitivity of the multi-lepton analysis is increased by event categorization depending on lepton multiplicity:

- **Three leptons** (3ℓ) where all W s decay leptonically.
- **Two same-sign leptons** ($2\ell_{ss}$) where only one of the W s from the Higgs boson decays leptonically. It is to be noted that the opposite-sign lepton pair final state is challenged by much larger amount of background dominated by $t\bar{t}$ process and hence they are not considered in the analysis.

The above final states are collected in the experiment as a combination of signal and large number of background processes. To increase the signal-to-background ratio in the selected sample, detailed studies about the signal process has been performed with simulated samples using state-of-the-art techniques. Following sections discuss the various steps involved in extracting the information on tH production from data.

5.2.1 Background estimation

The backgrounds to the tH process can be divided into two major categories: irreducible backgrounds containing genuine “prompt” and isolated leptons (i.e. originating from on-shell production and decay of W or Z boson) and reducible backgrounds where at least one of the leptons is “non-prompt” or “fake”. Irreducible backgrounds can be reliably estimated from MC simulations normalized to higher order cross sections, and the distribution can be compared with data in dedicated control regions enriched in these processes. All backgrounds involving only prompt leptons are estimated from MC: $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, WZ , ZZ , $W^\pm W^\pm qq$, $t\bar{t}t\bar{t}$, tZq , tZW , WWW , WWZ , WZZ , ZZZ . The diboson production processes WZ and ZZ are some times referred together as electroweak background in this analysis. Note, WW production resulting in opposite-sign leptons falls under the non-prompt background category. Photon conversions resulting in electron-positron pair production in processes like $t\bar{t}\gamma$, $t\gamma$, $W\gamma$ and $Z\gamma$ are also estimated from MC samples, if the leptons are identified as prompt.

Non-prompt leptons may either be genuine leptons produced in a b-hadron decay, or can originate from a light jet which is mis-identified during reconstruction. The main source of background for this analysis are the non-prompt leptons from the inclusive production of $t\bar{t}$ which has a cross section of ~ 832 pb at NLO at $\sqrt{s} = 13$ TeV, about 10^4 times larger than tH process. Since this process has many jets in the final state, one of them may result in a non-prompt lepton, and due to the large rate significantly contributes to the final selected events. In case of looser selections, Drell-Yan (DY) process can also substantially contribute to the non-prompt background. Moreover, in the two same-sign leptons final state, an additional background arises from the possible mis-measurement of the charge of a final state electron. Electrons have a large probability to emit a hard bremsstrahlung radiation that will subsequently undergo a strongly asymmetric conversion into e^+e^- pair. From this pair if the electron with the opposite charge to the original particle carries most of the initial momentum, the resulting track will have opposite curvature. As a result, the measured charge is not the one of the original electron. These fake and charge-flip backgrounds are of experimental origin, and are not well reproduced by the MC simulation. Data-driven methods are used to estimate their contribution to the selected sample. These background estimation methods for tH multi-lepton analysis are exactly same as in $t\bar{t}H$ multi-lepton analysis [137,138] and have been used in this work unaltered. A brief overview is given here.

Fake rate

The fake rate (FR) is the probability for a non-prompt lepton (passing a lenient *fakeable* criteria) to be selected as a prompt lepton (passing a stringent *tight* criteria). The FR is expressed as:

$$\text{FR} = \frac{N_{\text{tight}}}{N_{\text{loose}}} \quad (5.4)$$

where N_{loose} is the number of lepton candidates that pass the fakeable object selection, and N_{tight} is the number of fakeable leptons that pass the tight selection. The FR is measured in several background-enriched datasets, collected using dedicated triggers, to which the residual prompt lepton contribution is subtracted using the MC simulation. The non-prompt contribution in the signal region is then estimated by applying this fake

rate to events from an orthogonal side-band region with a looser lepton selection. Events from the side-band region are weighted according to the probability of the leptons to pass the tight criteria. If N_{00} , N_{01} or N_{10} , N_{11} are the actual number of events in the $2\ell ss$ channel when zero, one, two prompt leptons are present, and N_{ff} , N_{fp} or N_{pf} , N_{pp} are the measured number of events where two, one, zero leptons fail the tight selection criteria, then:

$$\begin{pmatrix} N_{pp} \\ N_{pf} \\ N_{fp} \\ N_{ff} \end{pmatrix} = \begin{pmatrix} p_1 p_2 & p_1 f_2 & f_1 p_2 & f_1 f_2 \\ 0 & p_1(1-f_2) & 0 & f_1(1-f_2) \\ 0 & 0 & (1-f_1)p_2 & (1-f_1)f_2 \\ 0 & 0 & 0 & (1-f_1)(1-f_2) \end{pmatrix} \begin{pmatrix} N_{11} \\ N_{10} \\ N_{01} \\ N_{00} \end{pmatrix} \quad (5.5)$$

where p_i is the prompt rate (probability of a prompt lepton to pass the tight requirement, measured using MC truth), and f_i is the fake rate for the i^{th} lepton. Here it is assumed that rate of prompt leptons failing the tight requirement is negligible compared to the contribution from fakes. Simplifying Eq.(5.5):

$$N_{00} = \frac{N_{ff}}{(1-f_1)(1-f_2)}, \quad N_{01} = \frac{N_{fp} - \frac{f_2}{1-f_2} N_{ff}}{p_2(1-f_1)}, \quad N_{10} = \frac{N_{pf} - \frac{f_1}{1-f_1} N_{ff}}{p_1(1-f_2)}, \quad (5.6)$$

the background contribution due to fakes can be obtained from the measured number of events in the side-band region:

$$\begin{aligned} N_{pp}^{bkg} &= p_1 f_2 N_{10} + f_1 p_2 N_{01} + f_1 f_2 N_{00} \\ &= \frac{f_1}{1-f_1} N_{fp} + \frac{f_2}{1-f_2} N_{pf} - \frac{f_1 f_2}{(1-f_1)(1-f_2)} N_{ff}. \end{aligned} \quad (5.7)$$

In the same way, for the 3ℓ channel, the non-prompt contribution is obtained using the expression:

$$\begin{aligned} N_{ppp}^{bkg} &= \frac{f_1}{1-f_1} N_{fpp} + \frac{f_2}{1-f_2} N_{pfp} + \frac{f_3}{1-f_3} N_{ppf} \\ &\quad - \frac{f_1 f_2}{(1-f_1)(1-f_2)} N_{ffp} - \frac{f_2 f_3}{(1-f_2)(1-f_3)} N_{pff} - \frac{f_3 f_1}{(1-f_3)(1-f_1)} N_{fpf} \\ &\quad + \frac{f_1 f_2 f_3}{(1-f_1)(1-f_2)(1-f_3)} N_{fff}. \end{aligned} \quad (5.8)$$

Charge-flip rate

Processes that yield opposite-sign di-lepton pairs ($t\bar{t}$, DY) may have occasionally one of the two leptons, specially electrons, with a wrong charge assigned if the lepton momentum is reasonably high. The background arising from such charge mis-assignment is estimated from data using a sample where there are two opposite-sign electrons from Z boson decays, which get reconstructed as having the same charge, but whose invariant mass peaks at the Z boson mass. The contribution to the $2\ell ss$ signal regions is then estimated by extrapolating from an equivalent opposite-sign di-lepton ($2\ell os$) selection. For each $2\ell os$ event, the lepton p_T and η are used to determine the event's mis-identification probability. This probability is then used in a way similar to the fake-rate, to determine the charge-flip background to the $2\ell ss$ final state from a side-band region.

5.2.2 Multivariate analysis

In high energy physics, quite often interesting signal rate is much lower than background processes. Use of selection and counting method may not yield a reasonably pure sample since the correlations between signal characteristics are not always visible. Multivariate statistical analysis is generally employed to interpret large ensemble of observations, exploiting the correlations between the observables. A set of measurements on a final state can be thought of a multivariate population or distribution. Each measurement can then be treated as random with a probability describing the population. *Machine learning* [139] algorithms or classifiers can be trained on the ensemble to predict the probability of a given observation. In this thesis the multivariate analysis (MVA) tools are provided by the TMVA [140] package within the ROOT framework [141]. TMVA implements a number of classifiers, each with different strategy to separate one observation from the other based on discriminating variables given as input. The boosted decision trees (BDT) [142,143] classifier has been used in this thesis to separate signal events from background. A decision tree splits the phase space of information using binary outcome (yes/no) of many requirements until a stop criterion defined a-priori, is fulfilled. The decision can be boosted using several trees, derived on the same ensemble. The final

probability outcome is a weighted average of individual decisions. The boosting used in this context is GradientBoost which uses binomial log-likelihood loss function [144] for classification. A representative binary tree structure and multivariate separation are shown in Fig. 5.7.

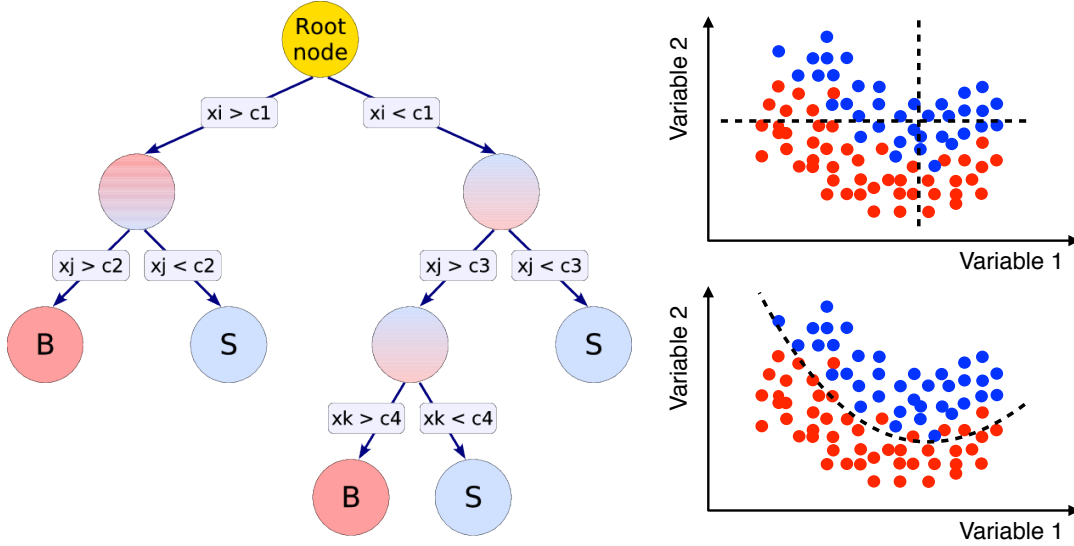


Figure 5.7: Schematic view of a decision tree [140] (left) and multivariate classification (right). x_i are discriminating variables used for binary splitting the phase-space. Each split uses the variables that gives the best separation between signal (S) and background (B). The same variable may be used multiple times while others may not be used at all. The right figure shows simple classification of events with two variables having correlation. The events are not very separable via straight cuts, however a higher dimensional variable can achieve the task.

In this analysis the tHq signal is discriminated against two major types of background: prompt and non-prompt. Use of different trainings allows for higher sensitivity to the underlying physics. For example, in case of prompt backgrounds, the training is more sensitive to the jet kinematics in the event whereas, for non-prompt backgrounds, it is more sensitive to lepton kinematics.

5.2.3 Statistical interpretation

Most of the data analyses in particle physics involves comparing observed distributions to predictions from theory. This is done using the probabilistic approach: whether the observed data is in agreement with the predicted probability of a given *hypothesis*. If data consists of n measured values of a variable $\vec{x} = (x_1, x_2, \dots, x_n)$ and there are a set of hypotheses H_0, H_1 etc., each hypothesis is described by a probability density function (p.d.f.), $f(\vec{x}|H_0), f(\vec{x}|H_1)$ etc. following Poisson statistics [145].

Parameter estimation

In general, if x is distributed according to a p.d.f. $f(x; a)$ where a is an unknown parameter of the theory, the likelihood function for a is expressed as:

$$\mathcal{L}(a) = \prod_{i=1}^n f(x_i; a) \quad (5.9)$$

corresponding to n observations of x , which is also the joint p.d.f. of the whole data sample. If the hypothesized p.d.f. and parameter values are correct, a high probability for the data is expected. The maximum likelihood estimators (MLEs) for the parameters are those that maximize the likelihood function: $\hat{a} = \underset{a}{\operatorname{argmin}}[\mathcal{L}(x|a)]$. The value of a that maximizes $\mathcal{L}(a)$ also maximizes its logarithm:

$$\left. \frac{d \ln \mathcal{L}(a)}{da} \right|_{a=\hat{a}} = 0. \quad (5.10)$$

The MLE has a bias but it decreases proportionally to the size of the data or n . For large data the negative log-likelihood (NLL) curve i.e., distribution of $\ln \mathcal{L}(a)$ vs. a is a parabola and identical to a χ^2 distribution. At the minimum:

$$-\ln \mathcal{L}(a) = -\ln \mathcal{L}(\hat{a}) - \frac{1}{2}(a - \hat{a})^2 \frac{d^2 \ln \mathcal{L}(a)}{d^2 a}. \quad (5.11)$$

Here $-1/\left\langle \frac{d^2 \ln \mathcal{L}(a)}{d^2 a} \right\rangle$ gives the variance of the MLE $\hat{a}$. Hence the error or standard deviation on the MLE is given by:

$$\sigma(\hat{a}) = \sqrt{-\frac{1}{\frac{d^2 \ln \mathcal{L}(a)}{d^2 a}}}. \quad (5.12)$$

Thus $-2\Delta \ln \mathcal{L}(a) = -2[\ln \mathcal{L}(a) - \ln \mathcal{L}(\hat{a})] = 1$ gives 1σ uncertainty on $\hat{a}$, $-2\Delta \ln \mathcal{L}(a) = 4$ gives 2σ uncertainty, and so on. In terms of χ^2 , with one parameter of interest, $-2\Delta \ln \mathcal{L}(a) = 1$ corresponds to 65% CL, $-2\Delta \ln \mathcal{L}(a) = 3.84$ corresponds to 95% CL etc.

Hypothesis testing

In case of hypothesis testing, i.e., to determine which theory better describes the observation, a *test statistic* is constructed following the Neyman-Pearson lemma [146]:

$$q_\mu = -2 \ln \frac{\mathcal{L}(n|\mu s + b)}{\mathcal{L}(n|b)}, \mu \geq 0 \quad (5.13)$$

where n is the observed number of events in data, μ is the signal strength modifier, s and b are number of signal and background events respectively. μ is defined as the ratio of observed and expected cross sections and regarded as a free parameter for the maximum likelihood fit to the data. The compatibility of the data with any hypothesis is determined by the *p-values*:

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|\mu) dq_\mu \quad (5.14)$$

$$1 - p_b = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|0) dq_\mu$$

where p_μ gives the probability of accepting the signal+background hypothesis ($\mu \neq 0$) and p_b gives the probability of accepting the background only ($\mu = 0$) hypothesis. If a decision boundary for q_μ value is chosen, there is a finite probability of p_μ to reject $\mu = 0$ hypothesis even if it is true (Type I error or false positive) or accepting a false $\mu = 0$ hypothesis with a probability p_b (Type II error or false negative). A simple illustration of the hypothesis test is shown in Fig. 5.8.

Asymptotic limits

At the LHC, signal rates are very small and several systematic uncertainties related to both theoretical prediction and experimental measurements exist, which can affect the observation. Current LHC experiments have agreed [147] to use the profile likelihood

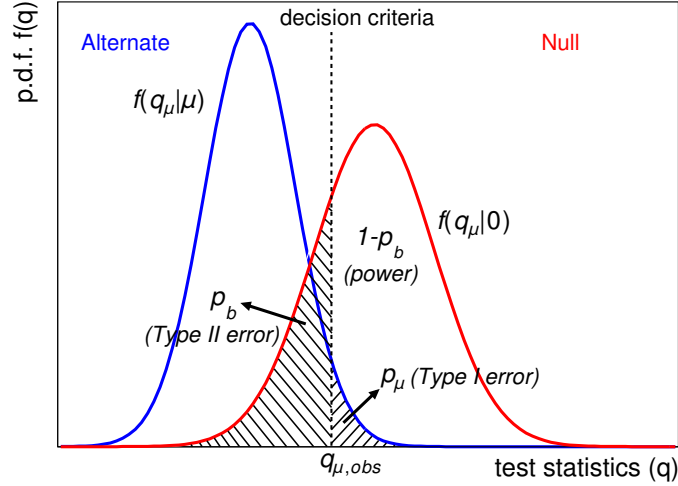


Figure 5.8: Illustration of hypothesis testing. The figure shows test statistics distributions for the *Null* (background only) and the *Alternate* (signal+background) hypotheses. The decision criteria is set by the observed value $q_{\mu,obs}$, which determines the *power* of the test as $(1-p_b)$ if the alternate hypothesis is rejected.

test statistic [148]:

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(n|\mu, \hat{\theta}_\mu)}{\mathcal{L}(n|\hat{\mu}, \hat{\theta})}, \quad 0 \leq \hat{\mu} \leq \mu \quad (5.15)$$

and CL_s estimate to set more conservative limits:

$$CL_s(\mu) = \frac{CL_{s+b}}{CL_b} = \frac{p_\mu}{1-p_b}. \quad (5.16)$$

θ are systematic uncertainties of the analysis or *nuisances* incorporated in the likelihood function as multiplicative error p.d.f.s. Profiling refers to the determination of nuisance parameter values $\hat{\theta}_\mu$ that maximize the likelihood for a particular μ . The parameters $\hat{\mu}$ and $\hat{\theta}$ give the global maximum of the likelihood. The profile likelihood test statistics is one-sided, which does not allow upward fluctuation of the signal or downward fluctuation of background to be rejected. Asymptotically the p.d.f. distributions of $\tilde{q}_\mu$ can be exactly approximated using mathematical theorems, simplifying the calculation of CL_s . $f(\tilde{q}_\mu|0)$ in the asymptotic limit is used to derive median expected limits and confidence intervals using the Asimov representative dataset. This ensures toy MC generation is not required to determine the statistical probability over a large ensemble. If for $\mu = 1$,

$CL_s \leq p$, the signal+background hypothesis is excluded with $(1 - p)$ CL_s confidence level. 95% CL upper limit on μ is quoted by adjusting μ until $CL_s = 0.05$ is reached. p is converted to an equivalent significance z , defined such that z standard deviation upward fluctuation of a Gaussian random variable has an upper tail area equal to p . Thus $p = 0.05$ corresponds to a significance of 1.96σ , whereas a 5σ significance requires $p = 5.7 \times 10^{-7}$.

5.2.4 Analysis workflow

Using the above mentioned techniques detailed analysis has been performed using the multi-lepton final state of tH process, presented in Chapter 6. Figure 5.9 pictorially depicts the steps followed in sequence.

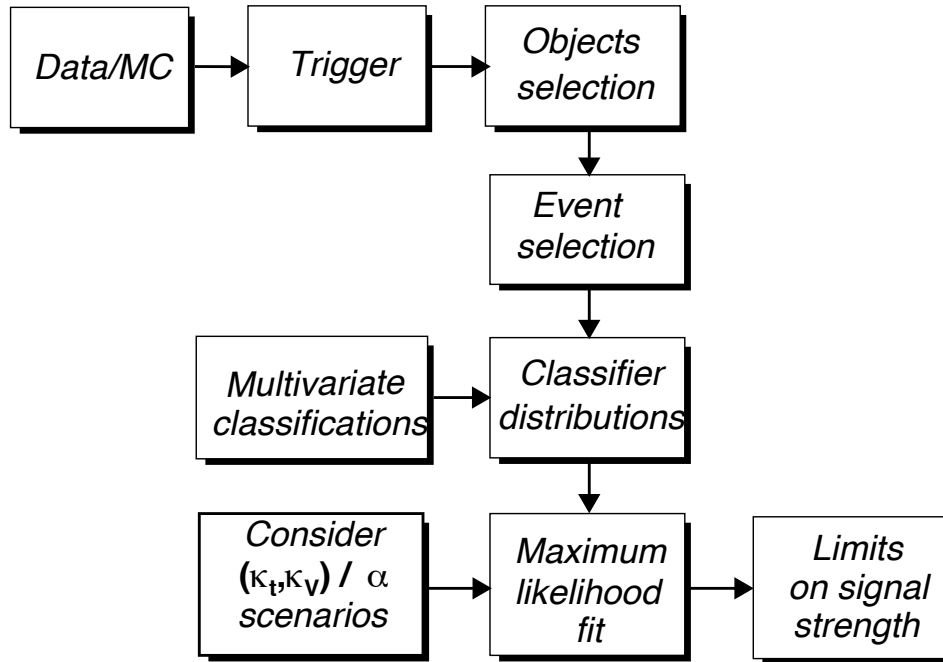


Figure 5.9: Illustration of analysis workflow. Detailed descriptions are presented in Chapter 6.

Chapter 6

tH multi-lepton at $\sqrt{s} = 13$ TeV

From the LHC Run 1 data a 95% CL upper limit on tHq production in the ITC scenario was extracted for the first time by the CMS experiment, considering Higgs decays to $b\bar{b}$, WW^* , $\tau\tau$ and $\gamma\gamma$. The observed (expected) upper limit on the $\sigma \times BR$ was set between 0.60 (0.45) pb and 1.00 (0.70) pb depending on the assumed di-photon branching fraction [69] using $L \sim 20 \text{ fb}^{-1}$. With increasing amount of data volume available in Run 2, additional values of (κ_t, κ_V) for the Higgs boson has been probed following the Higgs characterization model [123, 131]. In this chapter the study of the tH production process in the multi-lepton final states with CMS data collected during 2016 and 2017 amounting to $L \sim 77 \text{ fb}^{-1}$ has been described. This data volume is too small to observe the SM process but may be sufficient for BSM production of tH . Hence by measuring the cross section, the constraints on κ_t and α (see Chapter 5) have been obtained. It is to be noted that ATLAS experiment is yet to study the tH production mode with data from either run periods.

An important aspect of this analysis is that for determining the parameters of interest, $t\bar{t}H$ has been considered as a signal process since it involves the top-Higgs coupling as well. Hence the $t\bar{t}H$ cross section and kinematics vary as a function of κ_t and α . In addition, the multi-lepton final states of $t\bar{t}H$ and tH are largely overlapping; with evidence of $t\bar{t}H$ events in Run 2 data, additional sensitivity is provided to the tH final state analysis. In general, simulated events are generated with an arbitrary luminosity value and the yield for a particular process is obtained normalized to the measured luminosity L as

$N = \sigma L \epsilon A$, where ϵ are measurement efficiencies determined by the analysis and A are acceptance due to the detector constraints. The expected background yields are compared to data to determine the cross section of the signal process:

$$\sigma \times BR(\text{obs}) = \frac{N^{data} - N^{bkg}}{L \epsilon A} \quad (6.1)$$

and expressed in terms of signal strength μ using the maximum likelihood fit. Many times the efficiency of a particular selection are different in case of data and MC, in those cases scale factors are derived to correct the MC yield to match data and differences are accounted for using systematic uncertainties.

6.1 Analysis using 2016 data

6.1.1 Data and simulated samples

The data considered in this analysis were collected by the CMS experiment during 2016 and correspond to a total integrated luminosity of 35.9 fb^{-1} , when the CMS magnet was on. The 23Sep2016 re-reco (run periods B to G) and prompt (run period H) versions of the following categories of primary datasets have been used:

- DoubleMuon, SingleMuon, MuonEG, SingleElectron, DoubleEG.

The nomenclature of datasets refer to the presence of trigger objects in corresponding events, e.g., DoubleMuon dataset contains events that are triggered by two muons etc. The JSON file used to select the data events for which all the sub-detectors were fully operational are:

- Run B-G: Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt
- Run H: Cert_271036-284044_13TeV_PromptReco_Collisions16_JSON.txt.

The MC samples used in this analysis are produced with CMSSW version 80X, using the CUETP8M1 tune [115] for underlying events in PYTHIA. The tH signal samples were produced with MADGRAPH (MG5_aMC@NLO version 5.222), in leading-order mode using

PDF scheme of NNPDF, MLM matching, and are normalized to NLO cross sections, see Table 6.1. Each sample is generated with a set of event weights corresponding to different values of κ_t and κ_V , implemented using event-by-event parton level weights or *LHE weights*. The $t\bar{t}H$ sample has been generated at NLO, with Higgs decays to $b\bar{b}$ turned off to save computing resources and top quark decays using MadSpin [149, 150] to take into account spin-correlation of the decay products.

Table 6.1: tH and $t\bar{t}H$ samples with their cross section times branching fraction for semi-leptonic top quark decay (for tHq only) in the ITC scenario used in 2016 analysis. See Ref. [151] for more details.

Sample	$\sigma \times BR$ (pb)
THQ_Hinc1_13TeV-madgraph-pythia8_TuneCUETP8M1	0.2568
THW_Hinc1_13TeV-madgraph-pythia8_TuneCUETP8M1	0.1472
tHJetToNonbb_M125_13TeV_amcatnloFFX_madspin-pythia8_mWCutfix	0.2151

For discriminating with the signal, high statistics LO samples of inclusive $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$ processes listed in Table 6.2 are used in the training of BDT discriminators. For these samples no criteria has been applied on jets at generator level.

Table 6.2: Leading-order $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$ samples used in the BDT training in 2016 analysis. $t\bar{t}$ events where the top or anti-top or both decay semi-leptonically are available as different simulation samples, used together in the training. Also mentioned are the NLO cross sections of the samples.

Sample	$\sigma \times BR$ (pb)
t $\bar{t}W$ Jets_13TeV_madgraphMLM	0.6105
t $\bar{t}Z$ Jets_13TeV_madgraphMLM	0.7655
TTJets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.1754
TTJets_SingleLeptFromTbar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.1754
TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	87.3148

Different MC generators were used to generate the background processes as shown in Table 6.3. The dominant sources ($t\bar{t}W$, $t\bar{t}Z$) were produced using MG5_aMC@NLO inter-

faced to PYTHIA, and are scaled to NLO cross sections. Other processes are simulated using POWHEG interfaced to PYTHIA, or only PYTHIA as in the case of same-sign W production via double parton scattering. Data driven estimates have been used for non-prompt backgrounds such as inclusive $t\bar{t}$ and DY processes etc. as discussed in Chapter 5. The data and MC samples are both reconstructed using CMSSW version 8_0_25 and analyzed with the same.

6.1.2 Triggers

Since a collision event can pass multiple HLT paths or conditions, it may be recorded in more than one primary datasets. For example, an event with two muons may be included in either one or all of the datasets DoubleMuon, SingleMuon and MuonEG. Hence a set of primary HLT requirements are applied only to collision data events (not MC samples), in order to remove any such overlap. The list of triggers applied for exclusive event selection is given in Table 6.4. Events are chosen from each dataset passing any one of the corresponding trigger conditions, but failing all the triggers of other datasets ensuring exclusivity. This procedure of combining multiple datasets also maximizes the number of selected events.

In the next step, online-reconstructed events triggered by one, two, or three leptons are considered in the final state. Single-lepton triggers are included to boost the acceptance of signal events where the p_T of the sub-leading lepton falls below the threshold of the double-lepton triggers. Additionally, by including double-lepton triggers in the three lepton category, as well as single-lepton triggers, the trigger efficiency is enhanced, considering the logical “or” of the trigger decisions of all the individual triggers in a given category. Table 6.5 shows the lowest-threshold un-prescaled triggers present in the HLT menu for event selection in both MC and data in 2016. Here Mu and Ele refer to muon and electron objects and numbers next to the object name specify the p_T thresholds required by the particular trigger. “TrackId”, “CaloId” refer to particle identification based on tracker hits or calorimeter deposits respectively. In addition, there may be isolation required around the object or the track associated with the object, specified by the “Iso” or “TkIso” tags. “L”, “VL” and “VVL” stand for loose, very loose and very very loose type

Table 6.3: List of background MC samples used in 2016 analysis to extract the final yields and distributions. The cross sections mentioned are at NLO and NNLO accuracy, as available. In these samples prompt leptons produced from the decays of W , Z bosons are denoted using the notation L . Two $t\bar{t}Z$ samples used here correspond to different ranges of invariant mass of the di-lepton pair from Z/γ^* : 1-10 GeV and > 10 GeV. For conversion backgrounds $t\bar{t}\gamma$, $t\gamma$, $W\gamma$ and $Z\gamma$, the prompt photon decay to e^+e^- pair contributes to leptons in the final state. Production processes not observed in CMS are generally denoted as rares, the decays of which to leptonic final states have very low cross sections.

Process	Sample	$\sigma \times BR$ (pb)
$t\bar{t}W$	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
$t\bar{t}Z$	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2529
	TTLL_m1to10_LO_NoMS_for76X	0.0283
WZ	WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	5.063
ZZ	ZZTo4L_13TeV-powheg-pythia8	1.256
Conversion	TTGJets_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	3.697
	TGJets_TuneCUETP8M1_13TeV-amcatnlo-madspin-pythia8	2.967
	WGToLNuG_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	585.5
	ZGTo2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	131.3
Rares	WpWpJJ_EWK-QCD_TuneCUETP8M1_13TeV-madgraph-pythia8	0.03711
	WW_DoubleScattering_13TeV-pythia8	1.64
	tZq_l1_4f_13TeV-amcatnlo-pythia8_TuneCUETP8M1	0.0758
	ST_tWl1_5f_LO_13TeV-MadGraph-pythia8	0.01123
	ST_tWl1_5f_LO_13TeV-MadGraph-pythia8	0.01123
	TTTT_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.009103
	WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2086
	WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.1651
	WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.05565
	ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.01398

of the requirements. The v^* at the end of the trigger paths refer to different versions of the same trigger throughout the full data taking period, which depends on detector condition and calibration. The efficiencies of the combined trigger decision are determined to be ~ 1 using an unbiased dataset recorded using p_T^{miss} trigger for $t\bar{t}H$ multi-lepton

Table 6.4: Table of high-level triggers used in 2016 analysis for the overlap removal between datasets.

Dataset	Trigger
DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v*, HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v*, HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*, HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*, HLT_Mu23_TrkIsoVVL_Mu8_TrkIsoVVL_v*, HLT_Mu23_TrkIsoVVL_TkMu8_TrkIsoVVL_v*, HLT_Mu23_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*, HLT_Mu23_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*, HLT_Mu17_Mu8_SameSign_v*, HLT_Mu17_Mu8_SameSign_DZ_v*, HLT_Mu20_Mu10_SameSign_v*, HLT_Mu20_Mu10_SameSign_DZ_v*, HLT_DoubleMu8_Mass8_PFHT300_v*, HLT_TripleMu_12_10_5_v*, HLT_TripleMu_5_3_3_v*, HLT_TrkMu15_DoubleTrkMu5NoFiltersNoVtx_v*
DoubleEG	HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_DoubleEle8_CaloIdM_TrackIdM_Mass8_PFHT300_v*, HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL_v*,
MuonEG	HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*, HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL_v*, HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*, HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_v*, HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*, HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Mu8_Ele8_CaloIdM_TrackIdM_Mass8_PFHT300_v*, HLT_DiMu9_Ele9_CaloIdL_TrackIdL_v*, HLT_Mu8_DiEle12_CaloIdL_TrackIdL_v*
SingleMuon	HLT_IsoMu20_v*, HLT_IsoTkMu20_v*, HLT_IsoMu22_v*, HLT_IsoTkMu22_v*, HLT_IsoMu24_v*, HLT_IsoTkMu24_v*, HLT_Mu45_eta2p1_v*, HLT_Mu50_v*, HLT_TkMu50_v*
SingleElectron	HLT_Ele25_WPTight_Gsf_v*, HLT_Ele25_eta2p1_WPLoose_Gsf_v*, HLT_Ele25_eta2p1_WPTight_Gsf_v*, HLT_Ele27_WPTight_Gsf_v*, HLT_Ele27_eta2p1_WPLoose_Gsf_v*, HLT_Ele45_WPLoose_Gsf_v*

analysis [137], the data/MC scale factor values are shown in Table 6.6.

Table 6.5: List of HLT paths considered in the 2016 analysis for event selection.

Object	Trigger
one electron	HLT_Ele27_WPTight_Gsf_v*, HLT_Ele25_eta2p1_WPTight_Gsf_v*
one muon	HLT_IsoMu24_v*, HLT_IsoTkMu24_v*, HLT_IsoMu22_eta2p1_v*, HLT_IsoTkMu22_eta2p1_v*,
two electrons	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
two muons	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v*, HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v*, HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*, HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*, HLT_Mu23_TrkIsoVVL_Mu8_TrkIsoVVL_v*, HLT_Mu23_TrkIsoVVL_TkMu8_TrkIsoVVL_v*, HLT_Mu23_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*, HLT_Mu23_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
one electron, one muon	HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_v*, HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*, HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*,
three electrons	HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL_v*
two electrons, one muon	HLT_Mu8_DiEle12_CaloIdL_TrackIdL_v*
one electron, two muons	HLT_DiMu9_Ele9_CaloIdL_TrackIdL_v*
three muons	HLT_TripleMu_12_10_5_v*, HLT_TripleMu_5_3_3_v*

6.1.3 Identification of leptons

This analysis uses two or three reconstructed leptons and jets, one of which is b-tagged. The selections of these objects are mostly based on recommendation by corresponding physics object group (POG). First, electrons and muons are pre-selected with loose or lenient criteria. In a second step, the quality of these leptons are evaluated using the “lepton MVA” developed for the $t\bar{t}H$ multi-lepton analysis [138] to be either fakeable or tight (capable of satisfying stringent criteria).

Table 6.6: Trigger efficiency scale factors and associated uncertainties, shown here rounded to the nearest percent for the 2016 analysis. They are determined using the ratio of events that pass any of the required trigger criteria to the total number of events in the p_T^{miss} dataset, after applying the selection of two or three leptons in the final state.

Category	Scale Factor
3ℓ	1.00 ± 0.03
$e\mu$	1.01 ± 0.01
$\mu\mu$	1.00 ± 0.01

Muon selection

The muon objects considered in this analysis are reconstructed if they are within a geometrical acceptance of $|\eta| < 2.4$ and have transverse momenta $p_T > 5$ GeV. The pre-selected muons must pass the loose muon selection, which considers all PF muons without additional quality criteria. The PF muons are selected among the reconstructed muon candidates by requiring matching track components in the muon system and tracker. A requirement on the distance of the muon track from the leading primary vertex serves to reject badly reconstructed tracks, or tracks originating from pileup vertices. This is achieved by requiring the significance of the impact parameter (SIP_{3D}) to be lower than 8, and the distance in the transverse plane (d_{xy}) and along z axis (d_z) to be lower than 0.05 and 0.1 cm respectively.

Electron selection

The electron objects used in this analysis are reconstructed if they are within a geometrical acceptance of $|\eta| < 2.5$ and have transverse momenta $p_T > 7$ GeV. The pre-selected electrons are also required to pass the basic quality criteria for reconstructed tracks. Similarly to muons, they must pass the following requirements : $SIP_{3D} < 8$, $d_{xy} < 0.05$ cm and $d_z < 0.1$ cm. In addition to this, the electron identification relies on a BDT discriminator which was trained on all electrons, regardless of the trigger. The BDT variable considers observables based on purely calorimetric or tracker measurements and their

comparisons.

Lepton isolation

Leptons from W and Z boson decays are typically isolated, i.e., away from other reasonably energetic particles unlike a lepton from b -hadron decay. Thus the sum of p_T of particles in a cone around a prompt lepton in η - ϕ space is expected to be very less compared to the p_T of the lepton candidate, $p_T(\ell)$:

$$I = \sum_{i, i \neq \ell} p_T(i)/p_T(\ell) < 1. \quad (6.2)$$

In the present analysis, the lepton isolation is quantified by taking into account the potential radiations from the candidate lepton, lying within a cone of decreasing radius with increasing p_T of the lepton, defined as:

$$R = \frac{10}{\min(\max(p_T(\ell), 50), 200)}. \quad (6.3)$$

Here, unit of $p_T(\ell)$ is GeV. Therefore the maximum radius of the isolation cone is 0.2, whereas the minimum is 0.05. This technique significantly improves the identification efficiency for isolated leptons in events having high H_T , as it considers local isolation even in boosted topologies.

The impact of pileup on isolation is mitigated using effective area corrections. If $\sum_R p_T(h^\pm)$, $\sum_R p_T(h^0)$ and $\sum_R p_T(\gamma)$ refer to the sum of transverse momenta of charged hadrons, neutral hadrons and photons respectively, within the cone of radius R , then

$$I_{mini} = \frac{\sum_R p_T(h^\pm) - \max(0, \sum_R p_T(h^0) + \sum_R p_T(\gamma) - \rho \mathcal{A}(\frac{R}{0.3})^2)}{p_T(\ell)} \quad (6.4)$$

where ρ is the pileup energy density of neutral particles, and $\mathcal{A}$ is the effective area (whose values are given in Table 6.7). Both pre-selected electrons and muons are required to have I_{mini} value less than 0.4.

Jet related variables

Since the most important source of lepton background in this analysis is the mis-identification of jets, several variables are constructed based on the properties of the jet that lies closest

Table 6.7: Effective areas for electrons and muons used in 2016 analysis. Similar to photon objects, effective area for leptons is the ratio of slopes obtained by linearly fitting isolation and ρ as a function of number of vertices. It varies with the η range due to increasing pileup activity in the forward regions.

$ \eta $ range	$\mathcal{A}(\mu)$	$\mathcal{A}(e)$
0.0-0.8	0.0735	0.1752
0.8-1.0	0.0619	0.1752
1.0-1.3	0.0619	0.1862
1.3-2.0	0.0465	0.1411
2.0-2.2	0.0433	0.1534
2.2-2.3	0.0577	0.1903
2.3-2.4	0.0577	0.2243
2.4-2.5	0.0577	0.2687

to the candidate lepton. In particular, PF jets reconstructed within $\Delta R < 0.5$ around the lepton are considered. The four variables constructed from the jet informations are as follows.

- p_T^{ratio} : the ratio between the p_T of the lepton and the p_T of the jet
- p_T^{rel} : the lepton p_T w.r.t. the axis of jet p_T subtracting the lepton

$$p_T^{\text{rel}} = \frac{[\vec{p}_T(\text{jet}) - \vec{p}_T(\ell)] \cdot \vec{p}_T(\ell)}{\|\vec{p}_T(\text{jet}) - \vec{p}_T(\ell)\|}$$

- jet CSV : the CSVv2 b-tagging discriminator value of the jet
- the number N_{charged} of charged particles within the jet

Lepton MVA discriminator

To reduce the contamination by non-prompt leptons, a multivariate discriminator has been developed for the purpose of $t\bar{t}H$ multi-lepton analysis, that optimally selects the

signal leptons while rejecting most of the fake ones. This variable is referred to as the lepton MVA and is based on BDT techniques. Below listed are the input variables considered for this discriminator:

1. lepton p_T
2. lepton η
3. SIP_{3D}
4. d_{xy}
5. d_z
6. Charged component of I_{mini} or $\sum_R p_T(h^\pm)$ contribution
7. Neutral component of I_{mini} or $\sum_R p_T(h^0) + \sum_R p_T(\gamma)$ contribution
8. jet CSV
9. number of charged tracks in the jet
10. p_T^{rel}
11. p_T^{ratio}
12. For muons: compatibility of track segments in various muon stations;
For electrons: electron MVA ID discriminator value

For both muons and electrons, the working point of the lepton MVA is chosen at 0.90 for this analysis with efficiency greater than 80% (85%) to select prompt electrons (muons) and greater than 98% fake rejection. Further details in training and performances are documented in Refs. [152, 153].

Additional requirements

In the di-lepton final state additional requirements on the quality of the charge assignment are applied to suppress opposite-sign events in which the charge of one of the leptons is mis-measured. For the electrons, consistency between the different measurements of the charge using ECAL supercluster and the tracker hits is required, while for the muons the track transverse momentum has to be well measured ($\Delta p_T/p_T < 0.2$). These cuts are referred to as “tight charge”. Moreover in order to suppress background electrons from photon conversions as much as possible, electrons with missing hits in the

innermost layer or associated with a successfully reconstructed conversion vertex are rejected [154].

6.1.4 Lepton selection

Three different sets of selection are defined both for the electron and muon identification: *loose*, *fakeable object* and *tight* with increasing stringency to select good quality leptons. As described in more detail later, these are used for event level vetoes, the fake background estimation, and the final signal selection respectively. For the fakeable objects the p_T is defined as $0.90 \times p_T(\text{jet})$, where the jet is the one associated with the lepton. This mitigates the dependence of the fake rate on the momentum of the fakeable object and thereby improves the precision of the method. Tables [6.8] and [6.9] list the full criteria for the different selections of muons and electrons.

Lepton selection efficiency

Efficiencies of reconstruction and selecting loose leptons are measured both for muons and electrons using a tag-and-probe method on both data and MC, using DY or $Z \rightarrow \ell^+ \ell^-$ events. Corresponding scale factors are derived from the ratio of efficiencies between data and MC and applied to the selected events in MC, following the $t\bar{t}H$ multi-lepton analysis. The loose selection efficiency scale factor is defined in the following way:

$$\rho(p_T, \eta) = \frac{\epsilon_{data}(p_T, \eta)}{\epsilon_{MC}(p_T, \eta)} \quad (6.5)$$

where $\epsilon(p_T, \eta)$ is the efficiency measured for a given lepton in data or MC as a function p_T, η of the lepton. Leptons with very low or very high p_T are not well measured due to insufficient hits in the detector, small bending angle, high probability of bremsstrahlung for electrons etc. The efficiencies of applying the tight selection criteria on the loose leptons are determined again by using a tag-and-probe method on a sample of DY-enriched events. The scale factors are used to correct the weight of the simulated events. The full simulation corrections are the product of all scale factors:

$$\rho = \prod_{j \in \text{leptons}} \prod_{i \in \text{efficiencies}} \rho_i(p_{Tj}, \eta_j). \quad (6.6)$$

Table 6.8: Requirements on each of the three muon selections in 2016 analysis. In the cases where the cut values change between the selections, those values are listed in the table. Otherwise, whether the cut is applied is indicated. For the two p_T^{ratio} and CSV rows, the cuts marked with a † are applied to leptons that fail the lepton MVA cut, while the loose cut value is applied to those that pass.

Cut	Loose	Fakeable object	Tight
$ \eta < 2.4$	✓	✓	✓
p_T	> 5 GeV	> 15 GeV	> 15 GeV
$ d_{xy} < 0.05$ (cm)	✓	✓	✓
$ d_z < 0.1$ (cm)	✓	✓	✓
$\text{SIP}_{3D} < 8$	✓	✓	✓
$I_{\text{mini}} < 0.4$	✓	✓	✓
is Loose Muon	✓	✓	✓
p_T^{ratio}	–	$> 0.5^\dagger / -$	–
jet CSV	–	$< 0.3^\dagger / < 0.8484$	< 0.8484
segment compatibility	–	$> 0.3^\dagger / -$	–
is Medium muon	–	–	✓
tight-charge	–	–	✓
lepton MVA > 0.90	–	–	✓

They are documented for the $t\bar{t}H$ multi-lepton analysis in Ref. [153] and are exactly equivalent for this analysis.

6.1.5 Taus

Hadronically decaying taus (τ_h) are reconstructed using the hadron-plus-strips (HPS) algorithm [155]. Typically the τ_h decays to either one or three charged pions or kaons and up to two neutral pions and tau neutrino. The charged particle tracks along with their impact parameters and energy deposits in the ECAL are utilized by the HPS algorithm to construct the “decay mode finding” discriminator, to identify τ_h candidates in 1- or 3-prong decay modes with or without additional π^0 's. The different τ_h decays with their

Table 6.9: Criteria for each of the three electron selections in 2016 analysis. In cases where the cut values change between selections, those values are listed in the table. Otherwise, whether the cut is applied is indicated. In some cases, the cut values change for different η ranges as the material traversed by the electron increases toward high η . These ranges are $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.479$, and $1.479 < |\eta| < 2.5$ and the respective cut values are given in the form (value₁, value₂, value₃). Cuts marked with a † are applied to leptons failing the lepton MVA cut, while the loose cut value is applied to those that pass.

Cut	Loose	Fakeable Object	Tight
$ \eta < 2.5$	✓	✓	✓
p_T	> 7 GeV	> 15 GeV	> 15 GeV
$ d_{xy} < 0.05$ (cm)	✓	✓	✓
$ d_z < 0.1$ (cm)	✓	✓	✓
$SIP_{3D} < 8$	✓	✓	✓
$I_{\text{mini}} < 0.4$	✓	✓	✓
MVA(HZZ) ID	–	$> (0.0, 0.0, 0.7)^\dagger / -$	–
$\sigma_{i\eta i\eta} < (0.011, 0.011, 0.030)$	–	✓	✓
$H/E < (0.10, 0.10, 0.07)$	–	✓	✓
$\Delta\eta_{in} < (0.01, 0.01, 0.008)$	–	✓	✓
$\Delta\phi_{in} < (0.04, 0.04, 0.07)$	–	✓	✓
$-0.05 < 1/E - 1/p < (0.010, 0.010, 0.005)$	–	✓	✓
p_T^{ratio}	–	$> 0.5^\dagger / -$	–
jet CSV	–	$< 0.3^\dagger / < 0.8484$	< 0.8484
tight-charge	–	–	✓
conversion rejection	–	–	✓
Number of missing hits	< 2	$== 0$	$== 0$
lepton MVA > 0.90	–	–	✓

approximate BR are given in Table [6.10](#).

In addition, they have to fulfil $p_T > 20$ GeV, $|\eta| < 2.3$ and loose working point of tau discriminator which has an efficiency of $\sim 60\%$ to identify the τ_h [\[156\]](#). The track quality criteria required of the selected taus are: $|d_{xy}| < 1$ mm and $|d_z| < 0.2$ cm. The hadronically

Table 6.10: Approximate branching fractions of 1- and 3- prong hadronic decay of τ [25].

Decay mode	BR (%)
$\tau^- \rightarrow h^- \nu_\tau$	11.5
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	26.0
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	9.5
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	9.8
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$	4.8
Other	3.2

reconstructed taus must be away from any muons/electrons passing the loose selection by $\Delta R > 0.4$.

6.1.6 Jets and b-tagging

The analysis uses anti- k_T (0.4) and PF jets (see Chapter 3), corrected for charged hadrons not coming from the primary vertex (charged hadron subtraction), and having JECs applied as a function of the jet E_T and η . Jets are only considered if they have a transverse momentum more than 25 GeV within $|\eta| < 2.4$ and 40 GeV within $2.4 < |\eta| < 4.7$. The selection of p_T threshold in the forward region is driven by better data/MC agreement, discussed in Appendix B. They pass the loose identification criteria, shown in Table 6.11, which has greater than 98% efficiency to select jets and $\sim 99.8\%$ noise rejection. Jets within $|\eta| < 2.4$ are also required to have non-zero charged hadron energy fraction and charged multiplicity, and charged EM fraction less than 0.99. In addition, the selected

Table 6.11: PF jet identification criteria used in 2016 analysis [117].

variable	$ \eta < 2.7$	$2.7 < \eta < 3.0$	$ \eta > 3.0$
Neutral hadron energy fraction	< 0.99	< 0.98	–
Neutral EM energy Fraction	< 0.99	> 0.01	< 0.90
Number of Constituents	> 1	–	–
Number of neutral particles	–	> 2	> 10

jets must be separated from any fakeable lepton and hadronic tau candidates by $\Delta R > 0.4$. The loose and medium working points of the CSVv2 b-tagging algorithm, defined by discriminator threshold values 0.5426 and 0.8484 respectively, are used to identify b-jets within $|\eta| < 2.4$.

Data/simulation differences in the b-tagging performance are corrected by applying per-jet weights to the simulation, dependent on the jet p_T , η , b-tagging discriminator and jet flavor [157]. The per-event weight is taken as the product of the per-jet weights, including those of the jets associated to the leptons. A set of b-tagging efficiency scale factors, SF, are provided to CMS analyses for b- and light-flavor jets, where:

$$\text{SF}(\text{CSV}, p_T, \eta) = \frac{\text{DATA} - \text{MC}_A}{\text{MC}_B} \quad (6.7)$$

where A/B is light-/heavy-flavor for heavy-flavor SF or A/B heavy-/light-flavor for light-flavor SF. In this analysis the shape of b-tagging discriminator distribution is corrected using per-event weight calculated as:

$$\text{SF}(\text{event}) = \prod_i^{N_{\text{jets}}} \text{SF}_{\text{jet}}. \quad (6.8)$$

6.1.7 Event selection

Common selection criteria for the 3ℓ and $2\ell\text{ss}$ final states are:

- Invariant mass of any loose lepton pair ($m_{\ell\ell} > 12$ GeV)
- One or more jets passing CSVv2 medium working point
- At least one jet failing CSVv2 loose working point to select the light quark jet

Three leptons (3ℓ)

All three W bosons from top quark and Higgs decay leptonically and flavor combinations $\mu\mu\mu$, $\mu\mu e$, μee and eee are considered:

- Exactly three tight leptons, the sum of the charges is ± 1
- Lepton p_T thresholds of 25/15/15 GeV respectively for the leading and sub-leading leptons

- No opposite sign same flavor lepton pair with invariant mass within 15 GeV of the Z boson mass: $|m_{\ell\ell} - 91.2| > 15$ GeV

Two leptons same-sign ($2\ell ss$)

Such a final state arises when $H \rightarrow WW^*$ and the W bosons from top and Higgs boson having the same charge decay leptonically while the oppositely charged W from the Higgs decays hadronically. Depending on the combinations of lepton flavor, the same-sign final state is sub-divided into two channels, $\mu^\pm\mu^\pm$ and $e^\pm\mu^\pm$, both requiring:

- Exactly two tight leptons of same-sign
- Lepton p_T thresholds of 25 and 15 GeV respectively for the leading and sub-leading leptons
- Selected tight leptons must be tight charge consistent

The same-sign di-electron channel suffers from larger charge mis-identification background and does not add sensitivity and is therefore not included in the analysis.

Applying the selections and performing the background estimations described earlier, the yields shown in Table 6.12 are obtained. As can be seen, the selected sample in each case is very much dominated by the backgrounds due to fakes and charge mis-identifications contributing to about 50% or more of the overall numbers. The second most important background is from $t\bar{t}W$ followed by $t\bar{t}Z$. The signal yield is split into tHq , tHW and $t\bar{t}H$. For the first two processes, expected yields from ITC scenario also has been mentioned for comparison. To extract a possible signal in the selected sample, MVA is used.

6.1.8 Signal discrimination

Two BDT classifiers are trained on simulated tHq sample against two major background processes expected: events with prompt leptons from $t\bar{t}W$ and $t\bar{t}Z$ (also referred to as $t\bar{t}V$), and events with non-prompt leptons from inclusive $t\bar{t}$. Different training samples used are weighted proportionally to their respective cross sections. The trained algorithm

Table 6.12: Data yields and expected backgrounds after the event pre-selection for the three final states corresponding to an integrated luminosity 35.9 fb^{-1} , rounded off to nearest integral value.

Process	3ℓ	$e^\pm\mu^\pm$	$\mu^\pm\mu^\pm$
$t\bar{t}W$	23	97	68
$t\bar{t}Z$	32	65	26
WZ, ZZ	10	29	16
Rares	8	35	20
Conversion	7	24	–
Non-prompt	26	135	81
Charge flips	–	58	–
Total background	106	443	211
$t\bar{t}H$	18	35	24
tHq (SM)	1	2	1
tHW (SM)	1	1	1
Total (SM)	126	481	237
tHq (ITC)	8	27	18
tHW (ITC)	7	11	8
Total (ITC)	139	516	261
Data (35.9 fb^{-1})	127	525	280

is then used to classify events from the different MC samples and data. The discriminating variables can be put into three categories as follows.

- **Forward activity:** This analysis is optimized for the tHq signal process which has a forward light quark jet. None of the main backgrounds have any forward activity, which makes the forward jet kinematics a good discriminating variable.
- **Jet and b-jet multiplicity:** Both tHq and tHW processes contain one b-jet from the top quark decay, whereas $t\bar{t}V$ processes generally contain at least two hard b-jets. In case of the fake lepton background, since one of the b-jets most often produces the

fake lepton, it is no longer reconstructed as a b-jet, hence inclusive $t\bar{t}$ events most often contain only one b-tagged jet. However, the total number of jets in the event may be larger. Thus counting the number of jets, tagged or un-tagged, provides discrimination power. It is to be noted that in case of tHq which has a b quark in the initial state, a second b-jet may be reconstructed from the gluon-splitting in 4FS. However, most of the time it may not be tagged as it is in the forward region and b-tagging is relevant within $|\eta| < 2.4$.

- **Lepton kinematics and lepton charge:** The two leptons in the signal process are expected to be more back-to-back than those in the main backgrounds, which are essentially evenly distributed. The p_T of the trailing lepton will generally be harder in the signal processes, since it originates from an on-shell W boson, compared to fake leptons produced by the decay of a b-hadron. Finally, the signal process tHq shows a strong asymmetry in lepton charge due to the prevalent positive charges in the incoming protons. The same is true for $t\bar{t}V$ processes and WZ , but inclusive $t\bar{t}$ is a neutral state and will therefore produce leptons of each charge with equal probability. This may help discriminate against the fake background from the signal.

In Table 6.13 the input variables to the training are listed. Detailed discussion of the BDT trainings are documented in Appendix A. Figures 6.1, 6.2 and 6.3 show the distributions of the BDT discriminators “kinMVA” for the different trainings, left (right) plots showing tHq against $t\bar{t}$ ($t\bar{t}V$), in the channels 3ℓ , $e^\pm\mu^\pm$ and $\mu^\pm\mu^\pm$ respectively. The data/MC disagreement observed is attributed to statistical fluctuation of data.

6.1.9 Signal extraction

The two kinMVA distributions are used to evaluate the limits on the signal strength in a fit to the classifier shape. Figures 6.4 and 6.5 show the expected output distributions in a 2-dimensional plane of one training vs. the other for the 3ℓ and $2\ell ss$ channels respectively. The $e^\pm\mu^\pm$ and $\mu^\pm\mu^\pm$ channels are combined since the separation between signal and background are similar (see Figs 6.3 and 6.2). Note, main backgrounds to 3ℓ events are $t\bar{t}V$ processes, whereas in $2\ell ss$ inclusive $t\bar{t}$ process dominates. Each event is then

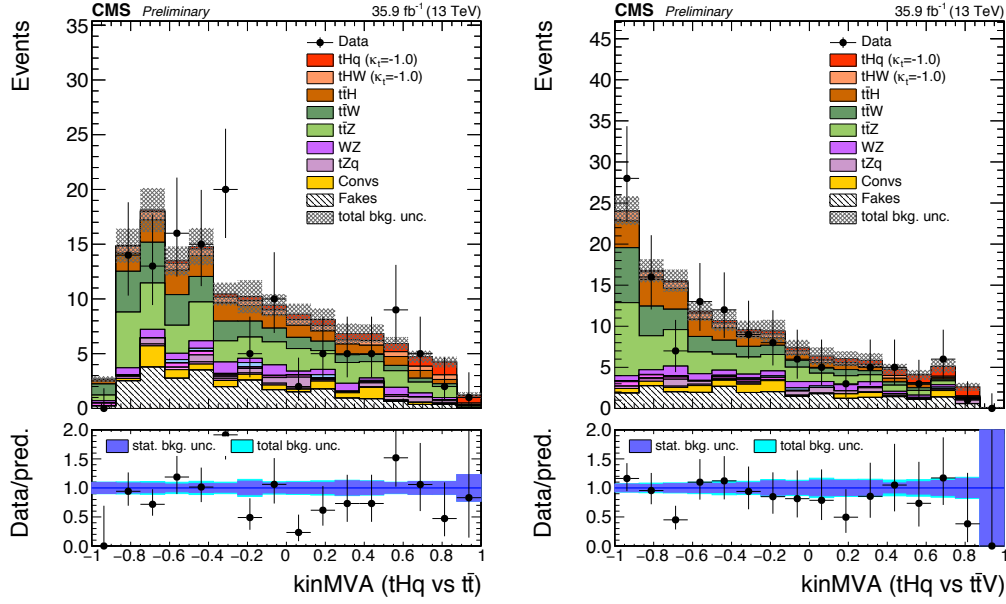


Figure 6.1: BDT output distributions in 2016 analysis for 3ℓ channel. The discriminator for tHq vs. $t\bar{t}$ ($t\bar{t}V$) is shown on the left (right). In both cases tHq events appear more towards the discriminator value 1.

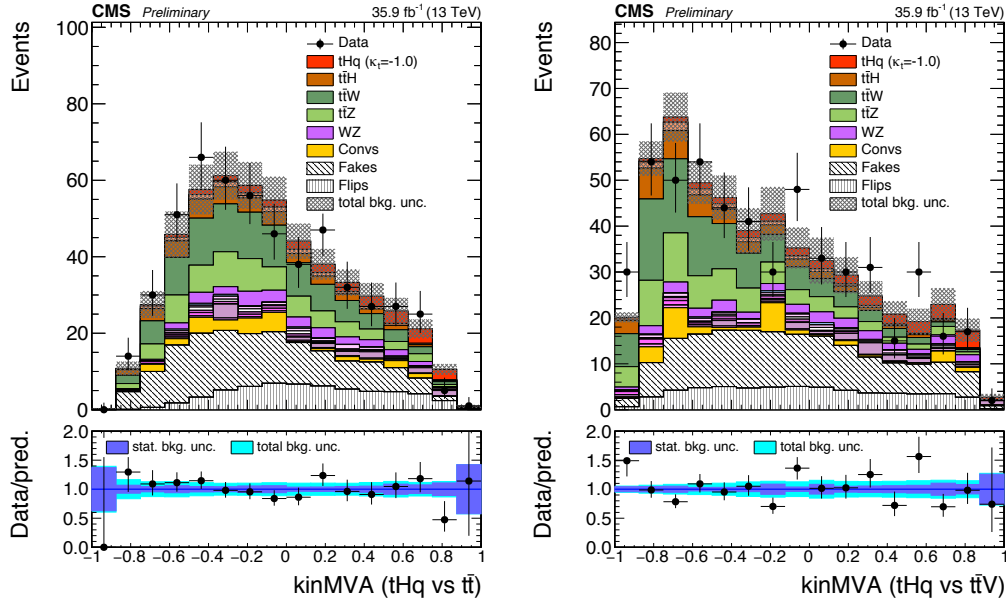


Figure 6.2: BDT output distributions in 2016 analysis for $e^\pm\mu^\pm$ channel. The discriminator for tHq vs. $t\bar{t}$ ($t\bar{t}V$) is shown on the left (right). In both cases tHq events appear more towards the discriminator value 1.

Table 6.13: List of input variables used for the MVA training.

Variable type	Description
Lepton kinematics	p_T of the trailing lepton
	Sum of lepton charges
	Minimum ΔR between any two fakeable leptons
	$\Delta\phi$ between highest p_T lepton pair
Jet multiplicity	Number of jets with $p_T > 25$ GeV and $ \eta < 2.4$
	Number of jets with $ \eta > 1.0$, non CSVv2 loose
Forward jet activity	Maximum $ \eta $ of any non CSVv2 loose jet with $p_T > 25$ GeV
	$\Delta\eta$ between forward light jet and hardest CSVv2 loose jet
	$\Delta\eta$ between forward light jet and second hardest CSVv2 loose jet (defaults to -1 in events with only one CSVv2 loose jet)
	$\Delta\eta$ between forward light jet and closest lepton

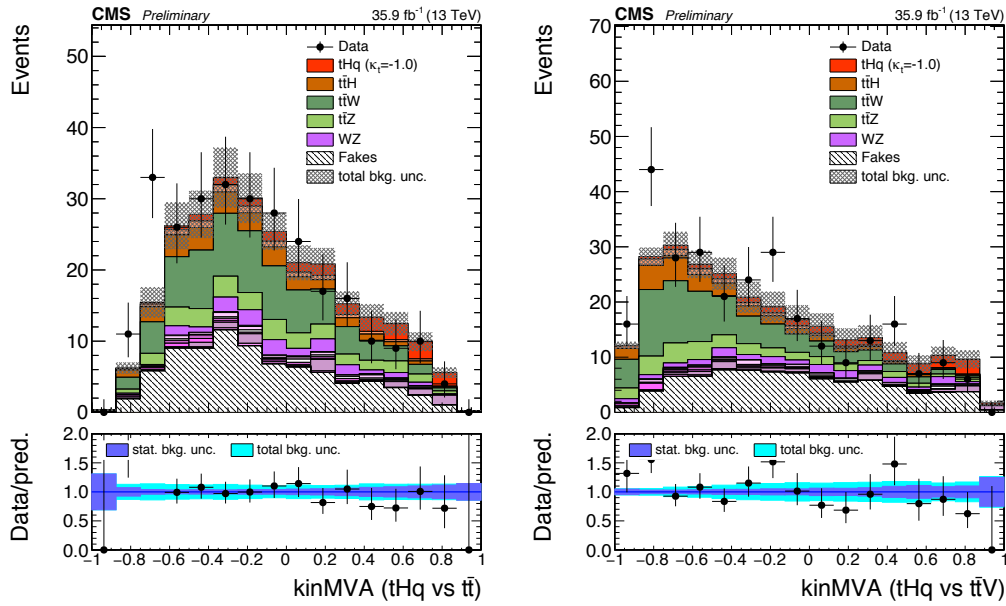


Figure 6.3: BDT output distributions in 2016 analysis for $\mu^\pm\mu^\pm$ channel. The discriminator for tHq vs. $t\bar{t}$ ($t\bar{t}V$) is shown on the left (right). In both cases tHq events appear more towards the discriminator value 1.

classified using bin boundaries into one of several 2-dimensional bins according to its position in the plane. The number of bins is chosen such that no bins are entirely empty for any process. The bin boundary positions have been studied to optimize the effect on the expected limit on signal strength. Several sizes and binning combinations were tested and optimized with respect to the expected limit. A set of ten bins for each channel provided the best expected limits, shown in Fig. 6.6. From this event categorization, a 1-dimensional histogram (BDT bin) of expected distribution is produced for each signal and background processes, and fitted to the observed data (or the Asimov dataset for expected limits).

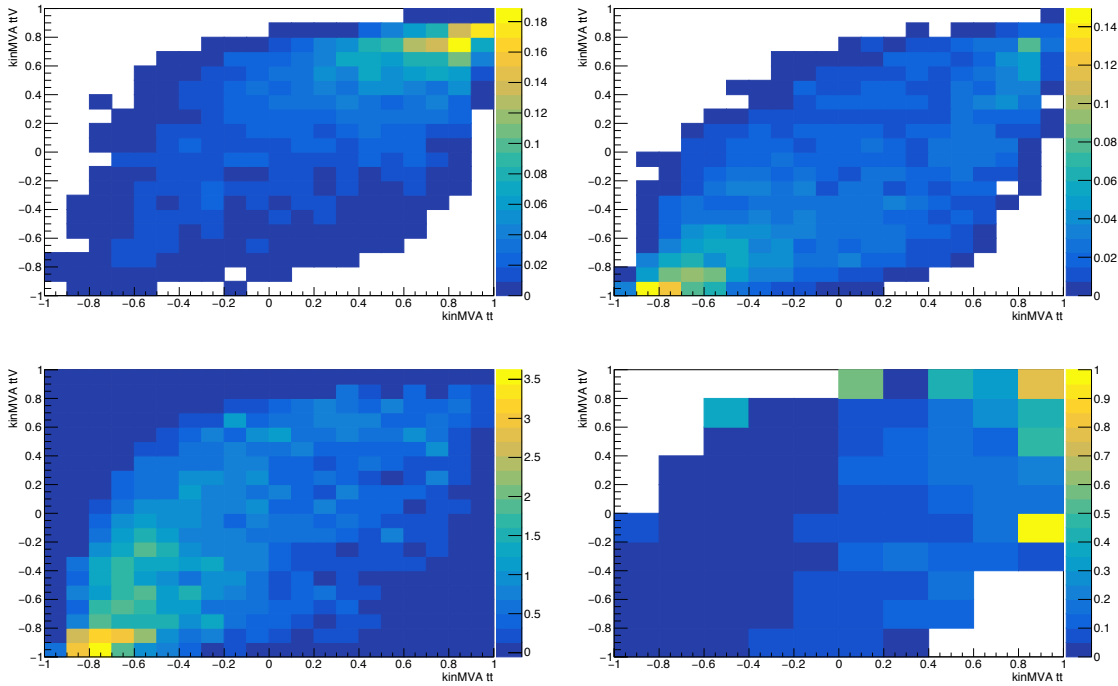


Figure 6.4: BDT classifier output planes (training vs $t\bar{t}$ on x -axis and vs $t\bar{t}V$ on y -axis) for the tHq and tHW signals (top row, left and right respectively), and for the combined backgrounds (bottom left) for the 2016 analysis in 3ℓ channel. Bottom right shows the S/B ratio (combining tHq and tHW) in the same plane.

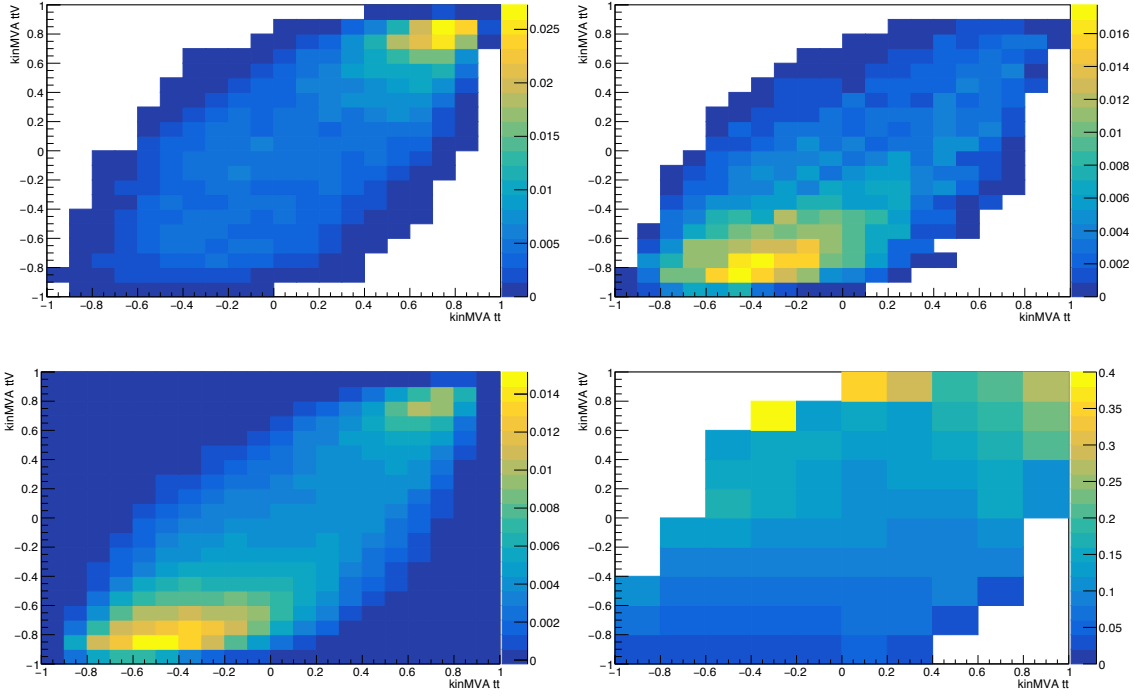


Figure 6.5: BDT classifier output planes (training vs $t\bar{t}$ on x -axis and vs $t\bar{t}V$ on y -axis) for the tHq and tHW signals (top row, left and right respectively), and for the combined backgrounds (bottom left) for the 2016 analysis in $2\ell ss$ channel. Bottom right shows the S/B ratio (combining tHq and tHW) in the same plane.

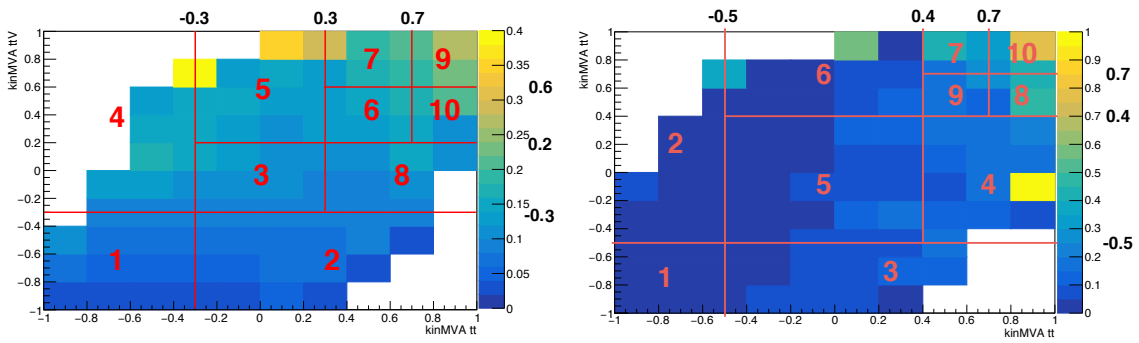


Figure 6.6: Binning overlaid on the S/B ratio map on the plane of classifier outputs for $2\ell ss$ channel (left) and 3ℓ channel (right) in 2016 analysis. The bold numbers along the axis refer to the bin boundaries.

6.1.10 Systematic uncertainties

Since all processes are statistical in nature, there is inherent uncertainty in determining the yields from data and simulation. Further, imperfect knowledge of the theory (such as cross sections) and experiment (such as measured luminosity) add to the error in measurement. Various sources of systematic uncertainties are thus considered, in order to account for the above effects in data and MC. Both the event yields and the shape of the BDT bin distribution used in the fit between data and MC can be impacted due to changes in either one of the quantities σ , L , ϵ or A of Eq.(6.1). A brief description of all systematic uncertainties considered in this analysis are as follows.

Experimental uncertainties

- **Luminosity:** The uncertainty associated with integrated luminosity measured in CMS experiment is $\pm 2.5\%$ [158]. It appears as a normalization-only uncertainty is applied to the total predicted yields.
- **Pileup:** Event weights are corrected with scale factors that account for the differences in the pileup distributions between data and MC samples (before any event selection is applied). To estimate the corresponding uncertainty, the p-p inelastic cross section (69.2 mb) is varied by $\pm 5\%$ and the weights are re-derived. The impact on the result was found to be negligible by the $t\bar{t}H$ analysis [137], and this source of uncertainty was thus not considered.
- **Trigger efficiency:** Per-event scale factors are applied depending on the final state. The uncertainties are 1-3% as given in Table 6.6.
- **Lepton identification efficiency:** The uncertainties attached to the probability for electrons (muons) to pass the loose and tight selection criteria (Tables 6.8 and 6.9) are determined to be 2% and 3% respectively for electrons and muons.
- **b-tagging efficiency:** Uncertainties related to the b-tagging efficiency and mis-tag rate are provided [157] as a function of the jet p_T and η . To account for the different

individual sources of uncertainty, separate shapes are obtained by varying the b-tagging scale factor by the corresponding uncertainty.

- Jet energy scale: Uncertainties related to the jet energy scale are provided [117] as a function of the jet p_T and η . To account for this source of uncertainty, events are re-analyzed after the four-momenta of the jets have been varied within these uncertainties, as well as the b-tagging efficiency scale factor.
- Forward jet: The forward jet η distribution is poorly modelled in simulation. To estimate the effect of a mis-modelled forward jet distribution, the events in simulation are reweighted based on the normalized data/MC ratio in a $t\bar{t}$ control region to thereby derive an alternative shape of the BDT output distributions that reflects a hypothetical perfect data/MC agreement, see Appendix B.

Theory uncertainties

- PDF, α_s (cross section normalization): Normalization-only uncertainties are applied to the predicted yields of the $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, tHq and tHW processes, to account for the uncertainties due to the choice of PDF, α_s and missing higher orders on the NLO theory calculations [159], provided along with the generated events.
- Interaction energy scale (shape and acceptance): Uncertainties on the interaction energy scale affect the shape and acceptance of the distributions used for signal extraction. Part of this effect is assessed by varying the renormalization scale (μ_R) and factorization scale (μ_F) between 1/2 and 2 (with the constraint $1/2 < \mu_R/\mu_F < 2$), and fitting linearly the difference between the nominal shape (used for signal extraction) and the varied shapes, following the $t\bar{t}H$ multi-lepton analysis [137].

Background estimation

- Electroweak background: The normalization uncertainty for the WZ background in the $2\ell ss$ channel is conservatively taken to be 50% due to the presence of larger number of jets. In the 3ℓ channel, the corresponding statistical uncertainty is taken to be 10%, the background uncertainty 20% and the theoretical uncertainty 7%.

- Rare SM backgrounds: Smaller backgrounds which have not yet been observed at the LHC are considered with a 50% normalization uncertainty.
- Conversion background: The contribution of photon conversions estimated from MC is considered to have 30% normalization uncertainty.
- Charge-flips: The mis-identified electron charge background is considered with a 30% normalization uncertainty from the measurement of the charge-flip probability.
- Non-prompt backgrounds: The main sources of uncertainty for these backgrounds arise from statistical fluctuations, the subtraction of the prompt MC contribution, and the difference in composition between the enriched side-band and the application region. The combined uncertainty is taken into account using alternative shapes obtained using a variable fake rate. In addition, a closure test is also performed to ensure consistency of the fake rates between data and MC. The BDT discriminator distributions due to fakes in the signal regions are evaluated in two ways using simulated $t\bar{t}$ events in two steps: first is by requiring to pass the tight lepton criteria and in second step passing the fakeable criteria and extrapolated using fake rate measured in QCD multi-jet events. The ratio is fitted to a constant and the resulting value is used as normalization uncertainty for the closure.

Of these sources of systematic uncertainties, the ones with largest impact on the final result are found to be those related to the normalization of the non-prompt backgrounds, the scale variations for the $t\bar{t}V$ and $t\bar{t}H$ processes, and the lepton selection efficiencies.

6.1.11 Results

The 1-dimensional BDT bin distribution is fitted using the Higgs combine tool [147], considering all systematic uncertainties described above. Since the leptons in the final state can also arise from $H \rightarrow \tau\tau$ which involves the coupling y_τ , it is assumed that the coupling modifier κ_τ scales the same way as κ_t . The maximum likelihood fit (see Chapter 5) adjusts the background estimations within given uncertainties for the expected model

(SM or ITC). The best-fit $tH+t\bar{t}H$ combined signal strength observed for the standard model hypothesis is $1.8 \pm 0.3(\text{stat.}) \pm 0.6(\text{syst.})$ corresponding to an observed significance of 2.7σ (1.5σ expected) over a background only (no Higgs) hypothesis. For the ITC scenario ($\kappa_t = -\kappa_V = -1$), the best-fit signal strength is 0.7 ± 0.4 with observed significance of 1.7σ (2.5σ expected). Figures 6.7, 6.8 and 6.9 show the pre-fit (without normalization adjusted) and post-fit (with normalization adjusted) distributions of the BDT bins in channels 3ℓ , $e^\pm\mu^\pm$ and $\mu^\pm\mu^\pm$ respectively. The plots can be compared to note the adjusted total uncertainty values and background normalization.

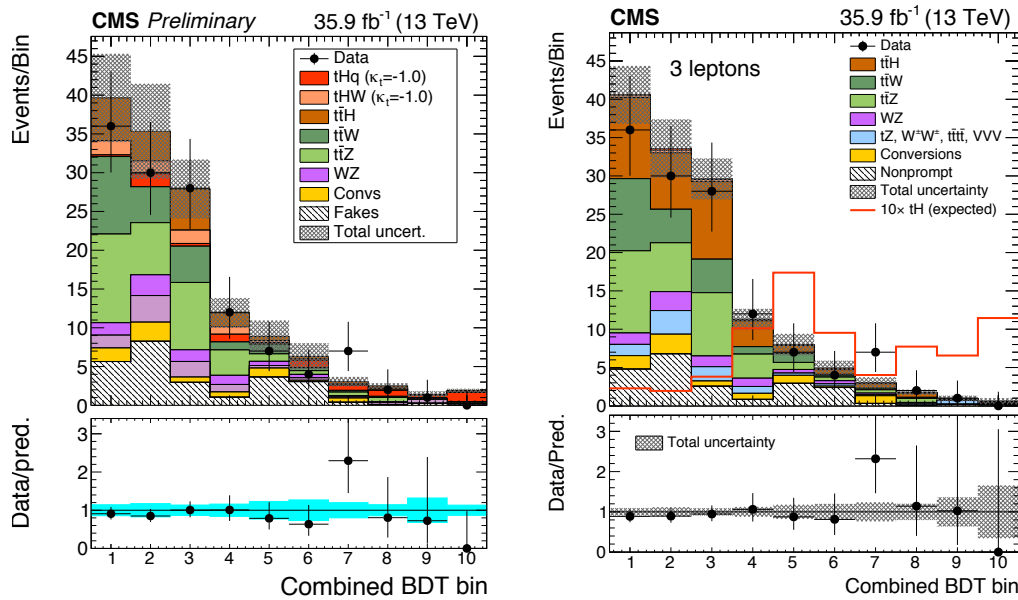


Figure 6.7: Pre-fit (left) and post-fit (right) BDT bin distributions in 2016 analysis for 3ℓ final state where data is plotted along with simulation. The post-fit figure shows the shape of tH signal with 10 times the SM expectation.

The upper limits on $\sigma \times BR$ for $tH+t\bar{t}H$ processes in the $H \rightarrow WW^*/ZZ^*/\tau\tau$ decay modes are obtained as a function of κ_t/κ_V along with the expectation, shown in Fig. 6.10. The observed upper limit is compared to the theoretical prediction of $tH+t\bar{t}H$ cross sections times branching fraction for $\kappa_V = 1.0$ to constrain the allowed ranges of κ_t . The values of κ_t outside the range $(-1.25, 1.6)$ are excluded with 95% CL.

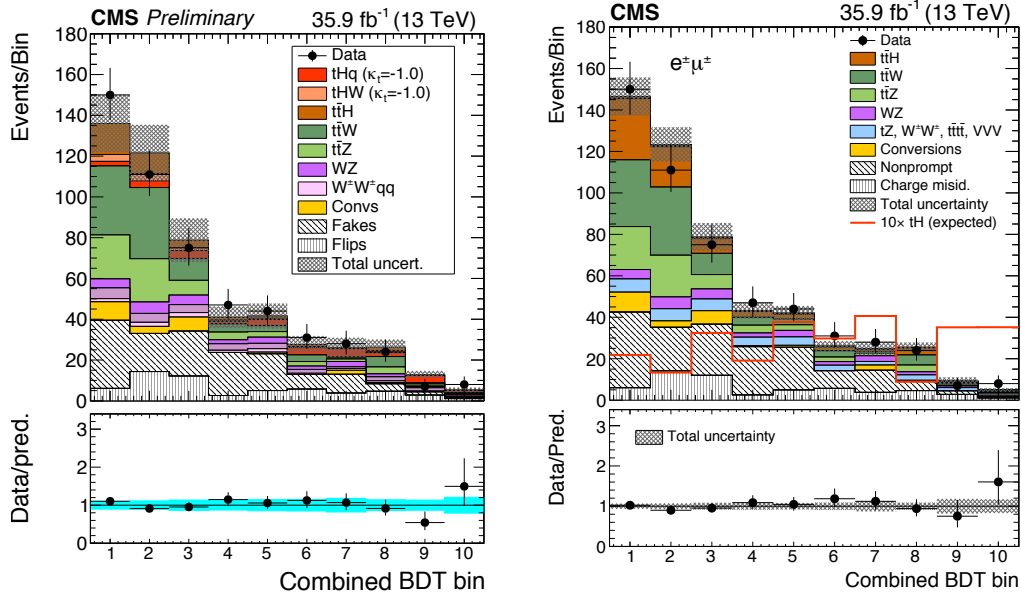


Figure 6.8: Pre-fit (left) and post-fit (right) BDT bin distributions in 2016 analysis for $2\ell ss e^\pm \mu^\pm$ final state where data is plotted along with simulation. The post-fit figure shows the shape of tH signal with 10 times the SM expectation.

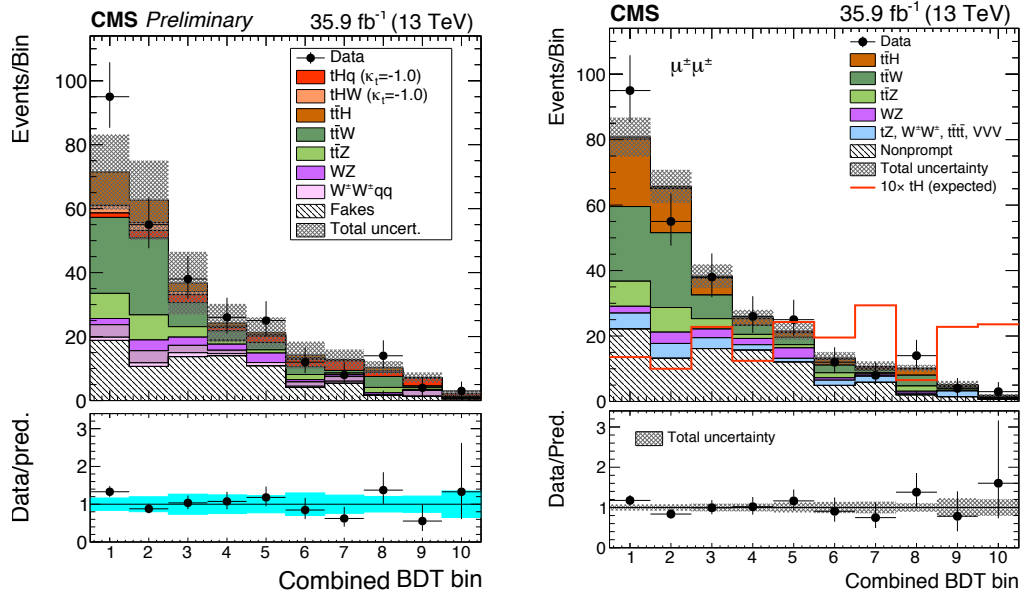


Figure 6.9: Pre-fit (left) and post-fit (right) BDT bin distributions in 2016 analysis for $2\ell ss \mu^\pm \mu^\pm$ final state where data is plotted along with simulation. The post-fit figure shows the shape of tH signal with 10 times the SM expectation.

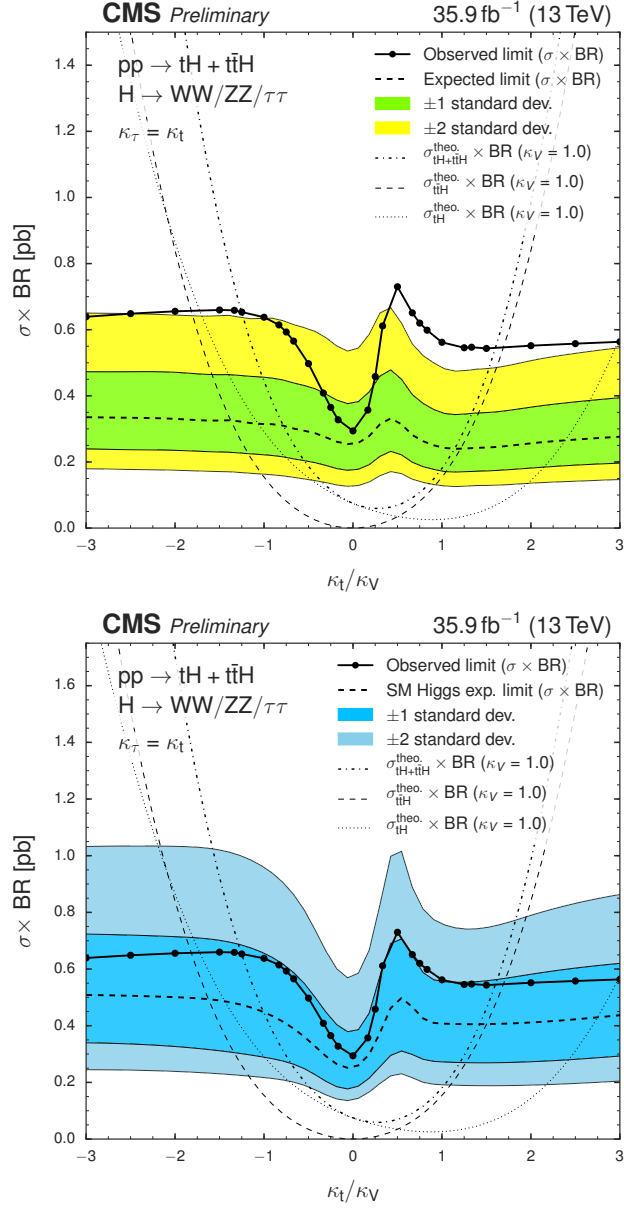


Figure 6.10: Observed and expected upper limits on $\sigma \times BR$ as a function of κ_t / κ_V in 2016 data. Top plot shows the limits in a background only (no Higgs) hypothesis, whereas the bottom plot shows the limits for signal+background hypothesis. The excess observed in the top figure is due to the presence of $t\bar{t}H$ events, which the background only hypothesis assumes as zero. The shape of observed curve denotes a maxima near $\kappa_t / \kappa_V = 1.0$ which is multiplied by the $\sigma \times BR$ prediction having minima near $\kappa_t / \kappa_V = 0.0$ resulting in the curved shape. The bottom figure is used to derive the constraints on the κ_t values for $\kappa_V = 1.0$ by comparing the observed limit with the theoretical prediction.

The limits on a tH only signal as a function of κ_t are obtained under resolved Higgs couplings hypothesis, where interference between κ_t and κ_V is accounted for in $H \rightarrow \gamma\gamma$ decay and $gg \rightarrow H$ production. The $t\bar{t}H$ process is included in the background assuming its cross section varies with κ_t with $\mu_{t\bar{t}H}$ fixed at 1.0. The observed (expected) upper limit on tH only signal strength μ_{tH} at 95% CL is **1.02 (0.92)** for the ITC scenario and **17.72 (15.31)** for the SM. Figure 6.11 shows the observed and expected upper limits on tH only cross section times branching fraction for different κ_t scenarios, assuming $\kappa_V=1.0$.

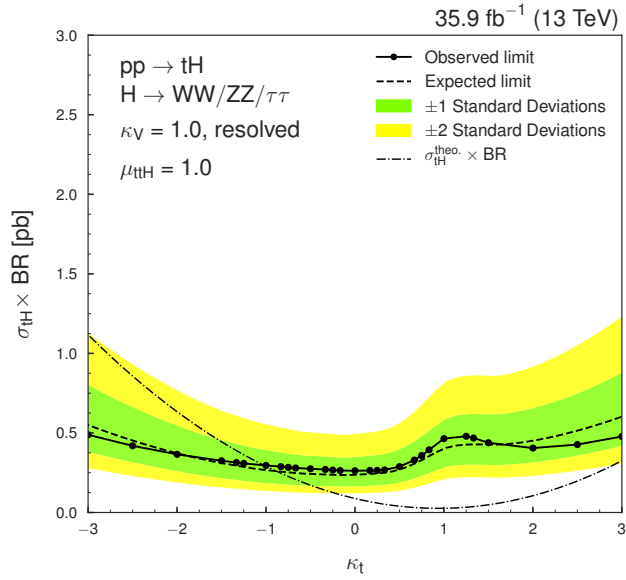


Figure 6.11: Upper limits on cross section times branching ratio of $tH, H \rightarrow WW^*/ZZ^*/\tau\tau$ signal as a function of κ_t considering background only hypothesis includes $t\bar{t}H$ production with a fixed signal strength $\mu_{t\bar{t}H} = 1.0$. The upper limits on μ_{tH} has a maxima near $\kappa_t=1.0$.

The observed limits are much closer to the expected values in this interpretation of the tH signal as compared to Fig. 6.10. In the earlier interpretation $t\bar{t}H$ is included in the signal hypothesis, whereas in Fig. 6.11 it is treated as background. The observation in Fig. 6.10 is supported by excess of events seen in the dedicated $t\bar{t}H$ analysis [137].

6.2 Analysis using 2017 data

6.2.1 Data and simulated samples

The 2017 analysis for tHq multi-lepton final state corresponds to an integrated luminosity of 41.5 fb^{-1} . The 17Nov2017 re-reco versions of the following datasets (runs B to F) are analyzed:

- DoubleMuon, SingleMuon, MuonEG, SingleElectron, DoubleEG.

The JSON file used to select the data events is:

- Cert_294927-306462_13TeV_E0Y2017ReReco_Collisions17_JSON.txt.

MC samples used in this analysis are generated using the CP5 tune [114] for underlying events. The signal samples tHq , tHW and $t\bar{t}H$ were generated with LHE reweighting based on κ_t , κ_V and including different α values in 2017, listed in Table 6.14. The generator used is MG5_aMC@NLO (version 5.242) in LO mode, samples have been scaled to the NLO cross sections shown in Table 6.14.

Table 6.14: Production cross sections (inclusive) for the LO tH and $t\bar{t}H$ samples in the ITC scenario used in 2017 analysis. The NLO cross sections as used in the analysis are mentioned. See Ref. [151] for more details.

Sample	σ [pb]
THQ_ctcvcp_4f_Hinc1_13TeV_madgraph_pythia8	0.7927
THW_ctcvcp_5f_Hinc1_13TeV_madgraph_pythia8	0.1472
TTH_4f_ctcvcp_TuneCP5_13TeV_madgraph_pythia8	0.5072

The background MC samples used for training against the signal are listed in Table 6.15.

Similar to 2016 analysis, different generators have been used to obtain the prompt lepton background samples, listed in Table 6.16. Since $W\gamma$ and $Z\gamma$ MC samples were unavailable in 2017, the conversion backgrounds have been estimated using W +jets and DY +jets samples. The tZq process has been observed by CMS experiment in 2017 [160], thus this

Table 6.15: LO $t\bar{t}W$ and $t\bar{t}Z$ samples and NLO $t\bar{t}$ samples used in the signal BDT training in 2017 analysis, mentioned together with the NLO cross sections.

Sample	$\sigma \times BR$ [pb]
ttWJets_TuneCP5_13TeV_madgraphMLM_pythia8	0.6105
ttZJets_TuneCP5_13TeV_madgraphMLM_pythia8	0.7655
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	87.3148
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	87.3148
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	364.35
TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	364.35
TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	831.76
TTJets_DiLept_TuneCP5_13TeV-madgraphMLM-pythia8	87.3148
TTJets_SingleLeptFromT_TuneCP5_13TeV-madgraphMLM-pythia8	182.1754
TTJets_SingleLeptFromTbar_TuneCP5_13TeV-madgraphMLM-pythia8	182.1754

background is considered separately from rare SM processes in this analysis. Also, as the MC sample for the SM process $t\bar{t}WW$ was made available, the same has been added to the backgrounds. Data driven estimate follows the $t\bar{t}H$ multi-lepton analysis [138], as in 2016.

The data and MC samples are both reconstructed using CMSSW version 9_4_6_patch1, and analyzed with the same.

6.2.2 Triggers

As the trigger menus evolve over the years, different HLT paths are available as compared to 2016. The overlap removing triggers between the primary datasets are listed in Table 6.17.

Table 6.18 shows the lowest-threshold, un-prescaled triggers present in HLT menus for both MC and data in 2017 used in this analysis. The efficiency of events to pass the trigger is measured in MC and in the data for the $t\bar{t}H$ multi-lepton [138], which have been implemented in this analysis as well. Small differences between the data and MC efficiencies are corrected by applying scale factors as shown in Table 6.19, as a function of the lepton flavor and p_T .

Table 6.16: List of background MC samples used to extract the final yields and distributions, together with NLO and NNLO cross sections, as applicable.

Process	Sample	$\sigma \times BR$ (pb)
$t\bar{t}W$	TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
$t\bar{t}Z$	TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8	0.2529
	TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8	0.05324
$t\bar{t}WW$	TTWW_TuneCP5_13TeV-madgraph-pythia8	0.011500
EWK	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	5.063
(WZ/ZZ)	ZZTo4L_13TeV_powheg_pythia8	1.256
Conversion	TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	4.09
	TGJets_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	1.018
	W1JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8	9503.91
	W2JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8	3258.45
	W3JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8	1162.278
	W4JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8	634.608
	DY1JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	878
	DY2JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	307
	DY3JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	112
	DY4JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	44.2
	DYJetsToLL_M-4to50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	202.8
	DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	53.7
	DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	5.66
	DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_13TeV-madgraphMLM-pythia8	1.852
tZq	tZq_11_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	0.0942
Rares	GluGluHTtoZZTo4L_M125_13TeV_powheg2_JHUGenV7011_pythia8	0.01212
	VHTtoNonbb_M125_13TeV-amcatnloFXFX-madspin-pythia8_Fall17	0.9561
	TTTT_TuneCP5_13TeV-amcatnlo-pythia8	0.009103
	TTTW_TuneCP5_13TeV-madgraph-pythia8	0.0007273
	TTWH_TuneCP5_13TeV-madgraph-pythia8	0.00114
	ST_tW11_5f_L0_TuneCP5_PSweights_13TeV_madgraph_pythia8_Fall17	0.01123
	WW_DoubleScattering_13TeV-pythia8_TuneCP5	0.1729
	WpWpJJ_EWK_QCD_TuneCP5_13TeV_madgraph_pythia8_Fall17	0.03711
	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.2086
	WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.1651
	WZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.05565
	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.01398
	WZG_TuneCP5_13TeV-amcatnlo-pythia8	0.04345

Table 6.17: Table of high-level triggers used in 2017 analysis for the overlap removal between datasets.

Dataset	Trigger
DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*, HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v*, HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v*, HLT_Mu19_TrkIsoVVL_Mu9_TrkIsoVVL_DZ_Mass3p8_v*, HLT_Mu19_TrkIsoVVL_Mu9_TrkIsoVVL_DZ_Mass8_v*, HLT_TripleMu_10_5_5_DZ_v*, HLT_TripleMu_12_10_5_v*, HLT_TripleMu_5_3_3_Mass3p8to60_DCA_v*, HLT_TripleMu_5_3_3_Mass3p8to60_DZ_v*
DoubleEG	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*, HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL_v*
MuonEG	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*, HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ_v*, HLT_Mu8_DiEle12_CaloIdL_TrackIdL_v*, HLT_Mu8_DiEle12_CaloIdL_TrackIdL_DZ_v*
SingleMuon	HLT_IsoMu24_v*, HLT_IsoMu24_eta2p1_v*, HLT_IsoMu27_v*
SingleElectron	HLT_Ele32_WPTight_Gsf_v*, HLT_Ele35_WPTight_Gsf_v*

6.2.3 Identification of leptons

The baseline object identification for the 2017 analysis remains same as in 2016 for leptons, as well as the isolation definition. The effective areas for lepton isolation have been updated, shown in Table 6.20.

More importantly, the lepton MVA discriminator has been retuned for 2017 data [138] and following are the changes to the variables of the associated jet.

- jet CSV : the DeepCSV b-tagging discriminator value of the jet
- p_T^{ratio} : the ratio between the p_T of the lepton and the p_T of the jet if jet CSV is

Table 6.18: List of HLT paths considered in the 2017 analysis for event selection.

Object	Trigger
one electron	HLT_E1e32_WPTight_Gsf_v* HLT_E1e35_WPTight_Gsf_v*
one muon	HLT_IsoMu24_v* HLT_IsoMu27_v*
two electrons	HLT_E1e23_E1e12_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_E1e23_E1e12_CaloIdL_TrackIdL_IsoVL_v*
two muons	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v* HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v*
one electron, one muon	HLT_Mu23_TrkIsoVVL_E1e12_CaloIdL_TrackIdL_IsoVL_v* HLT_Mu23_TrkIsoVVL_E1e12_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu12_TrkIsoVVL_E1e23_CaloIdL_TrackIdL_IsoVL_DZ_v*
three electrons	HLT_E1e16_E1e12_E1e8_CaloIdL_TrackIdL_v*
two electrons, one muon	HLT_Mu8_DiE1e12_CaloIdL_TrackIdL_v*
one electron, two muons	HLT_DiMu9_E1e9_CaloIdL_TrackIdL_DZ_v*
three muons	HLT_TripleMu_12_10_5_v*

Table 6.19: Trigger efficiency scale factors and associated uncertainties applied to the simulated events in 2017 analysis.

Channel	Leading lepton p_T	Scale factor
3ℓ	–	1.000 ± 0.050
$e\mu$	< 35 GeV	0.952 ± 0.008
	≥ 35 GeV and < 50 GeV	0.983 ± 0.003
	≥ 50 GeV	1.000 ± 0.001
$\mu\mu$	< 35 GeV	0.972 ± 0.006
	≥ 35 GeV	0.994 ± 0.001

non-zero. Otherwise

$$p_T^{\text{ratio}} = 1/(1 + I_{\text{rel}})$$

where I_{rel} is the relative isolation of the lepton within a cone of radius 0.4

Table 6.20: Effective areas for electrons and muons used in 2017 analysis.

$ \eta $ range	$\mathcal{A}(\mu)$	$\mathcal{A}(e)$
0.0-0.8	0.0566	0.1566
0.8-1.0	0.0562	0.1566
1.0-1.3	0.0562	0.1626
1.3-1.479	0.0363	0.1626
1.479-2.0	0.0363	0.1073
2.0-2.2	0.0119	0.0854
2.2-2.3	0.0064	0.1051
2.3-2.4	0.0064	0.1204
2.4-2.5	0.0064	0.1524

The working point of the lepton MVA for tight selection of leptons remains same at 0.90 with similar efficiency as in 2016.

6.2.4 Lepton selection

Three definitions of lepton selection are chosen similar to 2016 analysis: Tables 6.21 and 6.22 list the full criteria for 2017. The lepton selection efficiencies are obtained in a manner similar to 2016 analysis.

6.2.5 Taus

In 2017 analysis, τ_h candidates are required to have $p_T > 20$ GeV, $|\eta| < 2.3$, and a very loose working point of the tau discriminator with updated training [161]. The track quality criteria required for selected taus are: $|d_{xy}| < 1$ mm and $|d_z| < 0.2$ cm. The reconstructed hadronic taus must be separated from any muon or electron passing the loose selection by $\Delta R > 0.4$, a condition similar to 2016 analysis.

Table 6.21: Criteria for each of the three muons selection in 2017 analysis. Cuts marked with a † are applied to leptons failing the lepton MVA cut, while the loose cut value is applied to those that pass.

Cut	Loose	Fakeable object	Tight
$ \eta < 2.4$	✓	✓	✓
p_T	> 5 GeV	> 10 GeV	> 10 GeV
$ d_{xy} < 0.05$ (cm)	✓	✓	✓
$ d_z < 0.1$ (cm)	✓	✓	✓
$SIP_{3D} < 8$	✓	✓	✓
Relative $I_{\text{mini}} < 0.4$	✓	✓	✓
Loose PF muon	✓	✓	✓
p_T^{ratio}	–	$> 0.6^\dagger / -$	–
Jet CSV	–	$< 0.07^\dagger / < 0.4941$	< 0.4941
Segment compatibility	–	$> 0.3^\dagger / -$	–
Medium PF muon	–	–	✓
Lepton MVA > 0.90	–	–	✓

6.2.6 Jets and b-tagging

Due to the ECAL noise issue present in 2017 data, all jets are considered to have passed the tight identification criteria for 2017 analysis, shown in Table 6.23, which has greater than 99.5% efficiency to select jets and greater than 98% noise rejection. Jets within $|\eta| < 2.4$ are also required to have non-zero charged hadron energy fraction and charged multiplicity.

In addition, the selected jets must be separated from any fakeable lepton and hadronic tau candidates by $\Delta R > 0.4$. The loose and medium working points of the DeepCSV b-tagging algorithm, defined by discriminator threshold values 0.1522 and 0.4941 respectively, are used to identify b-jets within $|\eta| < 2.4$.

To mitigate the noise issues in the ECAL endcap and related to L1 trigger pre-firing in 2017 data, a higher p_T requirement of 60 GeV is placed on jets falling within $2.7 < |\eta| < 3.0$, detailed discussion in Appendix B. Additionally, corrections are applied to simulated

Table 6.22: Selection criteria applied to each of the three electrons in 2017 analysis. Cut values between ranges $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.479$, and $1.479 < |\eta| < 2.5$ are given in the form [value₁, value₂, value₃]. Cuts marked with a † are applied to leptons failing the lepton MVA cut, while the loose cut value is applied to those that pass.

Cut	Loose	Fakeable Object	Tight
$ \eta < 2.5$	✓	✓	✓
p_T	> 7 GeV	> 10 GeV	> 10 GeV
$ d_{xy} < 0.05$ (cm)	✓	✓	✓
$ d_z < 0.1$ (cm)	✓	✓	✓
$SIP_{3D} < 8$	✓	✓	✓
Relative $I_{\text{mini}} < 0.4$	✓	✓	✓
MVA ID $> (-0.86, -0.81, -0.72)$	✓ ¹	✓ ²	✓ ²
$\sigma_{i\eta i\eta} < (0.011, 0.011, 0.030)$	–	✓	✓
$H/E < 0.10$	–	✓	✓
$-0.04 < 1/E-1/p$	–	✓	✓
Conversion rejection	–	✓	✓
p_T^{ratio}	–	$> 0.6^\dagger / -$	–
Jet CSV	–	$< 0.07^\dagger / < 0.4941$	< 0.4941
Number of missing hits	< 2	$== 0$	$== 0$
Lepton MVA > 0.90	–	–	✓

¹ : $> (-0.13, -0.32, -0.08)$ if $p_T < 10$ GeV

² : > 0.50 if electron MVA ID > 0.90

Table 6.23: PF jet identification criteria used in 2017 analysis [117].

Variable	$ \eta < 2.7$	$2.7 < \eta < 3.0$	$ \eta > 3.0$
Neutral hadron energy fraction	< 0.90	–	> 0.02
Neutral EM energy Fraction	< 0.90	> 0.02 and < 0.99	< 0.90
Number of Constituents	> 1	–	–
Number of neutral particles	–	> 2	> 10

events to account for inefficiencies in the reconstruction of jets in the forward regions of the ECAL, caused by the L1 trigger pre-firing issue [162, 163].

6.2.7 Event selection

Common selection criteria for the 3ℓ and $2\ell ss$ final states remain same as in 2016 analysis, only change being the b-tagging discriminator:

- Invariant mass of any loose lepton pair > 12 GeV
- One or more jets passing DeepCSV medium working point
- At least one jet failing DeepCSV loose working point

The events selection are the same as those in 2016 analysis, except in 3ℓ channel the third lepton p_T threshold is relaxed to 10 GeV as it is observed to improve the efficiency. Applying the selections and performing the background estimations described earlier, the yields shown in Table 6.24 are obtained. As the amount of luminosities are comparable between 2016 and 2017, similar number of events are obtained within statistical fluctuations. Again, non-prompt and charge mis-identification are the most important backgrounds, followed by $t\bar{t}V$ processes.

6.2.8 Signal discrimination

All discriminating variables used in 2016 have been used for signal against background discrimination for 2017 data analysis as well, they are listed in Table 6.13. A preliminary study based on Receiver Operation Characteristic (ROC) integrals motivated the addition of the forward jet p_T as input variable. Similarly, a gain in discrimination power is observed when adding the resolved hadronic top-tagger [164] MVA score and sum of the p_T of all the jets selected in the event as input variables to the training of the BDT in the $2\ell ss$ region. Detailed discussion of the BDT training is documented in Appendix A. Figures 6.12, 6.13 and 6.14 show the distributions of the kinMVA variables for training of tHq against $t\bar{t}$ ($t\bar{t}V$) to the left (right), in the channels 3ℓ , $e^\pm\mu^\pm$ and $\mu^\pm\mu^\pm$ respectively.

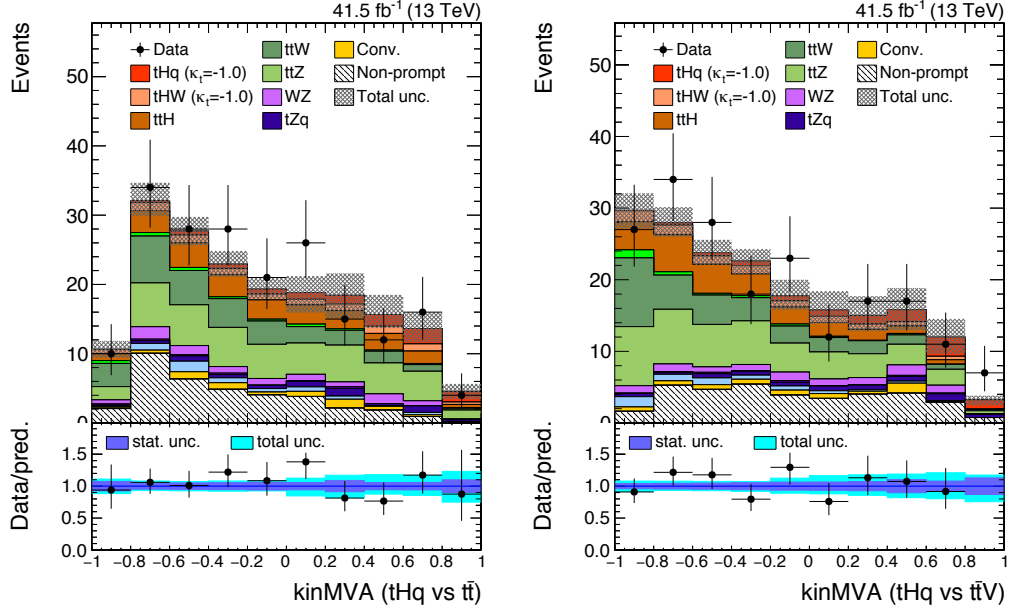


Figure 6.12: BDT output distributions in 2017 analysis for 3ℓ channel. The discriminator for tHq vs. $t\bar{t}$ ($t\bar{t}V$) is shown on the left (right). In both cases tHq events appear more towards the discriminator value 1.

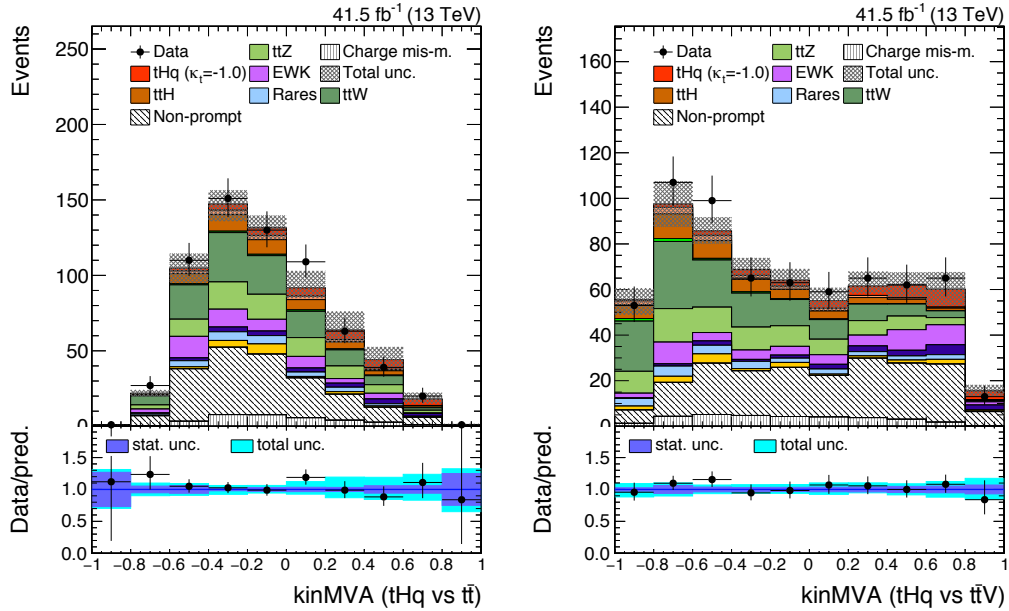


Figure 6.13: BDT output distributions in 2017 analysis for $e^\pm\mu^\pm$ channel. The discriminator for tHq vs. $t\bar{t}$ ($t\bar{t}V$) is shown on the left (right). In both cases tHq events appear more towards the discriminator value 1.

Table 6.24: Data yields and expected backgrounds after the event pre-selection for the three final states corresponding to an integrated luminosity 41.5 fb^{-1} , rounded off to nearest integral value.

Process	3ℓ	$e^\pm\mu^\pm$	$\mu^\pm\mu^\pm$
$t\bar{t}W$	30	123	80
$t\bar{t}Z$	44	77	28
$t\bar{t}WW$	3	5	3
WZ, ZZ	11	51	20
tZq	6	19	10
Rares	5	24	16
Conversion	6	16	–
Non-prompt	36	185	115
Charge flips	–	33	–
Total background	141	534	273
$t\bar{t}H$	24	44	28
tHq (SM)	1	2	1
tHW (SM)	1	1	1
Total (SM)	167	581	303
tHq (ITC)	10	34	22
tHW (ITC)	9	13	9
Total (ITC)	184	625	332
Data (41.5 fb^{-1})	194	651	354

6.2.9 Signal extraction

As in 2016 analysis, each event is now classified using its position in the 2-dimensional plane of the kinMVA variables, as presented in Fig. 6.15. Again, the $e^\pm\mu^\pm$ and $\mu^\pm\mu^\pm$ channels have been combined to decide the bin boundaries in $2\ell ss$ events since the signal to background separation is similar as shown in Figs. 6.2 and 6.3. The number of bins and bin-borders have been optimized for the new samples and training. A set of eight bins for

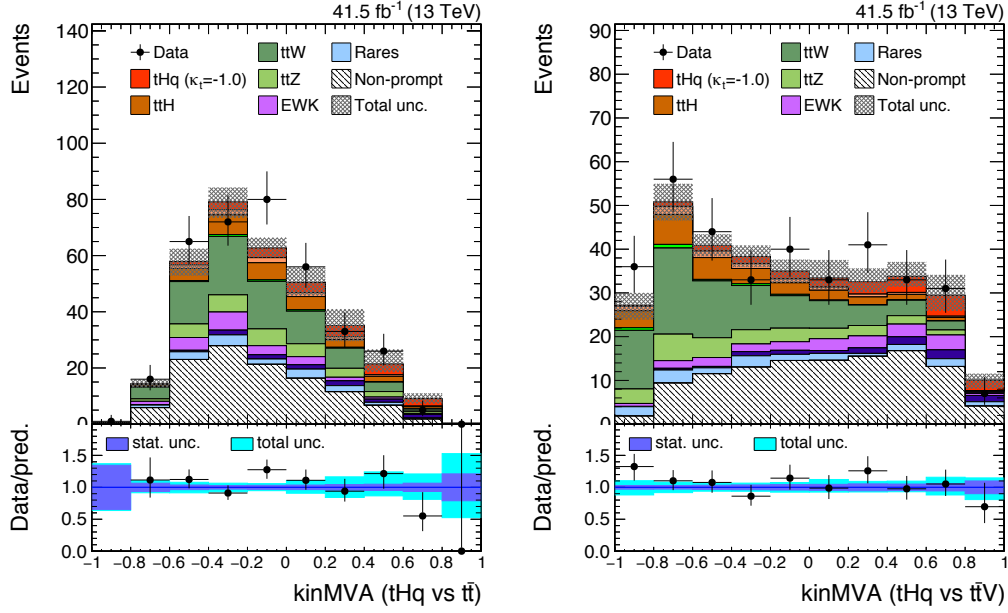


Figure 6.14: BDT output distributions in 2017 analysis for $\mu^+\mu^-$ channel. The discriminator for tHq vs. $t\bar{t}$ ($t\bar{t}V$) is shown on the left (right). In both cases tHq events appear more towards the discriminator value 1.

both 3ℓ and $2\ell ss$ channels are observed to be more sensitive in terms of the expected limit. Thus 1-dimensional histograms are obtained for each signal and background processes, and fitted to the observed data (or the Asimov dataset for expected limits).

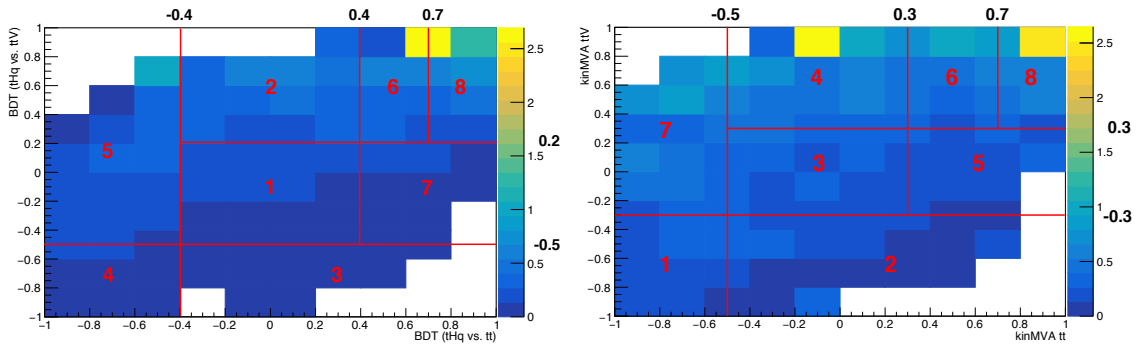


Figure 6.15: Binning overlaid on the S/B ratio map on the plane of classifier outputs for $2\ell ss$ channel (left) and 3ℓ channel (right) in 2017 analysis.

6.2.10 Systematic uncertainties

Following are the sources of uncertainty considered in this analysis. Many sources are common to the 2016 analysis, however, the numbers and shapes have been re-derived. Additional shape systematic has been added to take into account the pre-firing issue seen in 2017.

Experimental uncertainties

- Luminosity: The uncertainty in measuring luminosity in 2017 is $\pm 2.3\%$ [165], applied to the total predicted yields.
- Pileup: Similar to 2016 analysis, due to negligible effects this source of uncertainty was not considered.
- Trigger efficiency: Per-event scale factors applied depending on the flavors and p_T of the leptons as given in Table 6.19.
- Lepton identification efficiency: The uncertainties attached to the probability for electrons (muons) to pass the loose and tight selection criteria (Tables 6.21 and 6.22) are determined to be 3% for electrons with $p_T > 25$ GeV and 5% for electrons with $p_T < 25$ GeV (2% for all muons).
- b-tagging efficiency: Shape uncertainties to account for the different sources are obtained by varying the b-tagging scale factor [157].
- Jet energy scale: Shape uncertainties related to the jet energy scale as provided [117] as a function of the jet p_T and η , varying the four-momenta of the jets, as well as the b-tagging efficiency scale factor.
- Forward jet: Alternate shape of the BDT output distributions that reflects a hypothetical perfect data/MC agreement in the forward jet η is used as an uncertainty.
- L1 prefiring probability: In 2017 data, due to early L1 EGamma triggers, many uninteresting events were recorded while the subsequent interesting ones were rejected. Since this effect is not present in the MC simulation, a tool [162] provided

by the L1 detector performance group is used to reweight MC events, and derive uncertainties from 20% up/down variations (as a function of the p_T and η of jets and photons in the event). This effect is treated as a shape systematic.

Theory uncertainties

- PDF, α_s (cross section normalization): Normalization-only uncertainties are applied to the predicted yields of the $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, tHq , tHW and $t\bar{t}WW$ processes, provided along with the generated events.
- Interaction energy scale (shape and acceptance): Uncertainties on the interaction energy scale affecting the shape and acceptance of the distributions used for signal extraction are obtained using variations with μ_R and μ_F .

Background estimation

- Electroweak background: The normalization uncertainty for the diboson backgrounds in the $2\ell ss$ channel is taken to be 100% as WZ and ZZ backgrounds are combined together. In the 3ℓ channel, the corresponding statistical uncertainty is taken to be 10%, the background uncertainty 20% and the theoretical uncertainty 7%. The total normalization uncertainty for both WZ and ZZ in 3ℓ channel is considered to be 30% each.
- Rare SM backgrounds: Smaller backgrounds which have not yet been observed at the LHC are considered with a 50% normalization uncertainty.
- tZq background: The tZq cross section normalization uncertainty has been estimated to be 5% [160].
- Conversion background: The contribution of photon conversions estimated from MC is considered to have 30% normalization uncertainty.
- Charge-flips: The mis-identified electron charge background is considered with a 30% normalization uncertainty.

- Non-prompt backgrounds: Combined uncertainty of statistical fluctuations, subtraction of the prompt MC contribution etc. estimated using shape systematics. In addition, non-prompt estimation is compared against fake rate measured in QCD MC, where the ratio is fit to a constant and the resulting value is used as normalization uncertainty.

6.2.11 Results

The 1-dimensional BDT bin distribution is fitted using the Higgs combine tool [147], considering all systematic uncertainties described above. Figures 6.16, 6.17 and 6.18 show the pre-fit (without normalization adjusted) and post-fit (with normalization adjusted) distributions of the BDT bins in channels 3ℓ , $e^\pm\mu^\pm$ and $\mu^\pm\mu^\pm$ respectively. The best-

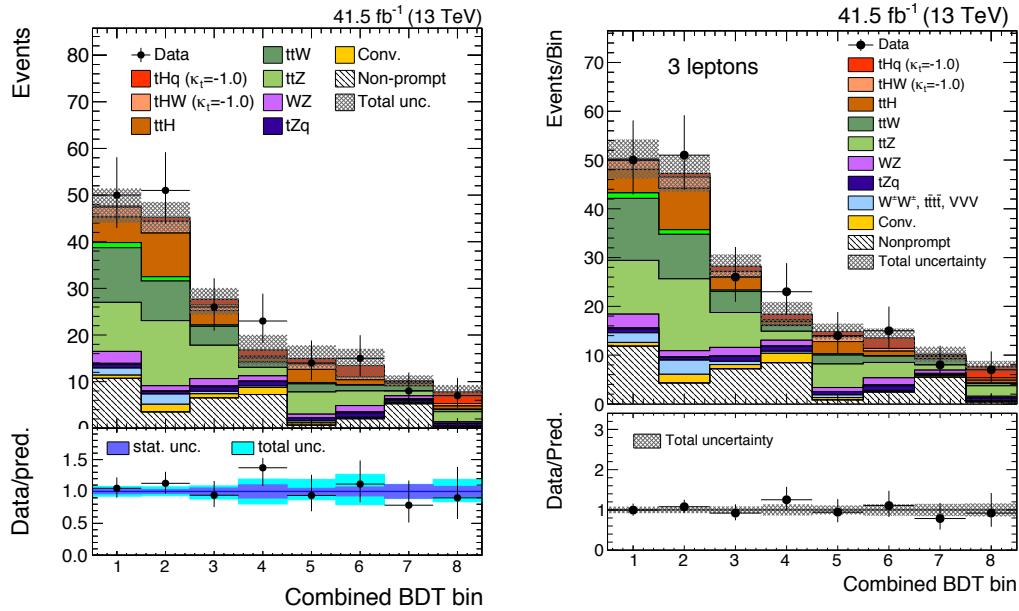


Figure 6.16: Pre-fit (left) and post-fit (right) BDT bin distributions in 2017 analysis for 3ℓ final state where data is plotted along with simulation. Here both pre-fit and post-fit figures show the tH signal in the ITC scenario against SM backgrounds.

fit $tH+t\bar{t}H$ combined signal strength observed for the inverted top-Yukawa coupling is $0.91^{+0.47}_{-0.43}$ with observed significance of 2.5σ (2.8σ expected). For the SM hypothesis, the best-fit signal strength is $0.91^{+0.47}_{-0.43}$ with observed significance of 2.5σ (2.8σ expected).

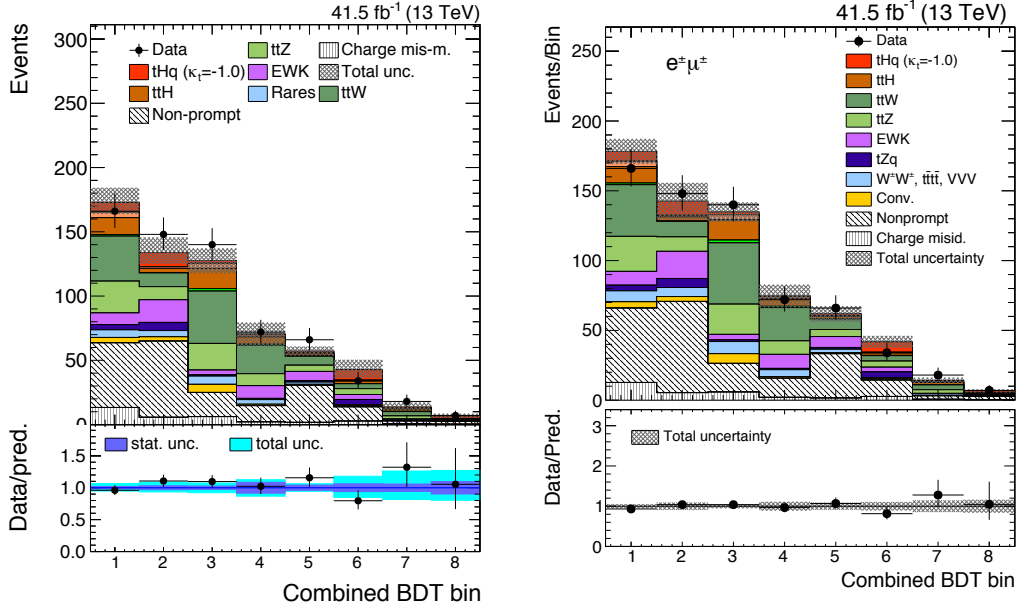


Figure 6.17: Pre-fit (left) and post-fit (right) BDT bin distributions in 2017 analysis for $2\ell_{ss} e^\pm \mu^\pm$ final state where data is plotted along with simulation. Here both pre-fit and post-fit figures show the tH signal in the ITC scenario against SM backgrounds.

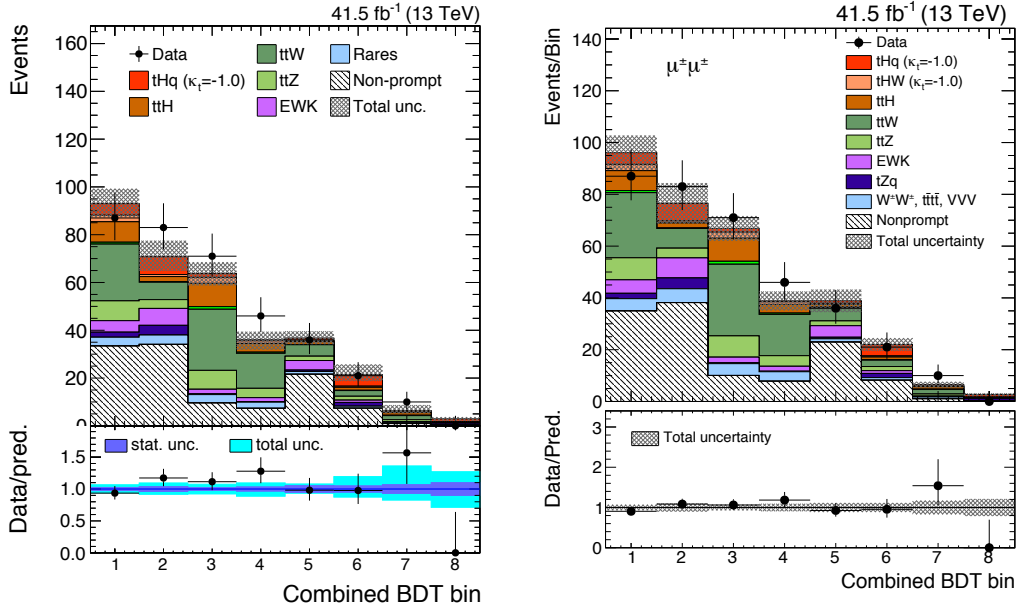


Figure 6.18: Pre-fit (left) and post-fit (right) BDT bin distributions in 2017 analysis for $2\ell_{ss} \mu^\pm \mu^\pm$ final state where data is plotted along with simulation. Here both pre-fit and post-fit figures show the tH signal in the ITC scenario against SM backgrounds.

The expected and observed upper limits on the cross section times branching ratio as a function of the κ_t are also shown in Fig. 6.19. Compared to Fig. 6.11, the expected uncertainty band is much wider, despite a larger integrated luminosity. As the momentum threshold on forward jets is loosened in the 2017 analysis, except in the small interval between 2.7–3.0 in pseudorapidity, more number of jets are accepted in the final state. This effect in conjunction with a more conservative prescription of jet energy scale uncertainty may explain the wider standard deviations of the expected limits. In fact, the dominant systematic uncertainty in 2017 analysis is determined to be due to the variation in jet energy scale, whereas it was non-prompt leptonic background estimation in 2016 analysis.

The limits as a function of κ_t are obtained with a tH only, resolved Higgs couplings hypothesis, where $t\bar{t}H$ is included in the background assuming a κ_t dependent cross section with $\mu_{t\bar{t}H} = 1.0$. The observed (expected) upper limit on tH only signal strength μ_{tH} at 95% CL is **1.56 (1.25)** for the ITC scenario and **28.09 (22.56)** for the SM.

The 95% CL uncertainties of a parameter of interest, here κ_t , is determined using the profile likelihood ratio as the test statistic (see Section 5.2.3). This is a more model-independent approach as compared to the one taken in 2016 analysis, where additional assumptions on other fermion couplings had to be made. A scan of the Likelihood ratio $\mathcal{L} = \mathcal{L}(\kappa_t)/\mathcal{L}(\hat{\kappa}_t)$ is performed as a function of κ_t , where the best-fit value is $\hat{\kappa}_t$. Constraints on κ_t is derived from a distribution of the test statistic $-2\Delta \ln \mathcal{L} = -2[\ln \mathcal{L}(\kappa_t) - \mathcal{L}(\hat{\kappa}_t)]$. The 95% CL allowed ranges of κ_t is determined to be within **(-1.4, -0.7)** and **(0.7, 1.95)**, see Fig. 6.19.

In the same way, a Likelihood scan against the CP phase angle is obtained as shown in Fig. 6.20. The data is able to exclude the completely CP odd scenario ($\cos(\alpha) = 0$) at 95% CL in this analysis using 41.5 fb^{-1} of data. It is expected that with the addition of more data and analysis of other Higgs boson decay modes, extended range of CP-violating scenarios will be excluded. Figure 6.20 also shows the expected and observed upper limits on the combined cross section times branching ratio of $tH+t\bar{t}H$ as a function of $\cos(\alpha)$, assuming a background only (no Higgs) hypothesis. The consistent excess is again attributed to the presence of $t\bar{t}H$ events in data.

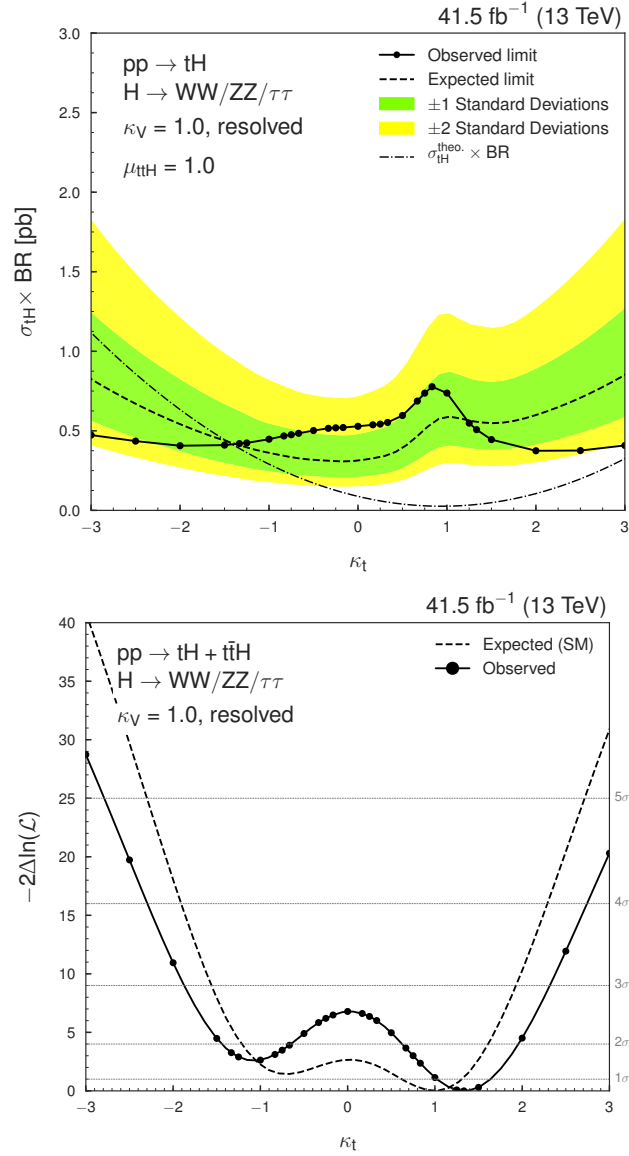


Figure 6.19: Upper limits on cross section times branching ratio of $tH, H \rightarrow WW^*/ZZ^*/\tau\tau$ signal as a function of κ_t is shown on top. Here the background only hypothesis considers $t\bar{t}H$ production with a fixed signal strength $\mu_{t\bar{t}H} = 1.0$. The upper limits on μ_{tH} has a maxima near $\kappa_t = 1.0$. Slight excess of events is observed compared to 2016 analysis, likely due to statistical fluctuation. Bottom plot shows likelihood scan as a function of κ_t for the resolved couplings model considering tH and $t\bar{t}H$ as signal. The maximum likelihood is obtained for the value $\kappa_t = 1.3$ where the minima of the curve occurs. Note that data is much more sensitive to rule out the scenario $\kappa_t = 0$ due to the presence of $t\bar{t}H$ events.

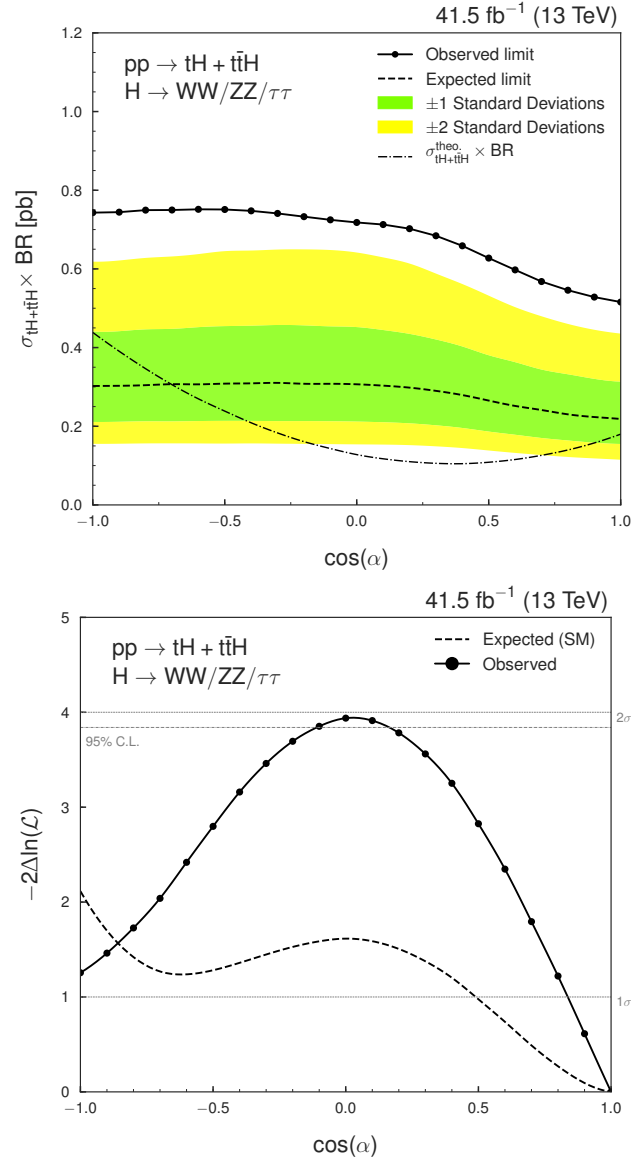


Figure 6.20: Upper limits on cross section times branching ratio of $tH+t\bar{t}H$, $H \rightarrow WW^*/ZZ^*/\tau\tau$ signal as a function of $\cos(\alpha)$ is shown on top. The excess of events observed is due to the presence of $t\bar{t}H$ events, considered as signal in this background-only interpretation. Bottom plot shows likelihood scan against the cosine of CP phase angle. Here the data is much more sensitive to rule out the completely CP-odd scenario at $\cos(\alpha) = 0$, compared to the expectation.

6.3 Combination of 2016 and 2017 results

The analyses described in earlier sections on tH multi-lepton final state for 2016 and 2017 run periods are combined to extract results from total integrated luminosity of 77.4 fb^{-1} . In this combination all experimental uncertainties are considered to be uncorrelated, except that due to forward jet modelling. The jet energy scale and b-tagging uncertainties etc. change between years depending on the used identification and selection criteria. As the forward jet modelling uncertainties follow the exact same procedure between 2016 and 2017, they are conservatively considered as fully correlated. Variations of PDF, α_s , interaction scale energy and other normalization uncertainties in background yield estimation that remain the same in both years are considered to be fully correlated. In the combined analysis the dominant uncertainty is determined to be due to the variation in the jet energy scale in the year 2017, followed by that due to fake rate determination of muons in the same year. Figure 6.21 shows the observed and expected upper limits on tH only signal strength and the likelihood scan as a function of κ_t . The observed limits are closer to the expected values in the κ_t range between -1.0 and 1.0 similar to the 2016 results; they are much more stringent at higher values of $|\kappa_t|$ as seen in 2017 analysis.

The observed (expected) upper limit on tH only signal strength μ_{tH} at 95% CL is **0.79 (0.72)** for the ITC scenario and **13.78 (12.69)** for the SM. If 2018 data corresponding to an integrated luminosity of $\sim 60 \text{ fb}^{-1}$ is included the results expected from complete Run 2 data can be derived. Thus, with full Run 2 luminosity of 137 fb^{-1} , the expected 95% CL median upper limit on tH only signal strength is **0.58** for ITC scenario and **10.38** for SM based on a luminosity projection, keeping all systematic uncertainties the same. Obtained limits on the tH only signal strength (μ_{tH}) is shown in Table 6.25 for the years 2016, 2017, their combination and projection to Run 2 luminosity. The corresponding expected upper limits on $\sigma \times BR$ for the tH process are summarized in Table 6.26.

From the likelihood scan with $tH+t\bar{t}H$ as signal, $\kappa_t = -1$ for the resolved Higgs couplings can be excluded at 95% CL with the combined dataset of 2016 and 2017. The allowed ranges of κ_t is constrained within **(0.85, 1.85)**. This is much more stringent than the derived ranges from analysis of individual years data.

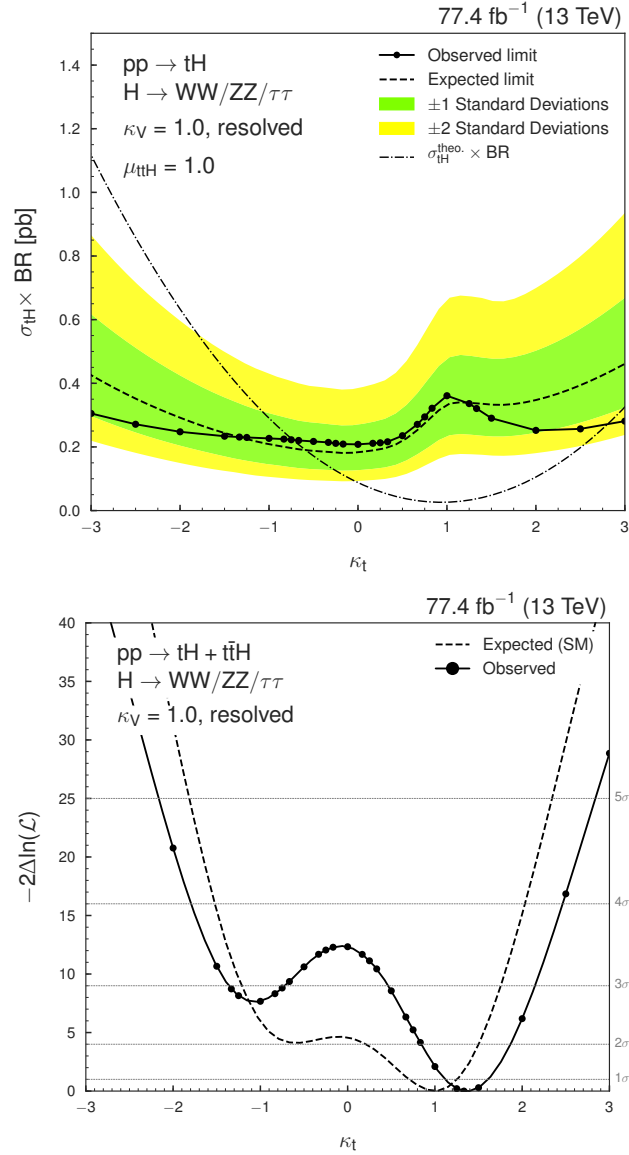


Figure 6.21: Combined (2016+2017) upper limits on cross section times branching ratio of tH signal as a function of κ_t is shown on top. The observed values are much closer to the median expected values with more amount of data. Here also the maxima of μ_{tH} is observed near $\kappa_t = 1.0$. Bottom plot shows likelihood scan as a function of κ_t for the resolved Higgs couplings hypothesis assuming $\kappa_V = 1.0$. The allowed ranges in the positive range have become narrower and the maximum likelihood occurs at $\kappa_t = 1.4$, negative ranges are ruled out at 95% CL.

Table 6.25: 95% CL observed (expected) upper limits on tH signal strength with $H \rightarrow WW^*/ZZ^*/\tau\tau$ obtained from multi-lepton channel with the SM and ITC scenarios. The 2016 data shows more stringent upper limit on tH production compared to 2017, this may be result of statistical fluctuation in the data since the luminosities are comparable. 2018 data, corresponding to integrated luminosity of $\sim 60 \text{ fb}^{-1}$ has been considered in Run2 projection.

Data	Upper limit on ITC	Upper limit on SM
2016 (35.9 fb^{-1})	1.02 (0.92)	17.72 (15.31)
2017 (41.5 fb^{-1})	1.56 (1.25)	28.09 (22.56)
Combined (77.4 fb^{-1})	0.79 (0.72)	13.78 (12.69)
Run 2 projection (137 fb^{-1})	(0.58)	(10.38)

Table 6.26: 95% CL observed (expected) upper limits on tH production cross section with $H \rightarrow WW^*/ZZ^*/\tau\tau$ obtained from multi-lepton channel with the SM and ITC scenarios in the 2016, 2017 analyses and their combination. 2018 data, corresponding to integrated luminosity of $\sim 60 \text{ fb}^{-1}$ has been considered in Run2 projection.

	ITC scenario	SM scenario
Theoretical $\sigma \times \text{BR}$	0.29 pb	0.03 pb
Data	Upper limit on ITC $\sigma \times \text{BR}$	Upper limit on SM $\sigma \times \text{BR}$
2016 (35.9 fb^{-1})	0.30 (0.27) pb	0.46 (0.40) pb
2017 (41.5 fb^{-1})	0.45 (0.36) pb	0.74 (0.59) pb
Combined (77.4 fb^{-1})	0.23 (0.21) pb	0.36 (0.33) pb
Run 2 projection (137 fb^{-1})	(0.17) pb	(0.27) pb

Table 6.27 summarizes the different observations in the combined $tH+t\bar{t}H$ multi-lepton final state using 2016 and 2017 analyses and their combination along with the expected results for full Run 2 luminosity. The simple luminosity projection for the likelihood scan is not included since the expected behavior is not as sensitive as data.

Table 6.27: 95% CL observed (expected) upper limits on $\sigma \times BR$ of combined $tH+t\bar{t}H$ production with $H \rightarrow WW^*/ZZ^*/\tau\tau$ decay in SM and ITC scenarios. Third column shows the observed 95% CL allowed range for κ_t values. Note, for 2016 analysis, the κ_t range has been obtained comparing the observed $\sigma \times BR$ with the theoretical values.

	ITC scenario	SM scenario	
Theoretical $\sigma \times BR$	0.44 pb	0.18 pb	
Data	Upper limit on ITC $\sigma \times BR$	Upper limit on SM $\sigma \times BR$	Allowed ranges of κ_t
2016 (35.9 fb ⁻¹)	0.64 (0.32) pb	0.56 (0.24) pb	(-1.25, 1.6)
2017 (41.5 fb ⁻¹)	0.74 (0.31) pb	0.52 (0.22) pb	(-1.4, -0.7) & (0.7, 1.95)
Combined (77.4 fb ⁻¹)	0.58 (0.23) pb	0.46 (0.18) pb	(0.85, 1.85)
Run 2 projection (137 fb ⁻¹)	(0.19) pb	(0.14) pb	–

For the ITC scenario, the best-fit $tH+t\bar{t}H$ signal strength is $0.72^{+0.35}_{-0.34}$ with observed significance of 2.1σ (3.5σ expected). The best-fit $tH+t\bar{t}H$ combined signal strength observed for the SM hypothesis is $1.75^{+0.56}_{0.52}$ corresponding to an observed significance of 3.4σ (2.1σ expected).

The results from the 2016 multi-lepton analysis have been combined with results from $H \rightarrow b\bar{b}$ and $H \rightarrow \gamma\gamma$ decay modes that use the same luminosity. In addition they have been projected to HL-LHC luminosity in order to explore the sensitivity with large amount of data, as discussed in next chapter.

Chapter 7

Combination of tH final states and future prospects

This chapter describes the combined results from the dedicated analyses of tH final states in $H \rightarrow WW^*/ZZ^*/\tau\tau$ channels using leptonic final states and $H \rightarrow b\bar{b}$ channels using two identified b-jets alongwith the reinterpretation of the $H \rightarrow \gamma\gamma$ analysis [166]. Dedicated analysis has been performed with the tH final state where $H \rightarrow b\bar{b}$ using 2016 data [167]. Further, the inclusive $H \rightarrow \gamma\gamma$ final state analysis for measuring Higgs properties using 2016 data [168] includes the SM tHq and tHW processes. Results from leptonic final states using 2016 and 2017 data has been presented in previous chapter.

Similar to the multi-lepton analysis, $H \rightarrow b\bar{b}$ final state exploits the sensitivity of the tHq production cross section depending on the ratio κ_t/κ_V . A BDT classifier is trained to discriminate against the dominant $t\bar{t}$ +heavy flavor (HF) jets background and the output distributions are used as input to the maximum likelihood fit. The reinterpretation of the di-photon analysis consists of correcting the event yields to tHq and tHW cross sections with scale factors that take into account the changed selection efficiency and acceptance as a function of κ_t/κ_V . Since the κ_t and κ_V interference is involved in $H \rightarrow \gamma\gamma$ decay, it adds additional sensitivity to the analysis in the resolved Higgs couplings hypothesis.

Limits are then set on the tH only and combined $tH+t\bar{t}H$ production cross section times BR for the decay modes $H \rightarrow WW^*/ZZ^*/\tau\tau/b\bar{b}/\gamma\gamma$ for different coupling scenarios. A negative log-likelihood scan is used to extract the allowed ranges of the κ_t for $\kappa_V=1.0$.

7.1 Results from combination based on 2016 data

Following results are based on 35.9 fb^{-1} data collected in 2016 [135], which greatly improves over Run 1 result where 95% CL observed (expected) upper limit was set between 0.60 (0.45) pb and 1.00 (0.70) pb on tHq production in ITC scenario alone using 19.7 fb^{-1} data at 8 TeV, depending on the assumed di-photon branching fraction of the Higgs boson [69]. Figure 7.1 shows the resulting limit on tH only cross section times branching ratio. The expected limit is calculated on a background-only data set, i.e., without tH contribution, but including a κ_t -dependent contribution from the $t\bar{t}H$ production. The $t\bar{t}H$ normalization is kept fixed in the fit, while the tH signal strength is allowed to float.

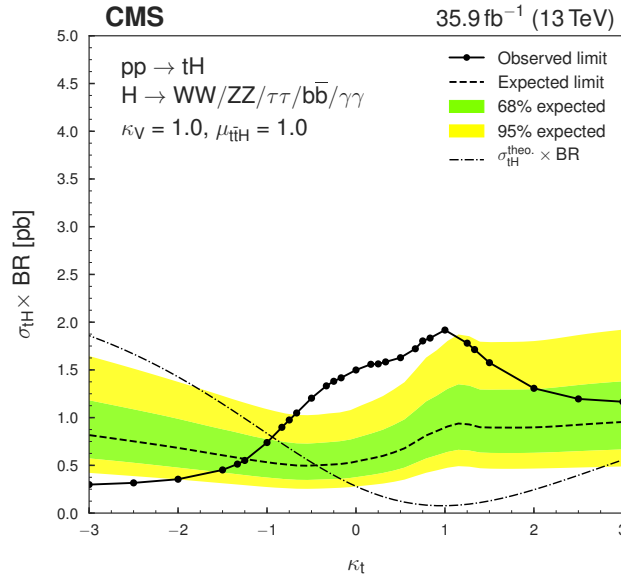


Figure 7.1: Observed and background-only expected cross section times branching ratio limits on combined tH production where $H \rightarrow WW^*/ZZ^*/\tau\tau/b\bar{b}/\gamma\gamma$, as a function of κ_t , for $\kappa_V = 1.0$. An excess of events is observed for the SM scenario $\kappa_t = 1.0$.

The observed (expected) upper limit at 95% CL is **0.89 (0.64)** in the ITC scenario whereas in case of SM it is **25.2 (12.0)** considering the tH only signal strength. Table 7.1 shows the corresponding upper limits on the cross section times branching fraction for the decays $H \rightarrow WW^*/ZZ^*/\tau\tau/b\bar{b}/\gamma\gamma$. The best-fit signal strength considering a combined $tH+t\bar{t}H$ signal and the significance of the measured values are also listed. Table 7.2 presents the

limits for the SM and ITC scenario for the individual channels.

Table 7.1: 95% CL observed (expected) upper limits on tH only $\sigma \times BR$ and best-fit signal strength along with the significance for $tH+t\bar{t}H$ signal.

Scenario	Upper limit on tH only	Best-fit $tH+t\bar{t}H$	Significance
	$\sigma \times BR$	signal strength	
ITC	0.74 (0.53) pb	$0.60^{+0.21}_{-0.18}$	3.9σ (5.9 σ)
SM	1.94 (0.92) pb	$2.00^{+0.54}_{-0.51}$	4.3σ (2.2 σ)

Table 7.2: 95% CL observed (expected) upper limits on the tH production cross section times $H \rightarrow WW/ZZ/\tau\tau/b\bar{b}/\gamma\gamma$ branching fraction for ITC scenario ($\kappa_t=-1.0$, top rows) and for an SM-like signal ($\kappa_t=1.0$, bottom rows), in pb. Limits can be compared to the expected product of tH cross sections and branching fractions of 0.83 and 0.077 pb for the inverted top-Yukawa coupling and for the SM, respectively.

Data	Upper limit on ITC	Upper limit on SM
	$\sigma \times BR$	$\sigma \times BR$
$b\bar{b}$	4.98 ($2.52^{+1.29}_{-0.81}$) pb	6.88 ($3.19^{+1.64}_{-1.02}$) pb
$\gamma\gamma$	0.84 ($0.88^{+0.46}_{-0.28}$) pb	3.68 ($2.03^{+1.05}_{-0.67}$) pb
$WW^*/ZZ^*/\tau\tau$	0.85 ($0.77^{+0.36}_{-0.24}$) pb	1.36 ($1.18^{+0.53}_{-0.35}$) pb
Combined	0.74 ($0.53^{+0.24}_{-0.16}$) pb	1.94 ($0.92^{+0.40}_{-0.27}$) pb

An excess of events compared to expected backgrounds is observed, however it is still compatible with the SM $tH+t\bar{t}H$ productions. Table 7.3 shows the impact of the most important groups of nuisance parameters on the $tH+t\bar{t}H$ signal yield along with pre-fit systematic uncertainties of the same groups, shown for comparison.

Combined likelihood as a function of κ_t for the resolved Higgs couplings is shown in Fig. 7.2. The data favors $\kappa_t = 1.0$ over $\kappa_t = -1.0$ by about 1.5σ while excluding values outside $(-0.9, -0.5)$ and $(1.0, 2.1)$ at 95% CL. The sensitivity is driven by the di-photon channel at negative values of κ_t and by the multi-lepton channel at positive values.

Table 7.3: Summary of the main sources of systematic uncertainty. $\Delta\mu/\mu$ corresponds to the relative change in $tH+t\bar{t}H$ signal yield induced by varying the systematic source within its associated uncertainty.

Source	Uncertainty [%]	$\Delta\mu/\mu$ [%]
e, μ selection efficiency	2–4	17
b-tagging efficiency	2–15	6
Jet energy calibration	2–15	3
Forward jet modeling	10–35	3
Integrated luminosity	2.5	10
Reducible background estimate	10–40	14
Theoretical sources	≈ 10	14
$t\bar{t}$ +HF normalization	≈ 50	7
PDFs	2–6	8

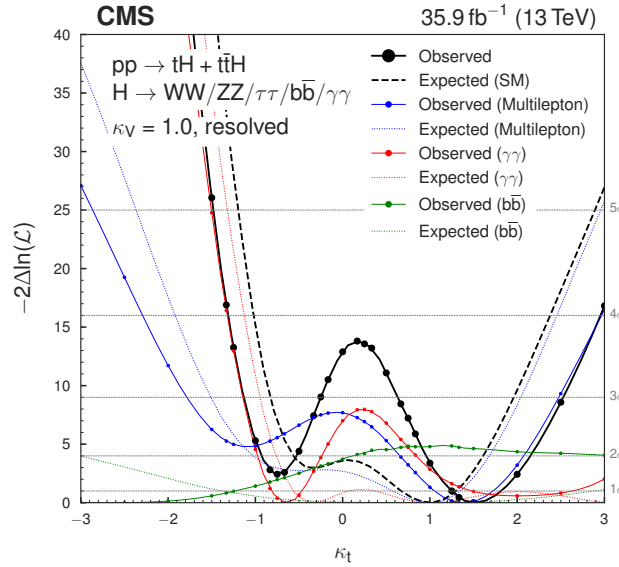


Figure 7.2: Scan of $-2\Delta\ln\mathcal{L}$ vs κ_t for the data (black line) and the individual channels (blue, red, and green), compared to the SM expectations (dashed lines).

From the above studies and interpretations it is inferred that observation of only tH pro-

duction process is very unlikely in case of SM scenario. The trend also shows that ITC scenario is becoming less probable with more data. However, the prospect of observation gets brighter in next few years time, as presented in next section.

7.2 Future prospects of tH search

From 2026 onwards, the LHC machine is expected to operate in a higher luminosity mode than originally planned two decades ago, where the instantaneous luminosity is expected to reach up to $7.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. This phase of the machine, after considerable upgrade of both the LHC and the detectors, is to be named as HL-LHC and is expected to last for ten years till 2036. The proposed timeline is shown schematically in Fig. 7.3.

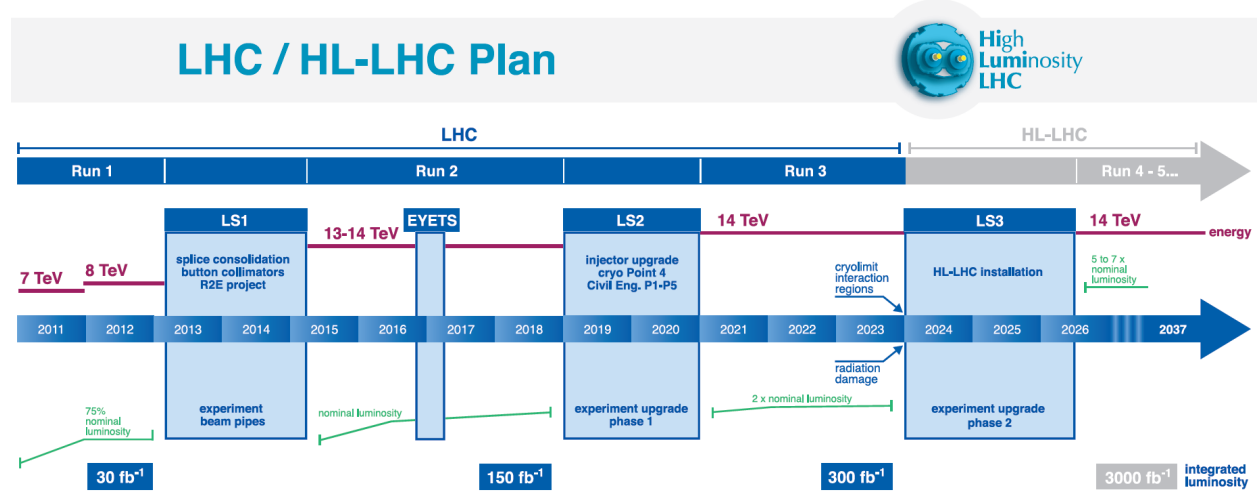


Figure 7.3: Timeline for the LHC and HL-LHC data taking and upgrade which shows the planned luminosity and energy within each run period [169].

A large amount of data at least 3000 fb^{-1} will be collected during ten years of run time during 2016-36, paving experimental access to rare processes. However, high luminosity operation will also result in severe detector degradation in performance and high pileup rates, about 140 – 200 per event, on average. Detector upgrades are therefore planned to mitigate the anticipated worsening of physics performance of the CMS experiment. The possible improvements in future results have been estimated during 2018 using simple

projection scenarios with the help of available statistical methods, in the absence of reliable detector simulations, in the context of HL-LHC workshop. The studies have been documented in CERN Yellow Report [51] submitted to European Strategy group as inputs for decision regarding further high energy physics facilities.

7.2.1 Projection scenarios

In order to estimate the physics potential of the CMS detector by the end of LHC Run 3 and at the HL-LHC, projections are presented in different scenarios for the evolution of systematic uncertainties with increased data samples and improved theoretical predictions. The baseline scenarios assume that the CMS upgrades will provide the same level of detector and trigger performance as in the Run 2 data taking period [170–174]. Uncertainties due to the limited number of simulated events in the current analyses are neglected, under the assumption that sufficiently large samples of events will be available in future analyses. The two scenarios evaluated are:

- **“Run 2 systematic uncertainties” scenario (S1):** All systematic uncertainties are kept constant with integrated luminosity. The performance of the CMS detector is assumed to be unchanged with respect to the reference analysis;
- **“YR18 systematic uncertainties” scenario (S2):** Theoretical uncertainties are scaled down by a factor of two, while experimental systematic uncertainties are scaled down with the square root of the integrated luminosity until they reach a defined minimum value based on estimates of the achievable accuracy with the upgraded detector.

Table 7.4 summarises the Run 2 uncertainties for which a minimum value is set in S2. Systematic uncertainties in the identification and isolation efficiencies for electrons and muons are expected to be reduced to approximately 0.5%. The τ_h identification uncertainty is assumed to be reduced to approximately 2.5%. The uncertainty in the overall JES is expected to reach approximately 1% precision for jets with $p_T > 30$ GeV, driven primarily by improvements in the absolute scale and jet flavor calibrations. The p_T^{miss} uncertainty is obtained by propagating the JES uncertainties in its computation, yielding a

reduction by up to a half of the Run 2 uncertainty. For the identification of b-tagged jets the uncertainty in the selection efficiency of b (c) quarks, and in misidentifying a light jet is expected to remain similar to the current level, with only the statistical component reducing with increasing integrated luminosity. The uncertainty in the integrated luminosity of the data sample could be reduced down to 1% thanks to a better understanding of the calibration and fit models employed in its determination, and making use of the finer granularity and improved electronics of the upgraded detectors.

Table 7.4: The sources of systematic uncertainty for which minimum values are applied in S2 [175].

Source	Component	Run 2 uncertainty	Projection minimum uncertainty
Muon ID		1–2%	0.5%
Electron ID		1–2%	0.5%
Photon ID		0.5–2%	0.25–1%
Hadronic tau ID		6%	2.5%
Jet energy scale	Absolute	0.5%	0.1–0.2%
	Relative	0.1–3%	0.1–0.5%
	Pileup	0–2%	Same as Run 2
	Method and sample	0.5–5%	No limit
	Jet flavour	1.5%	0.75%
	Time stability	0.2%	No limit
Jet energy res.		Varies with p_T and η	Half of Run 2
p_T^{miss} scale		Varies with analysis selection	Half of Run 2
b-Tagging	b-/c-jets (syst.)	Varies with p_T and η	Same as Run 2
	light mis-tag (syst.)	Varies with p_T and η	Same as Run 2
	b-/c-jets (stat.)	Varies with p_T and η	No limit
	light mis-tag (stat.)	Varies with p_T and η	No limit
L_{tot}		2.5%	1%

Theoretical uncertainties follow the prescriptions of the LHC Yellow Report 4 [45] in S1 and are halved in S2 to account for future theoretical developments. In both scenarios the intrinsic statistical uncertainty on any measurement scales with $1/\sqrt{R_L}$, where R_L is the

projected integrated luminosity divided by that of the Run 2 value: $R_L = L_{\text{projection}}/L_{\text{Run 2}}$.

7.2.2 Results

To obtain the projected constraints on κ_t , a scan of the test statistic $q(\kappa_t)$ is performed, assuming $\kappa_V = 1.0$ and using an Asimov dataset corresponding to the SM expectation ($\kappa_t = \kappa_V = 1.0$). The expected contribution from the $t\bar{t}H$ process is also modified consistently with κ_t . Figure 7.4 gives the likelihood scan for the S1 and S2 projections, which shows that at 300 fb^{-1} negative values of κ_t are excluded at more than the 99% CL in the S1 scenario. The κ_t ranges are even more constrained in S2. For 3000 fb^{-1} a negative value of κ_t is disfavored with a significance larger than five standard deviations.

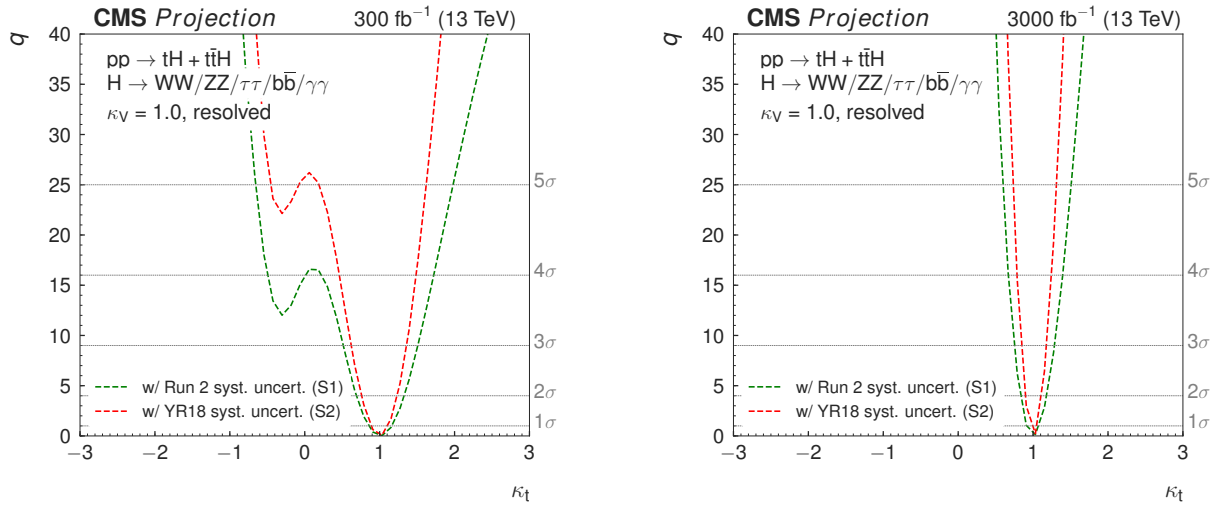


Figure 7.4: Scan of the test statistic q versus κ_t for the Asimov datasets corresponding to $\kappa_V=1.0$ with the two integrated luminosity scenarios in S1 (with Run 2 systematic uncertainties [135]) and S2 (with YR18 systematic uncertainties).

With the larger HL-LHC dataset it is therefore relevant to assess the sensitivity to the SM rate of tH production. In the Run 2 combination analysis with $L = 35.9 \text{ fb}^{-1}$, the observed upper limit on the tH signal strength relative to the SM (μ_{tH}) is 25 compared to the median expected value of 12, assuming the $t\bar{t}H$ signal strength ($\mu_{t\bar{t}H}$) is fixed to the SM value.

In Fig. 7.5 the variation of the expected upper limits on μ_{tH} is shown as a function of the integrated luminosity for the S1 and S2 scenarios. The limits are determined assuming a background-only hypothesis in which the $t\bar{t}H$ process is considered as following the SM expectation ($\mu_{t\bar{t}H} = 1$). In order to minimize further assumptions on the rate of $t\bar{t}H$ production, $\mu_{t\bar{t}H}$ is treated as a free parameter in the fit. In the S1 scenario the expected median upper limit on μ_{tH} at 300 (3000) fb^{-1} is determined to be 5.60 (2.35). The corresponding value in S2 at 300 (3000) fb^{-1} is 3.98 (1.51). With the 3000 fb^{-1} dataset and foreseen reduction in systematic uncertainties in S2, the expected upper limit on μ_{tH} improves by about a factor of eight with respect to the current values.

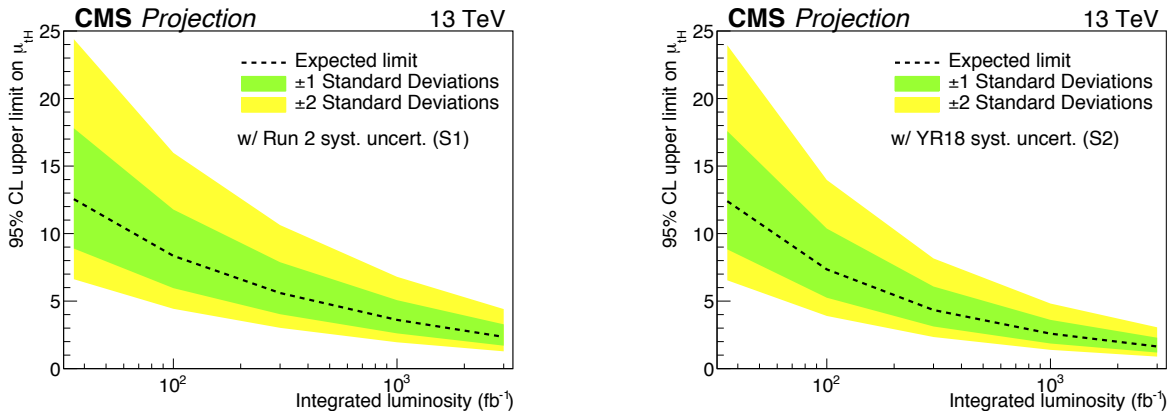


Figure 7.5: The variation of expected upper limit on μ_{tH} with integrated luminosity for two projection scenarios S1 (with Run 2 systematic uncertainties [135]) and S2 (with YR18 systematic uncertainties).

The evolution of the expected uncertainty on the measurement of μ_{tH} , assuming the SM rate, is given in Table 7.5. Values are given for two cases of background: one in which $\mu_{t\bar{t}H}$ is unconstrained in the fit, and one in which it is fixed to the SM value of 1. In the latter case the uncertainties are reduced by around 10% at 3000 fb^{-1} , indicating that a precise simultaneous measurement of the $t\bar{t}H$ signal strength will be needed to obtain the optimal sensitivity to the tH channel. In both cases it is found that the reduced systematic uncertainties in S2 improve the precision by up to 30%.

Table 7.5: The $\pm 1\sigma$ uncertainties on expected $\mu_{tH}=1$ for scenarios S1 (with Run 2 systematic uncertainties [135]) and S2 (with YR18 systematic uncertainties) at all three luminosities, considering also the case when $\mu_{t\bar{t}H}$ is fixed at the SM value 1.

	Scenario	$\mu_{t\bar{t}H}$ floating	$\mu_{t\bar{t}H}$ fixed
S1	35.9 fb ⁻¹	+6.2 -5.8	+5.8 -5.4
	300 fb ⁻¹	+2.9 -2.8	+2.5 -2.4
	3000 fb ⁻¹	+1.2 -1.2	+1.1 -1.0
S2	35.9 fb ⁻¹	+6.2 -5.8	+5.8 -5.3
	300 fb ⁻¹	+2.2 -2.2	+2.0 -2.0
	3000 fb ⁻¹	+0.9 -0.9	+0.8 -0.8

7.2.3 Significance of measurement

The expected significance for measuring $\mu_{tH} = 1.0$ at $L = 3000 \text{ fb}^{-1}$ in the S2 scenario is 1.3σ . A combination with a similar result from ATLAS would lead to a significance of 1.8σ . However, currently for Run 2 a dedicated tH analysis in the $H \rightarrow \gamma\gamma$ channel is not yet performed, which used to be the most sensitive channel in Run 1 [69]. The re-interpretation of the $H \rightarrow \gamma\gamma$ final state analysis is sub-optimal for observing the tHq production process. The Run 1 analysis was performed for the inverted top coupling scenario $\kappa_t = -1.0$, where for 19.7 fb^{-1} of data, the MC yield was 0.67 for tHq . If we consider the scaling of the tHq cross section from $\sqrt{s} = 8 \text{ TeV}$ to 13 TeV and also from $\kappa_t = -1.0$ to 1.0 , expected MC yield at $L = 3000 \text{ fb}^{-1}$ is 31 events. The number of background events extrapolated to 3000 fb^{-1} is 164. Assuming signal and background ratio improves by a factor of half, and further expecting that the number of events in the side-band region reduces by a factor of half by the HL-LHC era, number of background events reduces to 41. This leads to an optimistic estimation of significance of 2.8σ for observing the tH production process with the data collected by HL-LHC.

In the next chapter the findings from the tH analysis with Run 2 data have been summarized with future prospects discussed above.

Chapter 8

Conclusions and outlook

After the discovery of the Higgs boson, the particle physics community has shifted focus to precision measurements to characterize the resonance of its properties. Accurate determination of the Higgs boson properties can become a window to BSM physics by challenging the predicted values of the standard model. So far, no significant deviation from the SM has been observed within the expected uncertainties. In this thesis, a measurement of the top-Yukawa coupling has been presented using the tH production process in the multi-lepton final state where the Higgs boson decays to $WW^*/ZZ^*/\tau\tau$ and the top quark decays semi-leptonically. The search for the tH process is performed using 77.4 fb^{-1} p-p collision data collected by the CMS experiment in the years 2016 and 2017 at $\sqrt{s} = 13 \text{ TeV}$. This production is particularly sensitive to BSM phenomena that can alter the relative sign between the top-Higgs and W -Higgs couplings as predicted by the SM. A model independent study is carried out for probing different magnitudes and opposite signs of the top-Higgs coupling modifier κ_t . In addition, the CP properties of the top-Higgs interaction have also been studied in a similar way using the tH final state. To increase sensitivity, the multi-lepton final state is divided into two channels: three leptons and two same-sign leptons. The lepton selection and background estimation methods follow the $t\bar{t}H$ multi-lepton analyses [137, 138]. To extract the signal from the background, multivariate techniques have been used over a loose baseline selection of the final states. The t-channel production mode of tH has the unique kinematic property of a forward light quark jet. This feature is exploited to better discriminate against the

major background processes $t\bar{t}$ and $t\bar{t}V$. A combination of two boosted decision tree discriminators is used to obtain a final binned distribution of the event yields, used in the maximum likelihood fit. The combined dataset of 77.4 fb^{-1} collected by CMS experiment during 2016 and 2017 yields the following observed upper limit on the production cross section times branching ratio of the SM production rate of tH process:

$$\sigma(tH) \times BR(H \rightarrow WW^*/ZZ^*/\tau\tau) = 0.36 \text{ pb}$$

to be compared to the theoretical prediction of 0.03 pb. The inverted top coupling scenario is excluded with 95% CL using a likelihood scan of κ_t values. The allowed ranges of κ_t are within about (0.85, 1.85) with 95% CL. It is to be noted, the multi-lepton final state is dominated by the contribution from $t\bar{t}H$ multi-lepton process, which has similar kinematic properties. Most of the sensitivity to κ_t in this analysis is derived from the presence of $t\bar{t}H$ events, observed in Run 2 data [67]. The analysis is observed to have limited sensitivity to the variation of CP mixing angle. However, it is able to exclude the completely CP-odd scenario of the top-Higgs coupling sector with 95% CL.

A combination of results with other tH final states such as $H \rightarrow b\bar{b}$ and $H \rightarrow \gamma\gamma$ has been performed. The result is based on 35.9 fb^{-1} of integrated luminosity collected in 2016. In this analysis, the observed upper limit on the production cross section times branching ratio of the SM production rate of tH process is:

$$\sigma(tH) \times BR(H \rightarrow WW^*/ZZ^*/\tau\tau/b\bar{b}/\gamma\gamma) = 1.94 \text{ pb}$$

to be compared to the theoretical prediction of 0.077 pb. The values of κ_t outside (-0.9, -0.5) and (1.0, 2.1) are excluded with 95% CL.

Further, the combination results have been projected to higher luminosities and better understanding of systematics to be achieved in the HL-LHC era. It is expected that with 3000 fb^{-1} of data the analysis will become sensitive to the SM production rate. The measurement of the uncertainty in tH only signal strength is expected to improve by a factor eight. An optimistic estimation of the observed significance is $\sim 3\sigma$ with improved strategies and dedicated di-photon channel analysis. A very precise determination of the κ_t is thus expected in the presence of a SM like signal, values will be constrained within only the positive range.

In conclusion, in the era of precision physics, tH production plays a key role in the determination of the top-Higgs Yukawa coupling together with the $t\bar{t}H$ process. It remains to be verified how well the projection results compare with the actual LHC data: the future analyses may out-perform the prediction and achieve high precision as early as Run 3. Thus, search for new physics phenomena are expected to undergo interesting changes, pushing the boundaries of our knowledge.

Bibliography

- [1] S. L. Glashow, “Partial Symmetries of Weak Interactions,” *Nucl. Phys.* **22** (1961) 579–588.
- [2] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc.* **C680519** (1968) 367–377.
- [3] S. Weinberg, “A Model of Leptons,” *Phys. Rev. Lett.* **19** (1967) 1264–1266.
- [4] Y. Nambu and G. Jona-Lasinio, “Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I,” *Phys. Rev.* **122** (Apr, 1961) 345–358.
- [5] Y. Nambu and G. Jona-Lasinio, “Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. II,” *Phys. Rev.* **124** (Oct, 1961) 246–254.
- [6] J. Goldstone, A. Salam, and S. Weinberg, “Broken Symmetries,” *Phys. Rev.* **127** (1962) 965–970.
- [7] J. Goldstone, “Field Theories with Superconductor Solutions,” *Nuovo Cim.* **19** (1961) 154–164.
- [8] UA1 Collaboration, “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540\text{GeV}$,” *Phys. Lett.* **B122** no. 1, (1983) 103 – 116.
- [9] UA2 Collaboration, “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\bar{p}p$ collider,” *Phys. Lett.* **B122** no. 5, (1983) 476 – 485.

- [10] UA1 Collaboration, “Experimental observation of lepton pairs of invariant mass around $95\text{GeV}/c^2$ at the CERN SPS collider,” *Phys. Lett.* **B126** no. 5, (1983) 398 – 410.
- [11] UA2 Collaboration, “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN $\bar{p}p$ collider,” *Phys. Lett.* **B129** no. 1, (1983) 130 – 140.
- [12] ATLAS Collaboration, G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.* **B716** (2012) 1–29, [arXiv:1207.7214 \[hep-ex\]](#).
- [13] CMS Collaboration, S. Chatrchyan *et al.*, “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC,” *Phys. Lett.* **B716** (2012) 30–61, [arXiv:1207.7235 \[hep-ex\]](#).
- [14] E. Noether, “Invariante variationsprobleme,” *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918** (1918) 235–257. <http://eudml.org/doc/59024>.
- [15] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.
- [16] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.* **13** (1964) 508–509. [,160(1964)].
- [17] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles,” *Phys. Rev. Lett.* **13** (1964) 585–587. [,162(1964)].
- [18] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons,” *Phys. Rev.* **145** (1966) 1156–1163.
- [19] T. W. B. Kibble, “Symmetry breaking in nonAbelian gauge theories,” *Phys. Rev.* **155** (1967) 1554–1561. [,165(1967)].

- [20] T. van Ritbergen and R. G. Stuart, “Complete two loop quantum electrodynamic contributions to the muon lifetime in the Fermi model,” *Phys. Rev. Lett.* **82** (1999) 488–491, [arXiv:hep-ph/9808283 \[hep-ph\]](#).
- [21] T. van Ritbergen and R. G. Stuart, “On the precise determination of the Fermi coupling constant from the muon lifetime,” *Nucl. Phys.* **B564** (2000) 343–390, [arXiv:hep-ph/9904240 \[hep-ph\]](#).
- [22] M. Steinhauser and T. Seidensticker, “Second order corrections to the muon lifetime and the semileptonic B decay,” *Phys. Lett.* **B467** (1999) 271–278, [arXiv:hep-ph/9909436 \[hep-ph\]](#).
- [23] **MuLan** Collaboration, D. M. Webber *et al.*, “Measurement of the Positive Muon Lifetime and Determination of the Fermi Constant to Part-per-Million Precision,” *Phys. Rev. Lett.* **106** (2011) 041803.
- [24] E. Fermi, “An attempt of a theory of beta radiation. 1.,” *Z. Phys.* **88** (1934) 161–177.
- [25] **Particle Data Group** Collaboration, C. Patrignani *et al.*, “Review of Particle Physics,” *Chin. Phys.* **C40** no. 10, (2016) 100001.
- [26] **CMS** Collaboration, S. Chatrchyan *et al.*, “Observation of a New Boson with Mass Near 125 GeV in pp Collisions at $\sqrt{s} = 7$ and 8 TeV,” *JHEP* **06** (2013) 081, [arXiv:1303.4571 \[hep-ex\]](#).
- [27] G. ’t Hooft, C. Itzykson, A. Jaffe, H. Lehmann, P. K. Mitter, I. M. Singer, and R. Stora, “Recent Developments in Gauge Theories. Proceedings, Nato Advanced Study Institute, Cargese, France, August 26 - September 8, 1979,” *NATO Sci. Ser. B* **59** (1980) pp.1–438.
- [28] F. Zwicky, “On the Masses of Nebulae and of Clusters of Nebulae,” *Astrophys. J.* **86** (1937) 217–246.

- [29] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, “A direct empirical proof of the existence of dark matter,” *Astrophys. J.* **648** (2006) L109–L113, [arXiv:astro-ph/0608407 \[astro-ph\]](#).
- [30] **KamLAND** Collaboration, T. Araki *et al.*, “Measurement of neutrino oscillation with KamLAND: Evidence of spectral distortion,” *Phys. Rev. Lett.* **94** (2005) 081801, [arXiv:hep-ex/0406035 \[hep-ex\]](#).
- [31] **Daya Bay** Collaboration, F. P. An *et al.*, “Observation of electron-antineutrino disappearance at Daya Bay,” *Phys. Rev. Lett.* **108** (2012) 171803, [arXiv:1203.1669 \[hep-ex\]](#).
- [32] J. N. Bahcall and R. Davis, Jr., “On the Problem of Detecting Solar Neutrinos,” (1, 1964). <https://www.osti.gov/servlets/purl/4002837>.
- [33] **Super-Kamiokande** Collaboration, Y. Fukuda *et al.*, “Measurements of the solar neutrino flux from Super-Kamiokande’s first 300 days,” *Phys. Rev. Lett.* **81** (1998) 1158–1162, [arXiv:hep-ex/9805021 \[hep-ex\]](#). [Erratum: *Phys. Rev. Lett.* **81**, 4279 (1998)].
- [34] A. D. Sakharov, “Baryon asymmetry of the universe,” *Soviet Physics Uspekhi* **34** no. 5, (May, 1991) 417–421.
- [35] H. E. Haber, “Introductory low-energy supersymmetry,” in *Proceedings, Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles: Boulder, USA, June 1-26, 1992*, pp. 589–686. 1993. [arXiv:hep-ph/9306207 \[hep-ph\]](#).
- [36] H. P. Nilles, “Phenomenological aspects of supersymmetry,” in *Gauge theories, applied supersymmetry, quantum gravity. Proceedings, Workshop, Leuven, Belgium, July 10-14, 1995*, pp. 81–105. 1995. [arXiv:hep-ph/9511313 \[hep-ph\]](#).
- [37] H. Baer *et al.*, “Low-energy supersymmetry phenomenology,” [arXiv:hep-ph/9503479 \[hep-ph\]](#).

- [38] J. A. Bagger, “Weak scale supersymmetry: Theory and practice,” in *QCD and beyond. Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics, TASI-95, Boulder, USA, June 4-30, 1995*, pp. 109–162. 1996.
[arXiv:hep-ph/9604232 \[hep-ph\]](#).
- [39] D. J. Castaño, E. J. Piard, and P. Ramond, “Renormalization group study of the standard model and its extensions: The minimal supersymmetric standard model,” *Phys. Rev. D* **49** (May, 1994) 4882–4901.
- [40] M. Drees and S. P. Martin, “Implications of SUSY model building,”
[arXiv:hep-ph/9504324 \[hep-ph\]](#).
- [41] J. R. Ellis, G. Ridolfi, and F. Zwirner, “Radiative corrections to the masses of supersymmetric Higgs bosons,” *Phys. Lett.* **B257** (1991) 83–91.
- [42] J. R. Ellis, G. Ridolfi, and F. Zwirner, “On radiative corrections to supersymmetric Higgs boson masses and their implications for LEP searches,” *Phys. Lett.* **B262** (1991) 477–484.
- [43] **LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3, OPAL** Collaboration, R. Barate *et al.*, “Search for the standard model Higgs boson at LEP,” *Phys. Lett.* **B565** (2003) 61–75, [arXiv:hep-ex/0306033 \[hep-ex\]](#).
- [44] B. W. Lee, C. Quigg, and H. B. Thacker, “Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass,” *Phys. Rev.* **D16** (1977) 1519.
- [45] **LHC Higgs Cross Section Working Group** Collaboration, D. de Florian *et al.*, “Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector,”
[arXiv:1610.07922 \[hep-ph\]](#).
- [46] **LHC Higgs Cross Section Working Group** Collaboration, J. R. Andersen *et al.*, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties,” [arXiv:1307.1347 \[hep-ph\]](#).

- [47] **CMS Collaboration**, V. Khachatryan *et al.*, “Observation of the diphoton decay of the Higgs boson and measurement of its properties,” *Eur. Phys. J. C* **74** no. 10, (2014) 3076, [arXiv:1407.0558 \[hep-ex\]](#).
- [48] **CMS Collaboration**, S. Chatrchyan *et al.*, “Measurement of the properties of a Higgs boson in the four-lepton final state,” *Phys. Rev. D* **89** no. 9, (2014) 092007, [arXiv:1312.5353 \[hep-ex\]](#).
- [49] **CMS Collaboration**, “A measurement of the Higgs boson mass in the diphoton decay channel,” Tech. Rep. CMS-PAS-HIG-19-004, CERN, Geneva, 2019. <http://cds.cern.ch/record/2691211>.
- [50] **ATLAS, CMS Collaboration**, G. Aad *et al.*, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments,” *Phys. Rev. Lett.* **114** (2015) 191803, [arXiv:1503.07589 \[hep-ex\]](#).
- [51] **Physics of the HL-LHC Working Group Collaboration**, M. Cepeda *et al.*, “Higgs Physics at the HL-LHC and HE-LHC,” [arXiv:1902.00134 \[hep-ph\]](#).
- [52] **CMS Collaboration**, A. M. Sirunyan *et al.*, “Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector,” *Phys. Lett. B* **779** (2018) 283–316, [arXiv:1708.00373 \[hep-ex\]](#).
- [53] **ATLAS Collaboration**, M. Aaboud *et al.*, “Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector,” *Phys. Lett. B* **786** (2018) 59–86, [arXiv:1808.08238 \[hep-ex\]](#).
- [54] **CMS Collaboration**, A. M. Sirunyan *et al.*, “Observation of Higgs boson decay to bottom quarks,” *Phys. Rev. Lett.* **121** no. 12, (2018) 121801, [arXiv:1808.08242 \[hep-ex\]](#).
- [55] **LHC Higgs Cross Section Working Group Collaboration**, A. David, A. Denner, M. Duehrssen, M. Grazzini, C. Grojean, G. Passarino, M. Schumacher, M. Spira, G. Weiglein, and M. Zanetti, “LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle,” [arXiv:1209.0040 \[hep-ph\]](#).

- [56] **ATLAS, CMS** Collaboration, G. Aad *et al.*, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV,” *JHEP* **08** (2016) 045, [arXiv:1606.02266 \[hep-ex\]](#).
- [57] **CMS** Collaboration, V. Khachatryan *et al.*, “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV,” *Eur. Phys. J.* **C75** no. 5, (2015) 212, [arXiv:1412.8662 \[hep-ex\]](#).
- [58] B. A. Dobrescu and C. T. Hill, “Electroweak symmetry breaking via top condensation seesaw,” *Phys. Rev. Lett.* **81** (1998) 2634–2637, [arXiv:hep-ph/9712319 \[hep-ph\]](#).
- [59] R. S. Chivukula, B. A. Dobrescu, H. Georgi, and C. T. Hill, “Top Quark Seesaw Theory of Electroweak Symmetry Breaking,” *Phys. Rev.* **D59** (1999) 075003, [arXiv:hep-ph/9809470 \[hep-ph\]](#).
- [60] D. Delepine, J. M. Gerard, and R. Gonzalez Felipe, “Is the standard Higgs scalar elementary?,” *Phys. Lett.* **B372** (1996) 271–274, [arXiv:hep-ph/9512339 \[hep-ph\]](#).
- [61] S. Alekhin, A. Djouadi, and S. Moch, “The top quark and Higgs boson masses and the stability of the electroweak vacuum,” *Phys. Lett.* **B716** (2012) 214–219, [arXiv:1207.0980 \[hep-ph\]](#).
- [62] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, “Theory and phenomenology of two-Higgs-doublet models,” *Phys. Rept.* **516** (2012) 1–102, [arXiv:1106.0034 \[hep-ph\]](#).
- [63] A. Djouadi, “The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model,” *Phys. Rept.* **459** (2008) 1–241, [arXiv:hep-ph/0503173 \[hep-ph\]](#).

- [64] P. M. Ferreira, J. F. Gunion, H. E. Haber, and R. Santos, “Probing wrong-sign Yukawa couplings at the LHC and a future linear collider,” *Phys. Rev.* **D89** no. 11, (2014) 115003, [arXiv:1403.4736 \[hep-ph\]](#).
- [65] O. Eberhardt, “Two-Higgs-doublet model fits with HEPfit,” in *Proceedings, 2017 European Physical Society Conference on High Energy Physics (EPS-HEP 2017): Venice, Italy, July 5-12, 2017*. 2017. [arXiv:1709.09414 \[hep-ph\]](#).
<https://pos.sissa.it/314/281/pdf>.
- [66] D. Chowdhury and O. Eberhardt, “Update of Global Two-Higgs-Doublet Model Fits,” *JHEP* **05** (2018) 161, [arXiv:1711.02095 \[hep-ph\]](#).
- [67] CMS Collaboration, A. M. Sirunyan *et al.*, “Observation of $t\bar{t}H$ production,” *Phys. Rev. Lett.* **120** no. 23, (2018) 231801, [arXiv:1804.02610 \[hep-ex\]](#).
- [68] ATLAS Collaboration, M. Aaboud *et al.*, “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector,” *Phys. Lett.* **B784** (2018) 173–191, [arXiv:1806.00425 \[hep-ex\]](#).
- [69] CMS Collaboration, V. Khachatryan *et al.*, “Search for the associated production of a Higgs boson with a single top quark in proton-proton collisions at $\sqrt{s} = 8$ TeV,” *JHEP* **06** (2016) 177, [arXiv:1509.08159 \[hep-ex\]](#).
- [70] J. J. Thomson, “On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure,” *Phil. Mag. Ser.6* **7** no. 39, (1904) 237–265.
- [71] J. D. Cockcroft and E. T. S. Walton, “Disintegration of lithium by swift protons,” *Nature* **129** no. 3261, (1932) 649, <https://doi.org/10.1038/129649a0>.
- [72] ALICE Collaboration, K. Aamodt *et al.*, “The ALICE experiment at the CERN LHC,” *JINST* **3** (2008) S08002.

- [73] **ATLAS** Collaboration, G. Aad *et al.*, “The ATLAS Experiment at the CERN Large Hadron Collider,” [JINST 3 \(2008\) S08003](#).
- [74] **CMS** Collaboration, S. Chatrchyan *et al.*, “The CMS Experiment at the CERN LHC,” [JINST 3 \(2008\) S08004](#).
- [75] **LHCb** Collaboration, A. A. Alves, Jr. *et al.*, “The LHCb Detector at the LHC,” [JINST 3 \(2008\) S08005](#).
- [76] **CMS** Collaboration, “The CMS tracker system project: Technical Design Report,” Tech. Rep. CERN-LHCC-98-006, CMS-TDR-5, Geneva, 1997.
<http://cds.cern.ch/record/368412>.
- [77] **CMS** Collaboration, “CMS Technical Design Report for the Pixel Detector Upgrade,” Tech. Rep. CERN-LHCC-2012-016, CMS-TDR-11, Geneva, 2012.
<http://cds.cern.ch/record/1481838>.
- [78] **CMS** Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report,” Tech. Rep. CERN-LHCC-97-033, CMS-TDR-4, Geneva, 1997.
<http://cds.cern.ch/record/349375>.
- [79] P. Adzic *et al.*, “Energy resolution of the barrel of the CMS electromagnetic calorimeter,” [JINST 2 \(2007\) P04004](#).
- [80] **CMS** Collaboration, “The CMS hadron calorimeter project: Technical Design Report,” Tech. Rep. CERN-LHCC-97-031, CMS-TDR-2, Geneva, 1997.
<http://cds.cern.ch/record/357153>.
- [81] **CMS** Collaboration, “CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter,” Tech. Rep. CERN-LHCC-2012-015, CMS-TDR-010, Geneva, 2012. <http://cds.cern.ch/record/1481837>.
- [82] **CMS HCAL** Collaboration, S. Abdullin *et al.*, “Design, performance, and calibration of the CMS Hadron-outer calorimeter,” [Eur. Phys. J. C57 \(2008\) 653–663](#).

- [83] **CMS** Collaboration, S. Chatrchyan *et al.*, “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data,” *JINST* **5** (2010) T03012, [arXiv:0911.4991 \[physics.ins-det\]](#).
- [84] **CMS** Collaboration, “The CMS muon project: Technical Design Report,” Tech. Rep. CERN-LHCC-97-032, CMS-TDR-3, Geneva, 1997. <http://cds.cern.ch/record/343814>.
- [85] **CMS** Collaboration, “CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems,” Tech. Rep. CERN-LHCC-2000-038, CMS-TDR-6-1, Geneva, 2000. <http://cds.cern.ch/record/706847>.
- [86] **CMS** Collaboration, “CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger,” Tech. Rep. CERN-LHCC-2002-026, CMS-TDR-6, Geneva, 2002. <http://cds.cern.ch/record/578006>.
- [87] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Particle-flow reconstruction and global event description with the CMS detector,” *JINST* **12** no. 10, (2017) P10003, [arXiv:1706.04965 \[physics.ins-det\]](#).
- [88] S. Forte, L. Garrido, J. I. Latorre, and A. Piccione, “Neural network parametrization of deep inelastic structure functions,” *JHEP* **05** (2002) 062, [arXiv:hep-ph/0204232 \[hep-ph\]](#).
- [89] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP* **07** (2014) 079, [arXiv:1405.0301 \[hep-ph\]](#).
- [90] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP* **06** (2010) 043, [arXiv:1002.2581 \[hep-ph\]](#).

- [91] G. Marchesini and B. R. Webber, “Simulation of QCD jets including soft gluon interference,” *Nucl. Phys. B* **238** no. CERN-TH-3525, (Feb, 1983) 1–29. 39 p. <http://cds.cern.ch/record/143716>.
- [92] T. D. Gottschalk, “Backwards Evolved Initial State Parton Showers,” *Nucl. Phys.* **B277** (1986) 700–738.
- [93] T. Sjostrand, “A Model for Initial State Parton Showers,” *Phys. Lett.* **157B** (1985) 321–325.
- [94] J. Alwall *et al.*, “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions,” *Eur. Phys. J.* **C53** (2008) 473–500, [arXiv:0706.2569](https://arxiv.org/abs/0706.2569) [hep-ph].
- [95] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO,” *JHEP* **12** (2012) 061, [arXiv:1209.6215](https://arxiv.org/abs/1209.6215) [hep-ph].
- [96] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, “Parton Fragmentation and String Dynamics,” *Phys. Rept.* **97** (1983) 31–145.
- [97] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, “An Introduction to PYTHIA 8.2,” *Comput. Phys. Commun.* **191** (2015) 159–177, [arXiv:1410.3012](https://arxiv.org/abs/1410.3012) [hep-ph].
- [98] M. Bahr *et al.*, “Herwig++ Physics and Manual,” *Eur. Phys. J.* **C58** (2008) 639–707, [arXiv:0803.0883](https://arxiv.org/abs/0803.0883) [hep-ph].
- [99] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune,” *Eur. Phys. J.* **C74** no. 8, (2014) 3024, [arXiv:1404.5630](https://arxiv.org/abs/1404.5630) [hep-ph].
- [100] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Trans. Nucl. Sci.* **53** (2006) 270.
- [101] **CMS** Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET,” Tech. Rep. CMS-PAS-PFT-09-001, CERN, Geneva, Apr, 2009. <http://cds.cern.ch/record/1194487>.

- [102] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *ASME Journal of Basic Engineering* (1960) .
- [103] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems,” in *Proceedings of the IEEE*, pp. 2210–2239. 1998.
- [104] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, “Adaptive vertex fitting,” *J. Phys.* **G34** (2007) N343.
- [105] **CMS** Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *JINST* **9** no. 10, (2014) P10009, [arXiv:1405.6569 \[physics.ins-det\]](#).
- [106] **CMS** Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV,” *JINST* **13** no. 06, (2018) P06015, [arXiv:1804.04528 \[physics.ins-det\]](#).
- [107] **CMS** Collaboration, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV,” *JINST* **10** no. 06, (2015) P06005, [arXiv:1502.02701 \[physics.ins-det\]](#).
- [108] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, “Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC,” Tech. Rep. CMS-NOTE-2005-001, CERN, Geneva, Jan, 2005. <http://cds.cern.ch/record/815410>.
- [109] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,” *JHEP* **04** (2008) 063, [arXiv:0802.1189 \[hep-ph\]](#).
- [110] **CMS** Collaboration, V. Khachatryan *et al.*, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV,” *JINST* **12** no. 02, (2017) P02014, [arXiv:1607.03663 \[hep-ex\]](#).

- [111] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas,” *Phys. Lett. B* **659** (2008) 119–126, [arXiv:0707.1378 \[hep-ph\]](#).
- [112] CMS Collaboration, “Performance of heavy flavour identification algorithms in proton-proton collisions at 13 TeV at the CMS experiment,” <http://cds.cern.ch/record/2263801>.
- [113] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification,” *JHEP* **10** (2014) 059, [arXiv:1407.6013 \[hep-ph\]](#).
- [114] CMS Collaboration, A. M. Sirunyan *et al.*, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements,” [arXiv:1903.12179 \[hep-ex\]](#).
- [115] CMS Collaboration, V. Khachatryan *et al.*, “Event generator tunes obtained from underlying event and multiparton scattering measurements,” *Eur. Phys. J. C* **76** no. 3, (2016) 155, [arXiv:1512.00815 \[hep-ex\]](#).
- [116] “Cut based photon ID for Run 2.”
https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedPhotonIdentificationRun2#Working_points_for_94X_and_later.
- [117] “Jet ID for Run 2.”
<https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016>,
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID13TeVRun2017>,
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID13TeVRun2018>.
- [118] “Cut based electron ID for Run 2.”
https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2#Working_points_for_94X_and_later.
- [119] “Muon ID for Run 2.”
<https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2>.

- [120] CMS Collaboration, “Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector,” *JINST* **14** no. 07, (2019) P07004, [arXiv:1903.06078 \[hep-ex\]](#).
- [121] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, “Associated production of Higgs and single top at hadron colliders,” *Phys. Rev.* **D64** (2001) 094023, [arXiv:hep-ph/0106293 \[hep-ph\]](#).
- [122] “SM Higgs production cross sections at $\sqrt{s} = 13$ TeV (update in CERN Report4 2016).” <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV>.
- [123] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, “Higgs production in association with a single top quark at the LHC,” *Eur. Phys. J.* **C75** no. 6, (2015) 267, [arXiv:1504.00611 \[hep-ph\]](#).
- [124] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, “tWH associated production at the LHC,” *Eur. Phys. J.* **C77** no. 1, (2017) 34, [arXiv:1607.05862 \[hep-ph\]](#).
- [125] M. Farina, C. Grojean, F. Maltoni, E. Salvioni, and A. Thamm, “Lifting degeneracies in Higgs couplings using single top production in association with a Higgs boson,” *JHEP* **05** (2013) 022, [arXiv:1211.3736 \[hep-ph\]](#).
- [126] S. Biswas, E. Gabrielli, and B. Mele, “Single top and Higgs associated production as a probe of the Htt coupling sign at the LHC,” *JHEP* **01** (2013) 088, [arXiv:1211.0499 \[hep-ph\]](#).
- [127] S. Biswas, E. Gabrielli, F. Margaroli, and B. Mele, “Direct constraints on the top-Higgs coupling from the 8 TeV LHC data,” *JHEP* **07** (2013) 073, [arXiv:1304.1822 \[hep-ph\]](#).
- [128] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi, “Disentangling Higgs-Top Couplings in Associated Production,” *JHEP* **04** (2014) 004, [arXiv:1312.5736 \[hep-ph\]](#).

- [129] J. Yue, “Enhanced thj signal at the LHC with $h \rightarrow \gamma\gamma$ decay and $\mathcal{CP}$ -violating top-Higgs coupling,” *Phys. Lett.* **B744** (2015) 131–136, [arXiv:1410.2701 \[hep-ph\]](#).
- [130] A. Kobakhidze, L. Wu, and J. Yue, “Anomalous Top-Higgs Couplings and Top Polarisation in Single Top and Higgs Associated Production at the LHC,” *JHEP* **10** (2014) 100, [arXiv:1406.1961 \[hep-ph\]](#).
- [131] P. Artoisenet *et al.*, “A framework for Higgs characterisation,” *JHEP* **11** (2013) 043, [arXiv:1306.6464 \[hep-ph\]](#).
- [132] S. L. Glashow, J. Iliopoulos, and L. Maiani, “Weak Interactions with Lepton-Hadron Symmetry,” *Phys. Rev.* **D2** (1970) 1285–1292.
- [133] J. A. Aguilar-Saavedra, “Top flavor-changing neutral interactions: Theoretical expectations and experimental detection,” *Acta Phys. Polon.* **B35** (2004) 2695–2710, [arXiv:hep-ph/0409342 \[hep-ph\]](#).
- [134] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV,” *Eur. Phys. J.* **C79** no. 5, (2019) 421, [arXiv:1809.10733 \[hep-ex\]](#).
- [135] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *Phys. Rev.* **D99** no. 9, (2019) 092005, [arXiv:1811.09696 \[hep-ex\]](#).
- [136] **CMS** Collaboration, S. Chatrchyan *et al.*, “Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs,” *Phys. Rev. Lett.* **110** no. 8, (2013) 081803, [arXiv:1212.6639 \[hep-ex\]](#).
- [137] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV,” *JHEP* **08** (2018) 066, [arXiv:1803.05485 \[hep-ex\]](#).

- [138] CMS Collaboration, “Measurement of the associated production of a Higgs boson with a top quark pair in final states with electrons, muons and hadronically decaying τ leptons in data recorded in 2017 at $\sqrt{s} = 13$ TeV,” Tech. Rep. CMS-PAS-HIG-18-019, CERN, Geneva, 2018.
<http://cds.cern.ch/record/2649199>.
- [139] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of Research and Development* **3** no. 3, (July, 1959) 210–229.
- [140] A. Hocker *et al.*, “TMVA - Toolkit for Multivariate Data Analysis,” [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039) [physics.data-an].
- [141] I. Antcheva *et al.*, “ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization,” *Comput. Phys. Commun.* **180** (2009) 2499–2512, [arXiv:1508.07749](https://arxiv.org/abs/1508.07749) [physics.data-an].
- [142] J. Quinlan, “Simplifying decision trees,” *International Journal of Man-Machine Studies* **27** no. 3, (1987) 221 – 234.
<http://www.sciencedirect.com/science/article/pii/S0020737387800536>.
- [143] R. E. Schapire, “The strength of weak learnability,” *Machine Learning* **5** no. 2, (Jun, 1990) 197–227. <https://doi.org/10.1007/BF00116037>.
- [144] A. Wald, “Statistical decision functions,” *Ann. Math. Statist.* **20** no. 2, (06, 1949) 165–205. <https://doi.org/10.1214/aoms/1177730030>.
- [145] G. Cowan, *Statistical Data Analysis*. Clarendon (Oxford), Oxford, 1998 (ISBN: 0-19-850156-0 or 0-19-850155-2 in paperback).
- [146] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests of statistical hypotheses,” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **231** (1933) 289–337. <http://www.jstor.org/stable/91247>.

- [147] **ATLAS, CMS, LHC Higgs Combination Group** Collaboration, “Procedure for the LHC Higgs boson search combination in summer 2011,”
<https://cds.cern.ch/record/1379837>.
- [148] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C* **71** (2011) 1554,
[arXiv:1007.1727](https://arxiv.org/abs/1007.1727) [physics.data-an].
- [149] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations,” *JHEP* **03** (2013) 015, [arXiv:1212.3460](https://arxiv.org/abs/1212.3460) [hep-ph].
- [150] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, “Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations,” *JHEP* **04** (2007) 081, [arXiv:hep-ph/0702198](https://arxiv.org/abs/hep-ph/0702198) [HEP-PH].
- [151] B. Maier, “Singletophiggproduction13tev,” February, 2016. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV>.
- [152] M. Peruzzi, C. Mueller, B. Stieger, *et al.*, “Search for $t\bar{t}H$ in multilepton final states at 13 tev,” CMS Analysis Note CMS AN-16-211, CERN, 2016. http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2016_211_v4.pdf.
- [153] M. Peruzzi, F. Romeo, B. Stieger, *et al.*, “Search for $t\bar{t}H$ in multilepton final states at 13 tev,” CMS Analysis Note CMS AN-17-029, CERN, 2017. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2017/029.
- [154] **CMS** Collaboration, “Studies of Tracker Material,” Tech. Rep. CMS-PAS-TRK-10-003, CERN, 2010. <https://cds.cern.ch/record/1279138>.
- [155] **CMS** Collaboration, V. Khachatryan *et al.*, “Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS,” *JINST* **11** no. 01, (2016) P01019,
[arXiv:1510.07488](https://arxiv.org/abs/1510.07488) [physics.ins-det].

- [156] **CMS Collaboration**, A. M. Sirunyan *et al.*, “Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV,” *JINST* **13** no. 10, (2018) P10005, [arXiv:1809.02816 \[hep-ex\]](#).
- [157] “B tag recommendation for Run 2.” <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco>, <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X>.
- [158] **CMS Collaboration**, “CMS Luminosity Measurements for the 2016 Data Taking Period,”. <http://cds.cern.ch/record/2257069>.
- [159] **LHC Higgs Cross Section Working Group Collaboration**, S. H. et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. CERN Yellow Reports: Monographs. 2013.
<http://cds.cern.ch/record/1559921>.
- [160] **CMS Collaboration**, A. M. Sirunyan *et al.*, “Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV,” *Phys. Rev. Lett.* **122** no. 13, (2019) 132003, [arXiv:1812.05900 \[hep-ex\]](#).
- [161] Tau POG. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendation13TeV>.
- [162] L. Thomas. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe>.
- [163] L. Thomas. <https://indico.cern.ch/event/718554/contributions/3030794/attachments/1667627/2674094/ecalprefiring.pdf>.
- [164] **CMS Collaboration**, “Search for charged Higgs bosons decaying into top and a bottom quark in the fully hadronic final state at 13 TeV,” Tech. Rep. CMS-PAS-HIG-18-015, CERN, Geneva, 2019.
<http://cds.cern.ch/record/2682105>.

- [165] **CMS** Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV,” <https://cds.cern.ch/record/2621960>.
- [166] H. Bakhshiansohi, A. Giammanco, A. Jafari, and G. Krintiras, “Search for anomalous higgs yukawa coupling in single top quark associated higgs production with $h \rightarrow \gamma\gamma$ at $\sqrt{s} = 13$ tev,” CMS Analysis Note CMS AN-17-160, 2017. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2017/160.
- [167] **CMS** Collaboration, “Search for the $tH(H \rightarrow b\bar{b})$ process in pp collisions at $\sqrt{s} = 13$ TeV and study of Higgs boson couplings,” Tech. Rep. CMS-PAS-HIG-17-016, CERN, Geneva, 2018. <http://cds.cern.ch/record/2621422>.
- [168] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV,” [JHEP 11 \(2018\) 185](#), [arXiv:1804.02716 \[hep-ex\]](#).
- [169] HL-LHC Plan update. https://hilumilhcds.web.cern.ch/sites/hilumilhcds.web.cern.ch/files/HL_LHC_PlanUpdateJuly2015.PNG.
- [170] **CMS** Collaboration, “Technical Proposal for the Phase-II Upgrade of the CMS Detector,” Tech. Rep. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, 2015. <https://cds.cern.ch/record/2020886>.
- [171] **CMS** Collaboration, C. Collaboration, “The Phase-2 Upgrade of the CMS Tracker,” Tech. Rep. CERN-LHCC-2017-009. CMS-TDR-014, 2017. <https://cds.cern.ch/record/2272264>.
- [172] **CMS** Collaboration, C. Collaboration, “The Phase-2 Upgrade of the CMS Barrel Calorimeters Technical Design Report,” Tech. Rep. CERN-LHCC-2017-011. CMS-TDR-015, 2017. <https://cds.cern.ch/record/2283187>.

- [173] **CMS** Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter,” Tech. Rep. CERN-LHCC-2017-023. CMS-TDR-019, 2017.
<https://cds.cern.ch/record/2293646>.
- [174] **CMS** Collaboration, “The Phase-2 Upgrade of the CMS Muon Detectors,” Tech. Rep. CERN-LHCC-2017-012. CMS-TDR-016, 2017.
<https://cds.cern.ch/record/2283189>.
- [175] **CMS** Collaboration, “Sensitivity projections for Higgs boson properties measurements at the HL-LHC,” Tech. Rep. CMS-PAS-FTR-18-011, CERN, Geneva, 2018. <http://cds.cern.ch/record/2647699>.

Appendices

Appendix A

Training of BDT discriminators

Table A.1 lists the variable names as used in the trainings.

Table A.1: MVA input discriminating variables.

Variable name	Description
nJet25	Number of jets with $p_T > 25$ GeV, $ \eta < 2.4$
MaxEtaJet25	Maximum $ \eta $ of any (non-CSV-loose) jet with $p_T > 25$ GeV
totCharge	Sum of lepton charges
nJetEta1	Number of jets with $ \eta > 1.0$, failing loose b-tag
detaFwdJetBJet	$\Delta\eta$ between forward light jet and hardest loosely b-tagged jet
detaFwdJet2BJet	$\Delta\eta$ between forward light jet and second hardest loosely b-tagged jet
detaFwdJetClosestLep	$\Delta\eta$ between forward light jet and closest lepton
dphiHighestPtSSPair	$\Delta\phi$ of highest p_T same-sign lepton pair
minDRll	Minimum ΔR between any two leptons
Lep3Pt/Lep2Pt	p_T of the 3 rd lepton in 3ℓ , 2 nd for $2\ell ss$
SumJetsPt	Sum of the p_T of all selected jets in the event
BDThttTT	Hadronic top tagger score of the event

A.1 Training with 2016 samples

For all trainings, the signal is tHq in the ITC scenario ($\kappa_t=-1.0$). The $t\bar{t}V$ background is an admixture of $t\bar{t}W$ and $t\bar{t}Z$ processes, used in a rough proportion of their cross sections. The TMVA settings used in the BDT training are shown in Table A.2.

Table A.2: TMVA configuration used in the training.

TMVA.Types.kBDT
NTrees=800
BoostType=Grad
Shrinkage=0.10
!UseBaggedGrad
nCuts=50
MaxDepth=3
NegWeightTreatment=PairNegWeightsGlobal
CreateMVAPdfs

Figures A.1 and A.2 show the input variables distributions as seen by the MVA algorithm for the 3ℓ channel in the two different trainings. The input variables distributions for $2\ell ss$ channel for signal against the $t\bar{t}$ and $t\bar{t}V$ are shown in Figs. A.3 and A.4 respectively.

Note that splitting the training in two groups reveals that some variables show opposite behaviour for the two background sources; potentially screening the discrimination power if they were to be used in a single discriminant. For some other variables the distributions are similar in both background cases.

A.1.1 Input variables correlation

Figure A.5 show the linear correlation coefficients for signal and background in 3ℓ channel for the two training cases (the signal values are identical by construction). As expected, strong correlations appears for variables related to the forward jet activity. Same trend is seen in case of the $2\ell ss$ channel in Figure A.6.

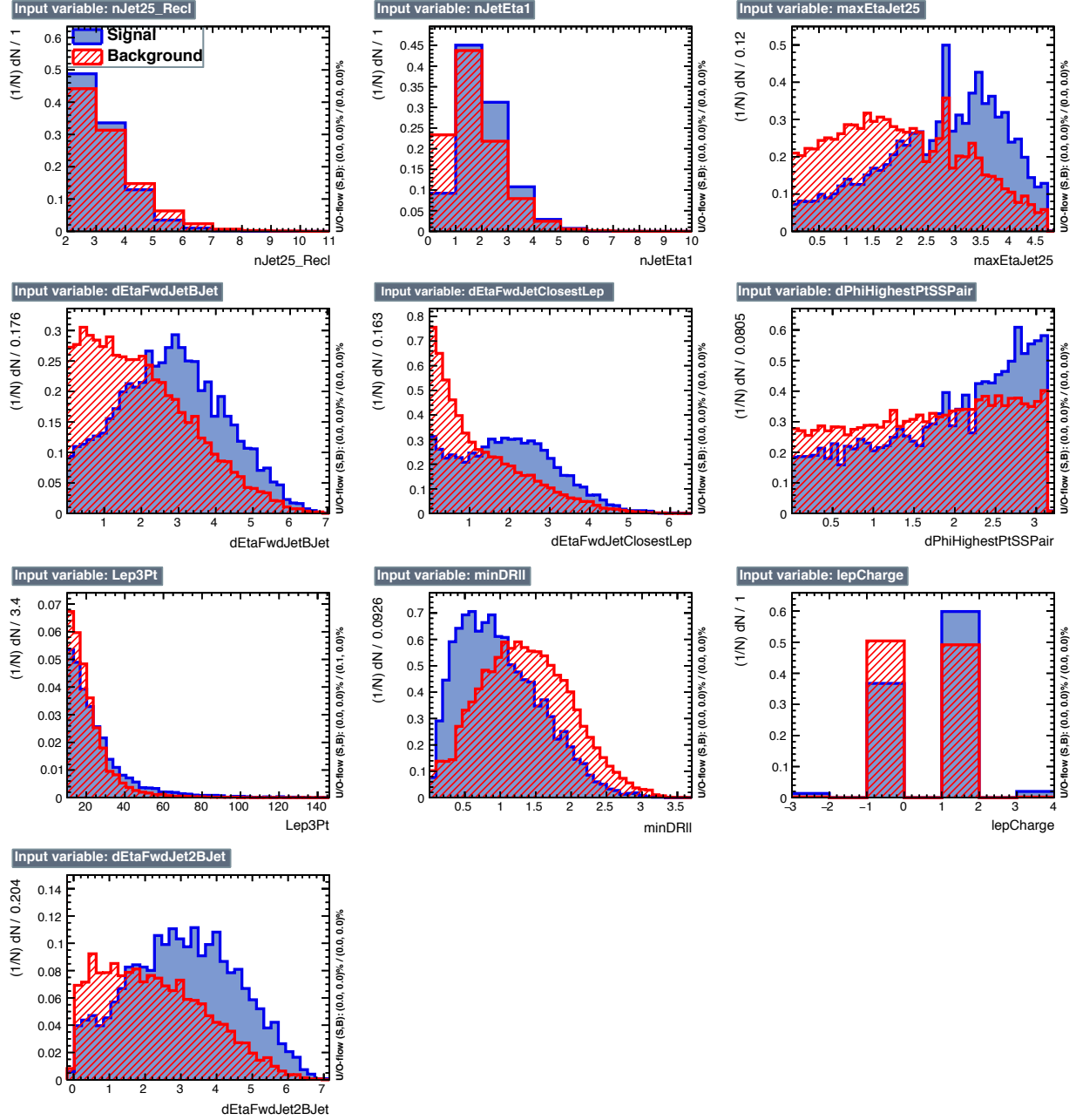


Figure A.1: BDT inputs as seen by TMVA (signal in blue, background in red) for the 3ℓ channel, discriminated against $t\bar{t}$ (non-prompt) background in 2016 analysis.

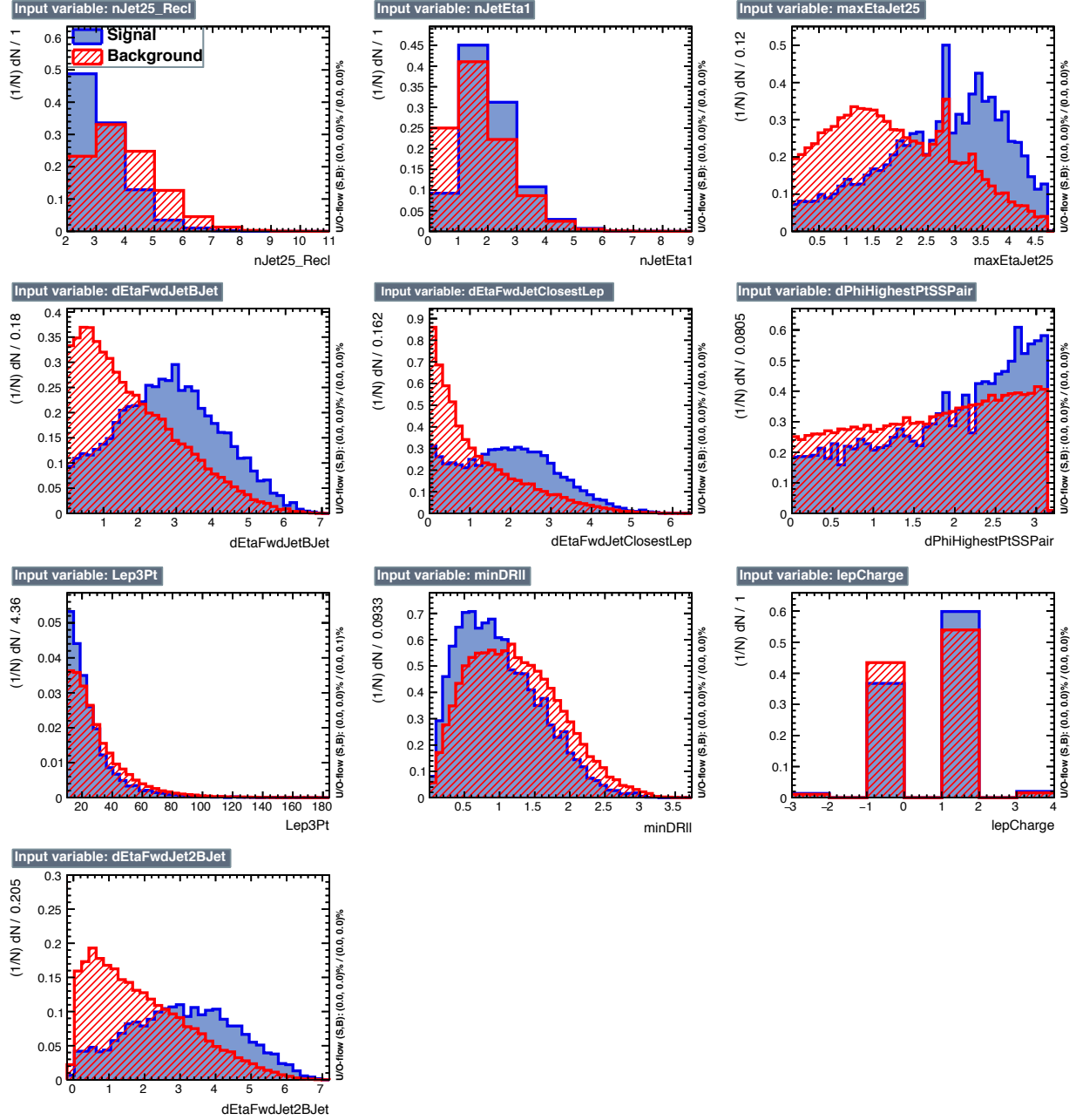


Figure A.2: BDT inputs as seen by TMVA (signal in blue, background in red) for the 3 ℓ channel, discriminated against $t\bar{t}V$ (prompt) background in 2016 analysis.

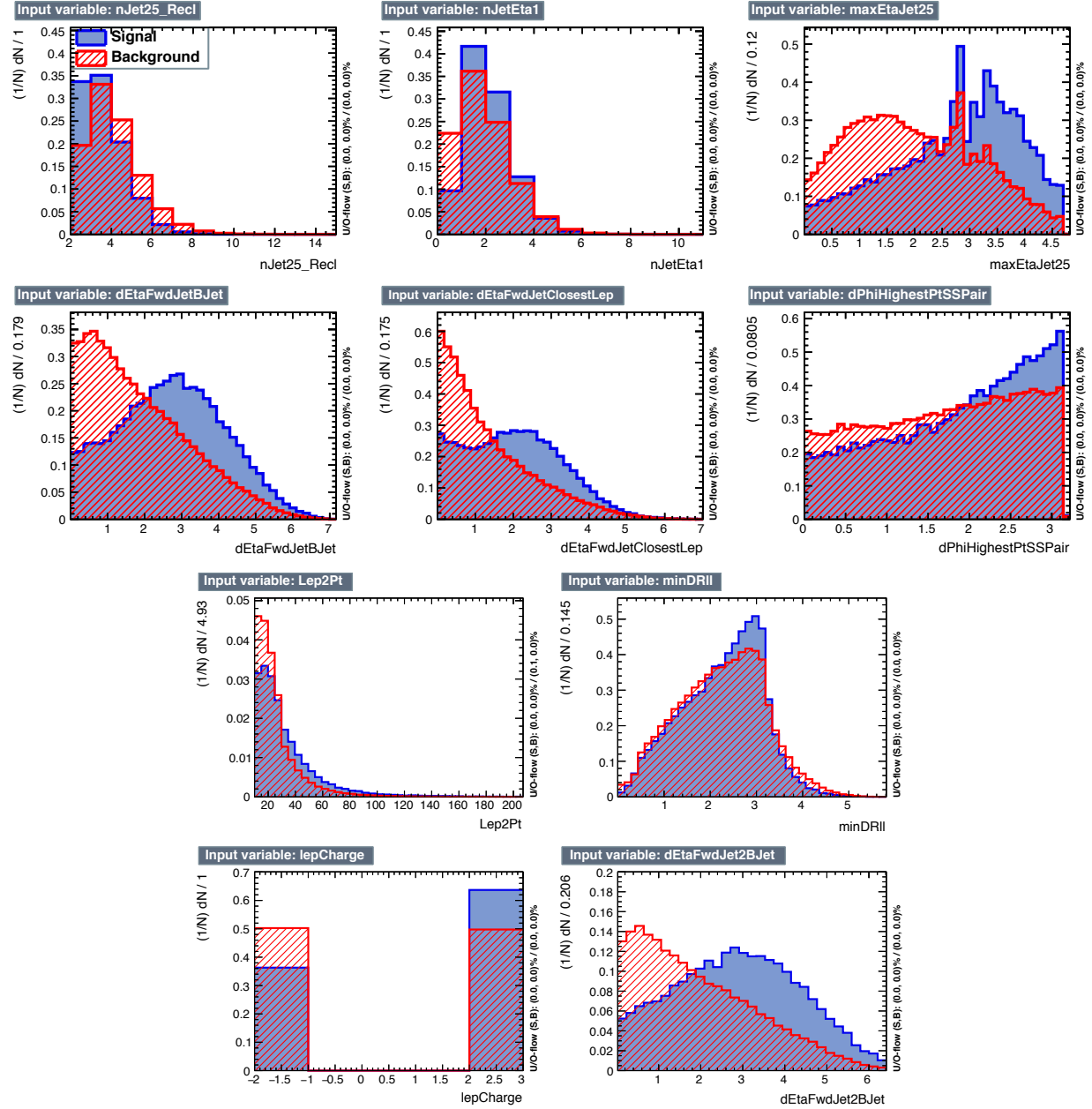


Figure A.3: BDT inputs as seen by TMVA (signal in blue, background in red) for the $2\ell ss$ channel, discriminated against $t\bar{t}$ (non-prompt) background in 2016 analysis.

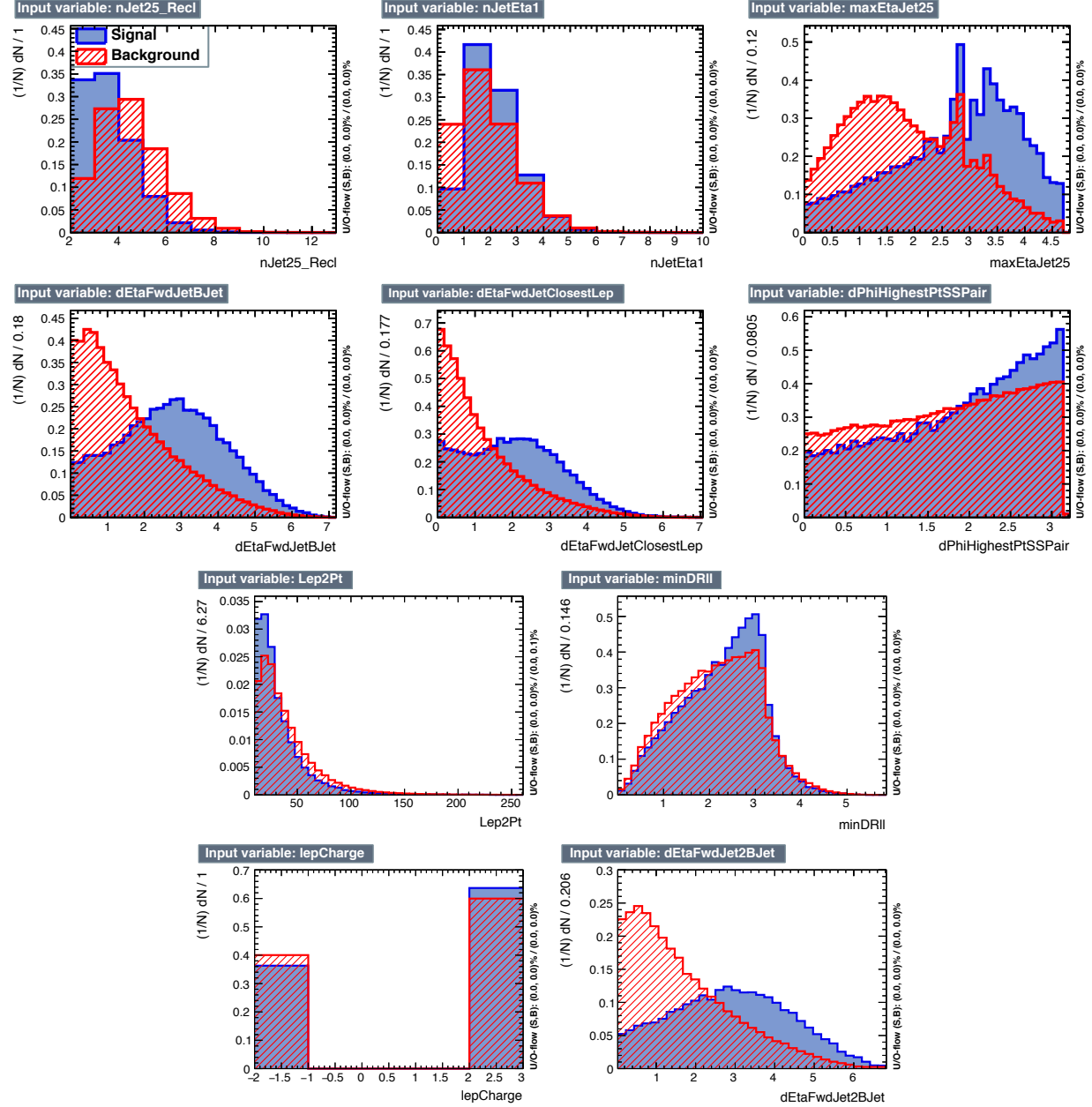


Figure A.4: BDT inputs as seen by TMVA (signal in blue, background in red) for the $2\ell ss$ channel, discriminated against $t\bar{t}V$ (prompt) background in 2016 analysis.

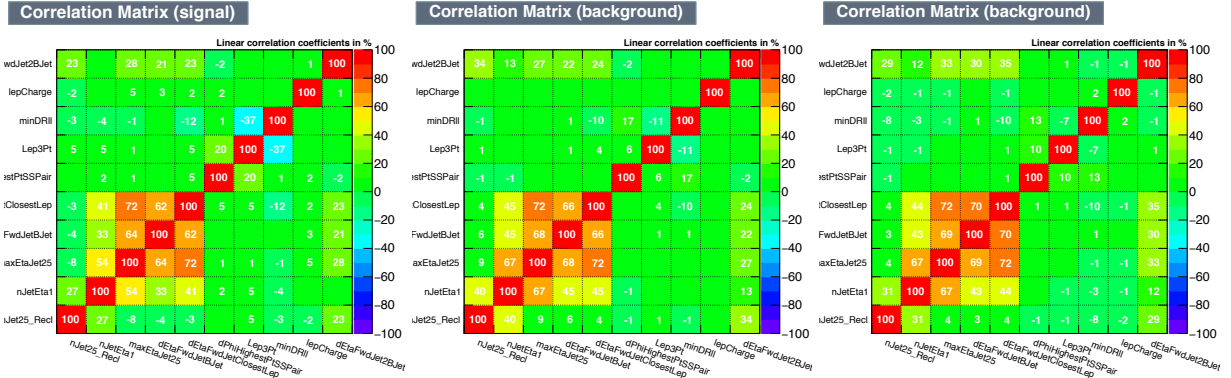


Figure A.5: Signal (left), $t\bar{t}$ background (middle), and $t\bar{t}V$ background (right.) correlation matrices for the input variables in the TMVA analysis for the 3 ℓ channel in 2016 analysis.

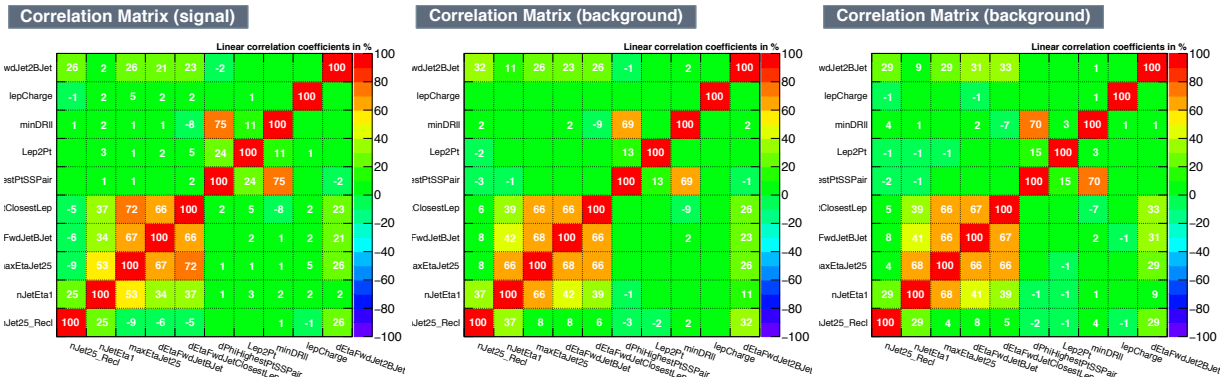


Figure A.6: Signal (left), $t\bar{t}$ background (middle), and $t\bar{t}V$ background (right.) correlation matrices for the input variables in the TMVA analysis for the 2 ℓ ss channel in 2016 analysis.

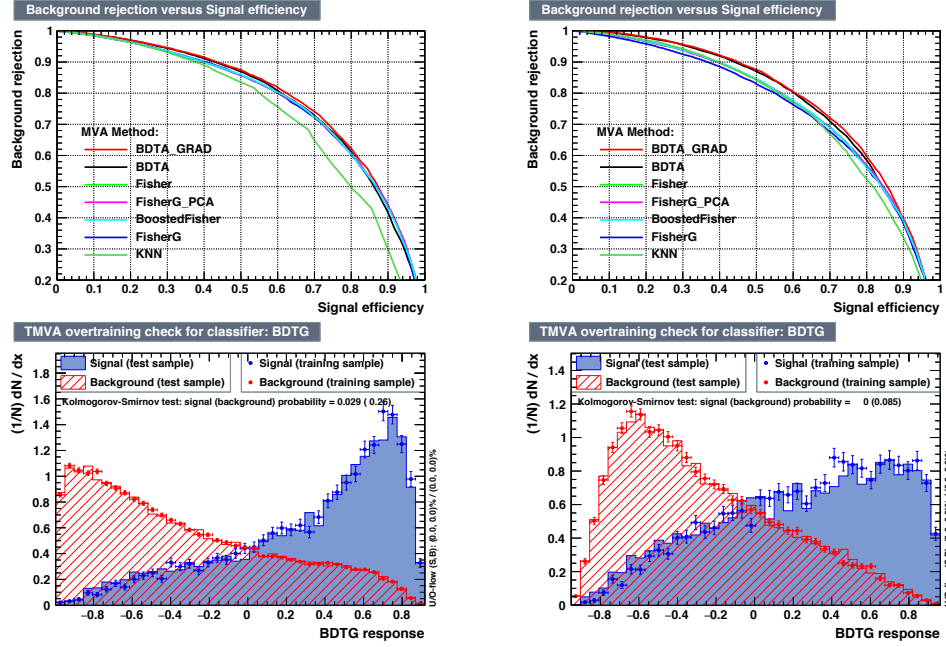


Figure A.7: Top: background rejection vs signal efficiency (ROC curves) for various MVA classifiers (top) in the 3ℓ channel against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Bottom: classifier output distributions for the gradient boosted decision trees, for training against $t\bar{t}V$ (left) and against $t\bar{t}$ (right).

A.1.2 Classifier response

The plots at the top in Figs. [A.7](#) and [A.8](#) show the background rejection as a function of the signal efficiency for $t\bar{t}$ and $t\bar{t}V$ trainings (ROC curves). Several MVA algorithms were evaluated to determine the most appropriate method for this analysis using the 3ℓ channel.

In both cases the gradient boosted decision tree (BDTA_GRAD or BDTG) classifier offers the best results, followed by an adaptive BDT classifier. The BDTG classifier was chosen for all subsequent analysis. The classifier output distributions for signal and backgrounds are shown on the bottom of Fig. [A.7](#). As expected, a good discrimination power is obtained using default discriminator parameter values, with minimal overtraining. TMVA provides a ranking of the input variables by their importance in the classification process, shown in Tables [A.3](#) and [A.4](#).

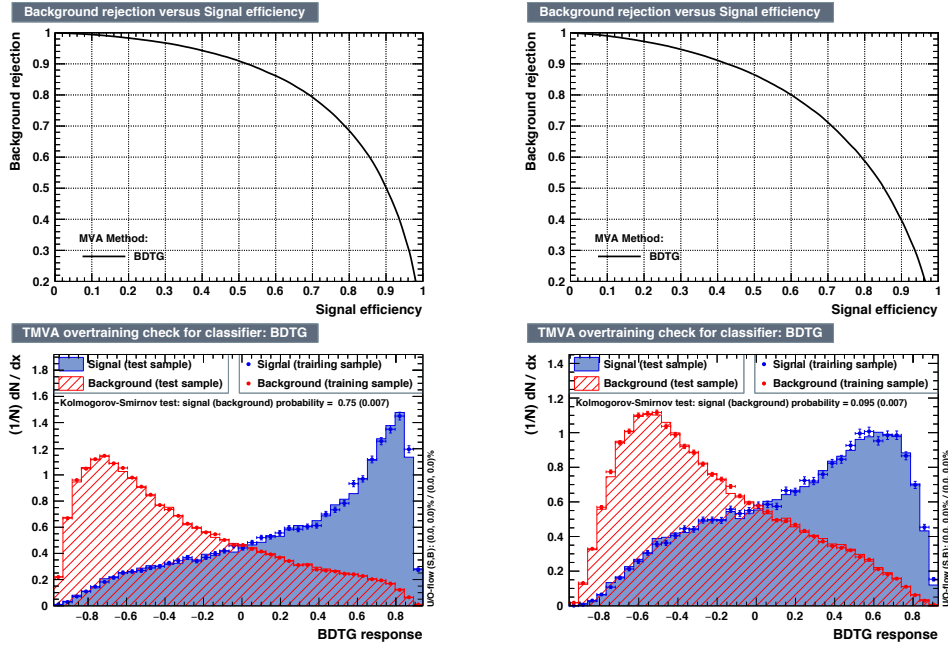


Figure A.8: Top: background rejection vs signal efficiency (ROC curve) in the $2\ell_{ss}$ channel for BDTG discriminator, against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Bottom: classifier output distribution, for training against $t\bar{t}V$ (left) and against $t\bar{t}$ (right).

Table A.3: TMVA input variables ranking for BDTG method in the 3ℓ channel.

	$t\bar{t}$ training		$t\bar{t}V$ training	
Rank	Variable	Importance	Variable	Importance
1	minDRll	1.329×10^{-1}	dEtaFwdJetBJet	1.264×10^{-1}
2	dEtaFwdJetClosestLep	1.294×10^{-1}	Lep3Pt	1.224×10^{-1}
3	dEtaFwdJetBJet	1.209×10^{-1}	maxEtaJet25	1.221×10^{-1}
4	dPhiHighestPtSSPair	1.192×10^{-1}	dEtaFwdJet2BJet	1.204×10^{-1}
5	Lep3Pt	1.158×10^{-1}	dEtaFwdJetClosestLep	1.177×10^{-1}
6	maxEtaJet25	1.121×10^{-1}	minDRll	1.143×10^{-1}
7	dEtaFwdJet2BJet	9.363×10^{-2}	dPhiHighestPtSSPair	9.777×10^{-2}
8	nJetEta1	6.730×10^{-2}	nJet25_Recl	9.034×10^{-2}
9	nJet25_Recl	6.178×10^{-2}	nJetEta1	4.749×10^{-2}
10	lepCharge	4.701×10^{-2}	lepCharge	4.116×10^{-2}

Table A.4: TMVA input variables ranking for BDTG method in the $2\ell ss$ channel.

Rank	$t\bar{t}$ training		$t\bar{t}V$ training	
	Variable	Importance	Variable	Importance
1	dEtaFwdJetClosestLep	1.394×10^{-1}	maxEtaJet25	1.357×10^{-1}
2	minDRll	1.359×10^{-1}	dEtaFwdJet2BJet	1.267×10^{-1}
3	maxEtaJet25	1.308×10^{-1}	dEtaFwdJetBJet	1.200×10^{-1}
4	dPhiHighestPtSSPair	1.116×10^{-1}	Lep2Pt	1.196×10^{-1}
5	Lep2Pt	1.111×10^{-1}	dEtaFwdJetClosestLep	1.145×10^{-1}
6	dEtaFwdJetBJet	1.067×10^{-1}	minDRll	1.077×10^{-1}
7	dEtaFwdJet2BJet	8.906×10^{-2}	nJet25_Recl	1.020×10^{-1}
8	nJetEta1	6.445×10^{-2}	dPhiHighestPtSSPair	8.232×10^{-2}
9	nJet25_Recl	6.254×10^{-2}	nJetEta1	5.948×10^{-2}
10	lepCharge	4.848×10^{-2}	lepCharge	3.198×10^{-2}

A.2 Training with 2017 samples

Here also for all trainings, the signal is tHq in the ITC scenario ($\kappa_t = -1.0$). The TMVA settings used in the BDT training are shown in Table A.5.

Table A.5: TMVA configuration used in the training.

```

TMVA.Types.kBDT
NTrees=200
BoostType=Grad
Shrinkage=0.10
nCuts=40
MaxDepth=4
NegWeightTreatment=PairNegWeightsGlobal

```

Background samples differ from the earlier training in generator settings and statistics which affects the training. Figures A.9 and A.10 show the input variables distributions as seen by the MVA algorithm for the 3ℓ channel in the two different trainings. The input variables distributions for $2\ell ss$ channel for signal against the $t\bar{t}$ and $t\bar{t}V$ are shown in Figs A.11 and A.12 respectively.

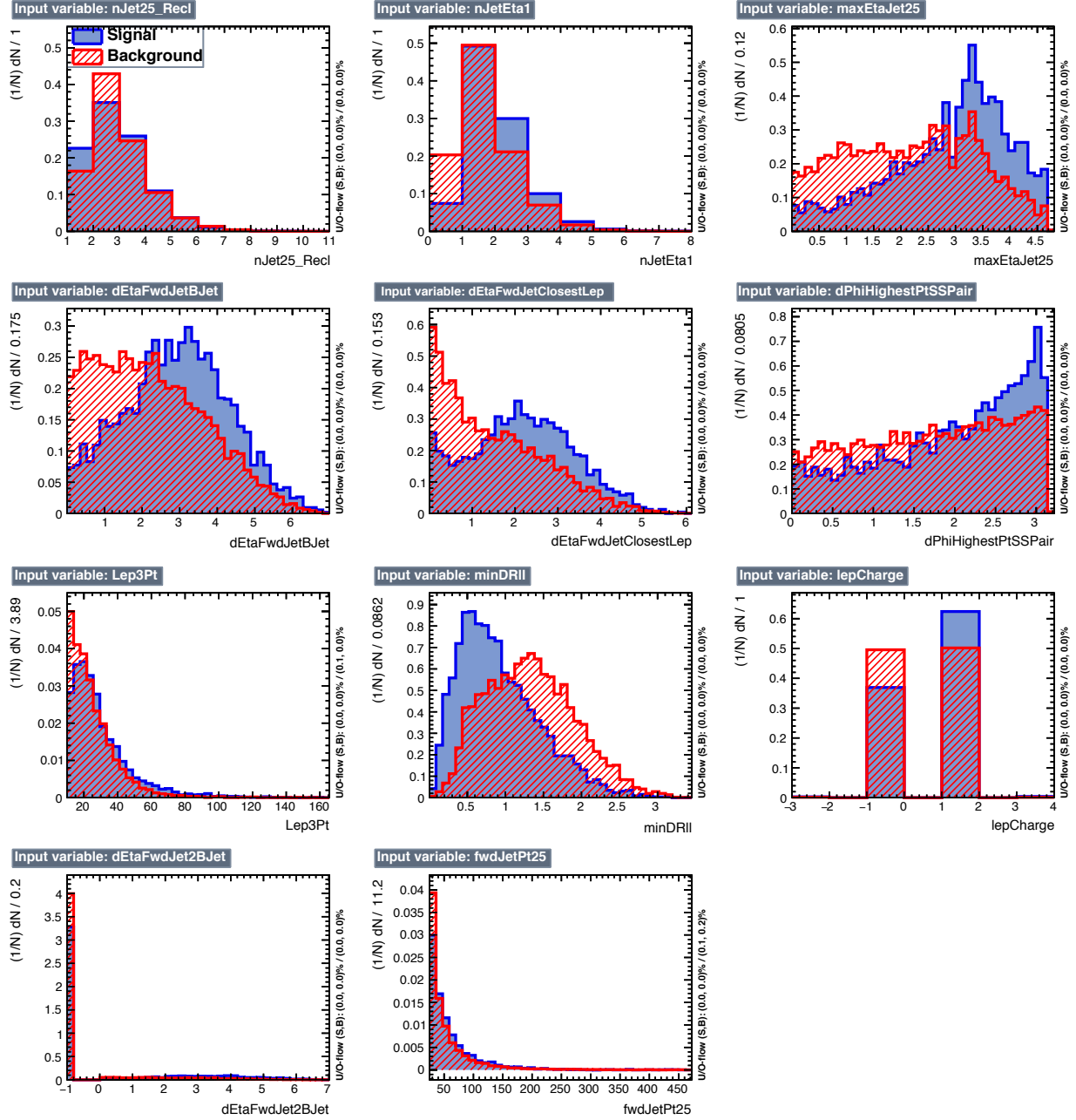


Figure A.9: BDT inputs as seen by TMVA (signal in blue, background in red) for the 3ℓ channel, discriminated against $t\bar{t}$ (non-prompt) background.

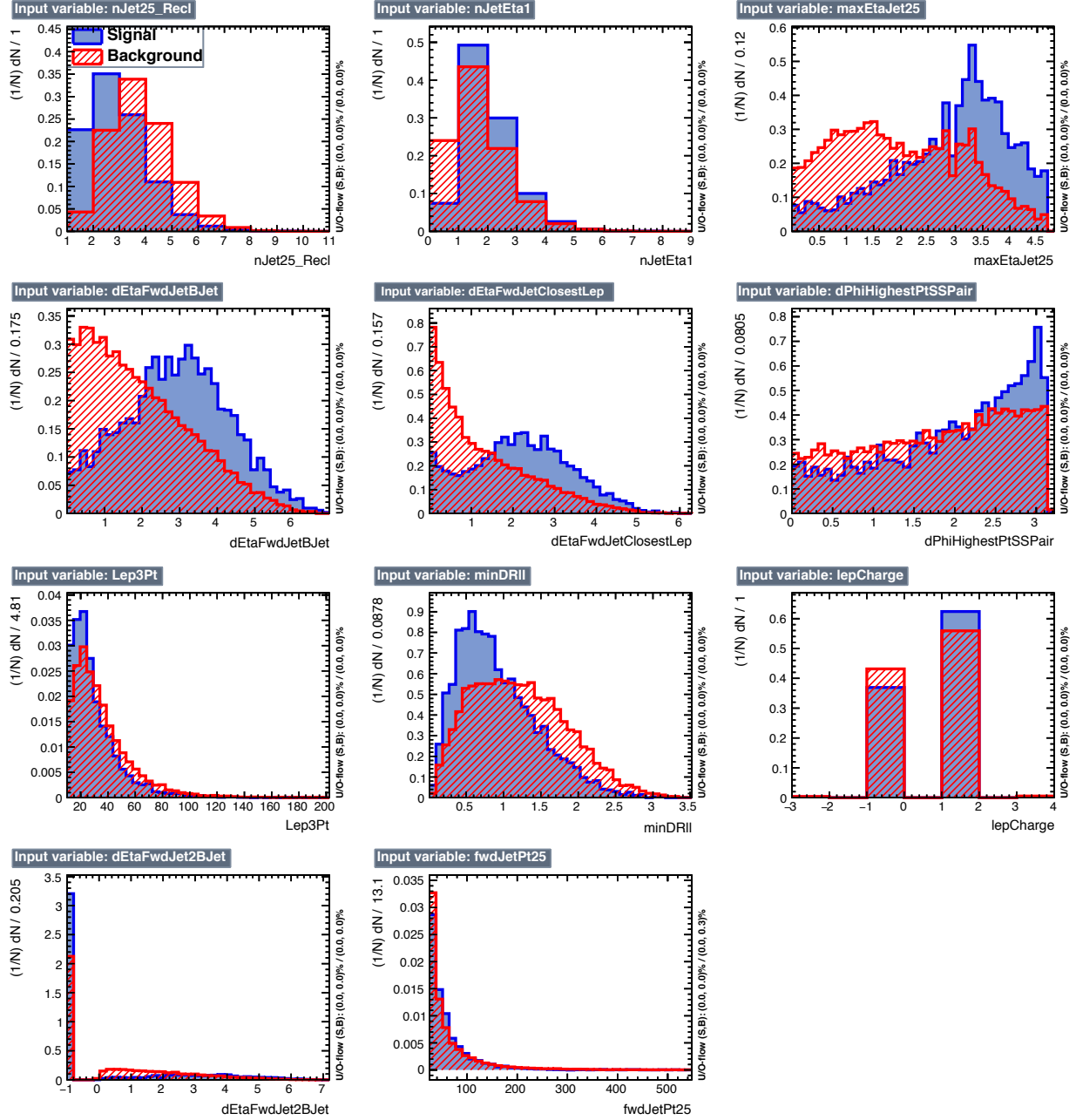


Figure A.10: BDT inputs as seen by TMVA (signal in blue, background in red) for the 3ℓ channel, discriminated against $t\bar{t}V$ (prompt) background.

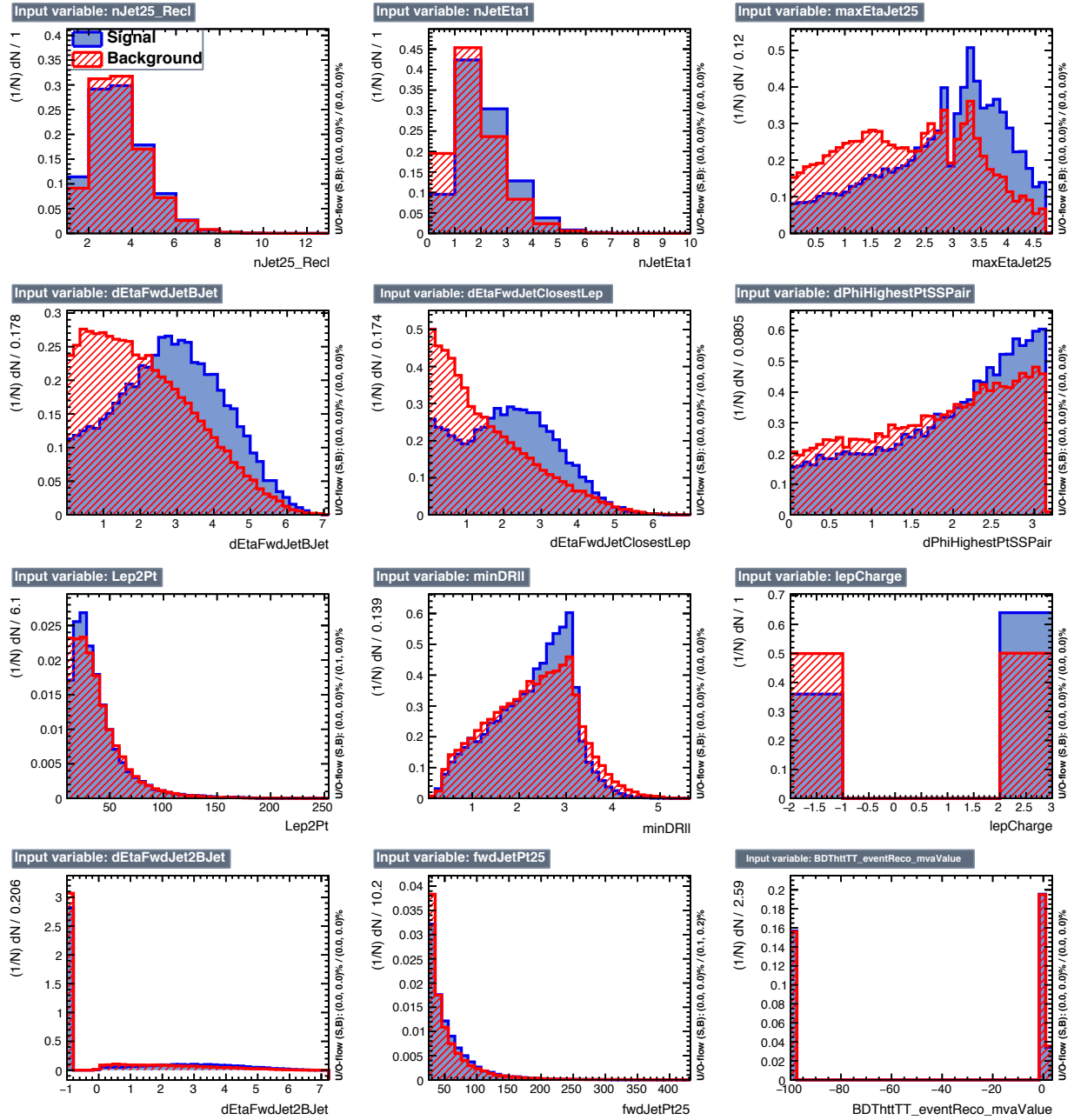


Figure A.11: BDT inputs as seen by TMVA (signal in blue, background in red) for the $2\ell ss$ channel, discriminated against $t\bar{t}$ (non-prompt) background.

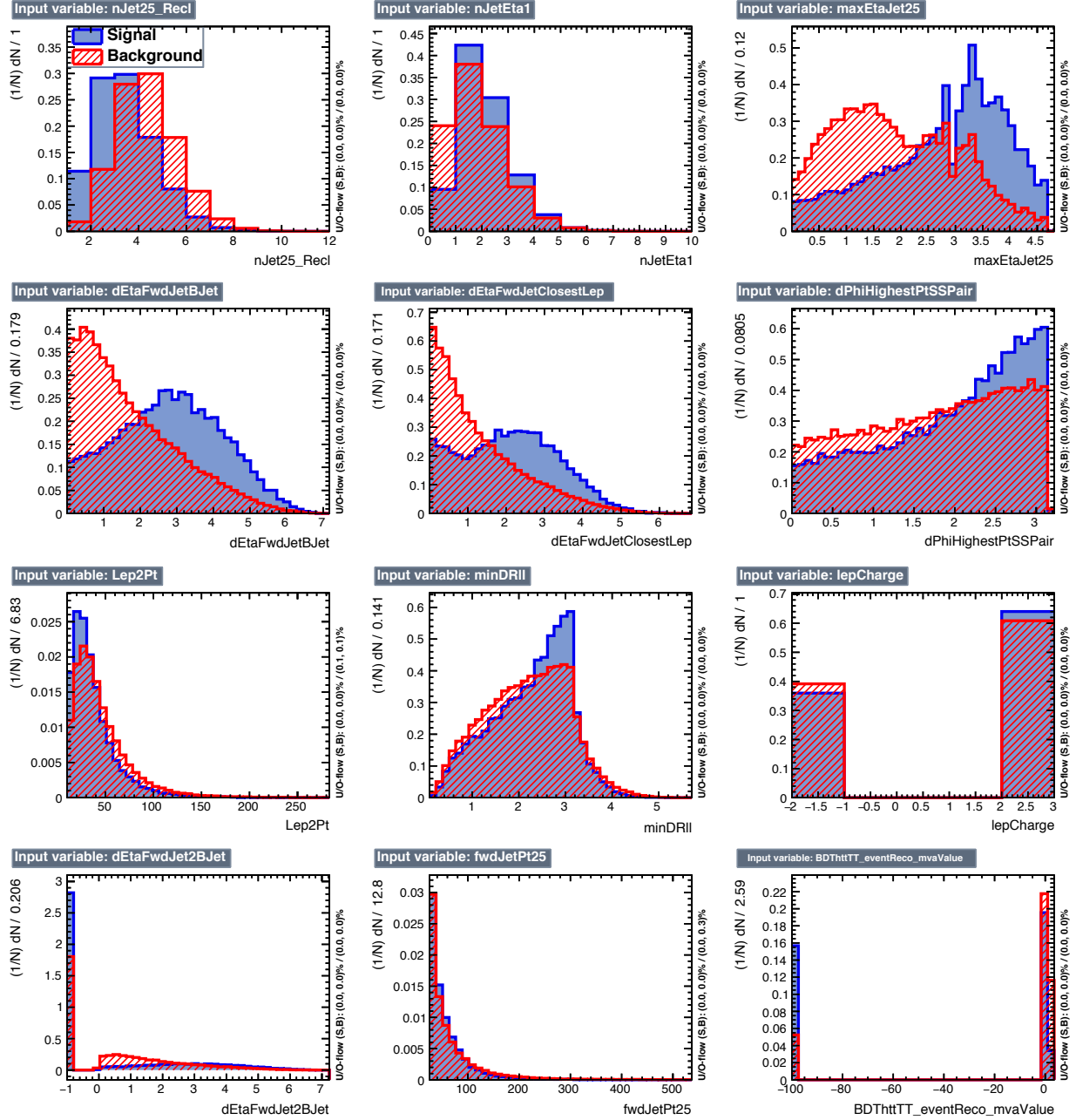


Figure A.12: BDT inputs as seen by TMVA (signal in blue, background in red) for the $2\ell ss$ channel, discriminated against $t\bar{t}V$ (prompt) background.

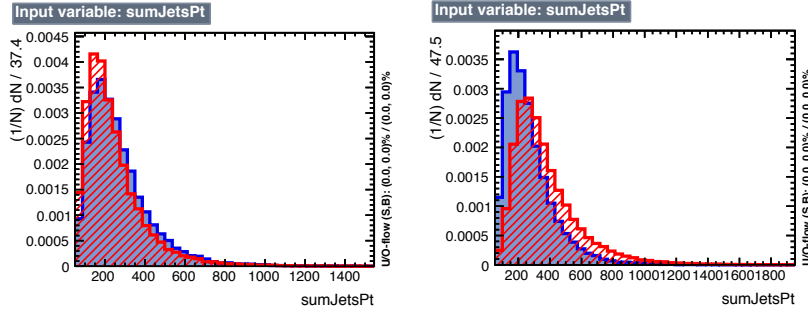


Figure A.13: The SumJetsPt variable input to BDT training as seen by TMVA (signal in blue, background in red) for the $2\ell ss$ channel, discriminated against $t\bar{t}$ (non-prompt) in the left and $t\bar{t}V$ (prompt) in the right. Clearly this variable has more discrimination power against the prompt background, which has more selected jet multiplicity.

A.2.1 Input variables correlation

Figures A.14 and A.15 show the linear correlation coefficients for signal and background for the two training cases in 3ℓ and $2\ell ss$ channels respectively. As expected, strong correlations appears for variables related to the forward jet activity. Same trend is seen in case of the $2\ell ss$ channel in Figure. [h!]

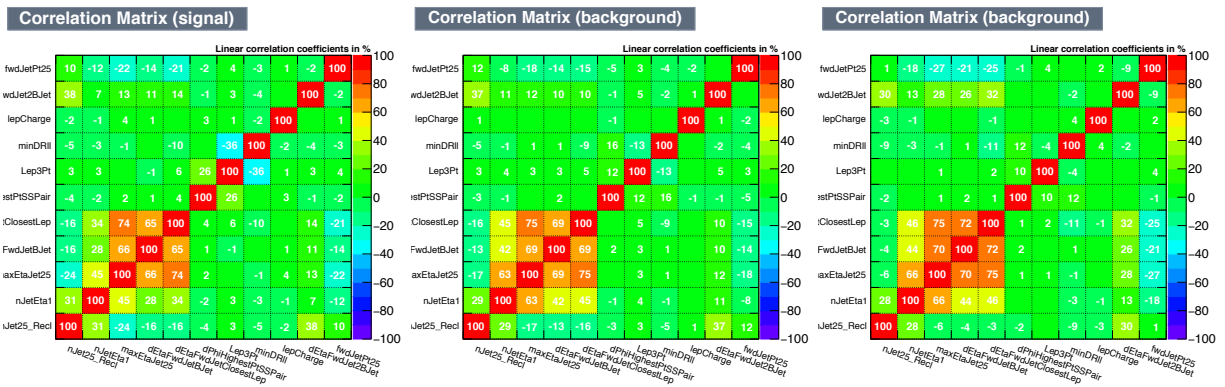


Figure A.14: Signal (left), $t\bar{t}$ background (middle), and $t\bar{t}V$ background (right.) correlation matrices for the input variables in the TMVA analysis for the 3ℓ channel.

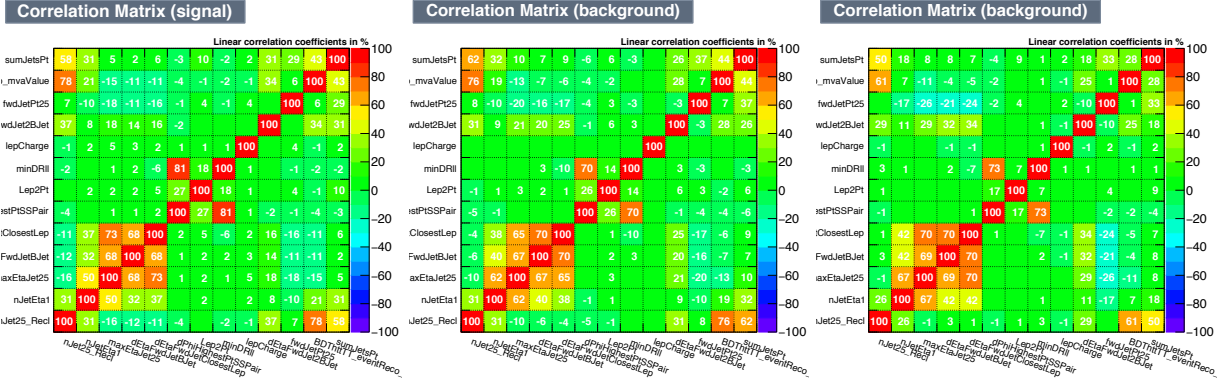


Figure A.15: Signal (left), $t\bar{t}$ background (middle), and $t\bar{t}V$ background (right.) correlation matrices for the input variables in the TMVA analysis for the $2\ell ss$ channel.

A.2.2 Classifier response

Figures A.16 and A.17 show the ROC curves and the classifier output distributions for signal and backgrounds. The ROC integrals are 0.804 (0.736) and 0.855 (0.838) for training against $t\bar{t}$ and $t\bar{t}V$ respectively in 3ℓ ($2\ell ss$) channel. [h!]

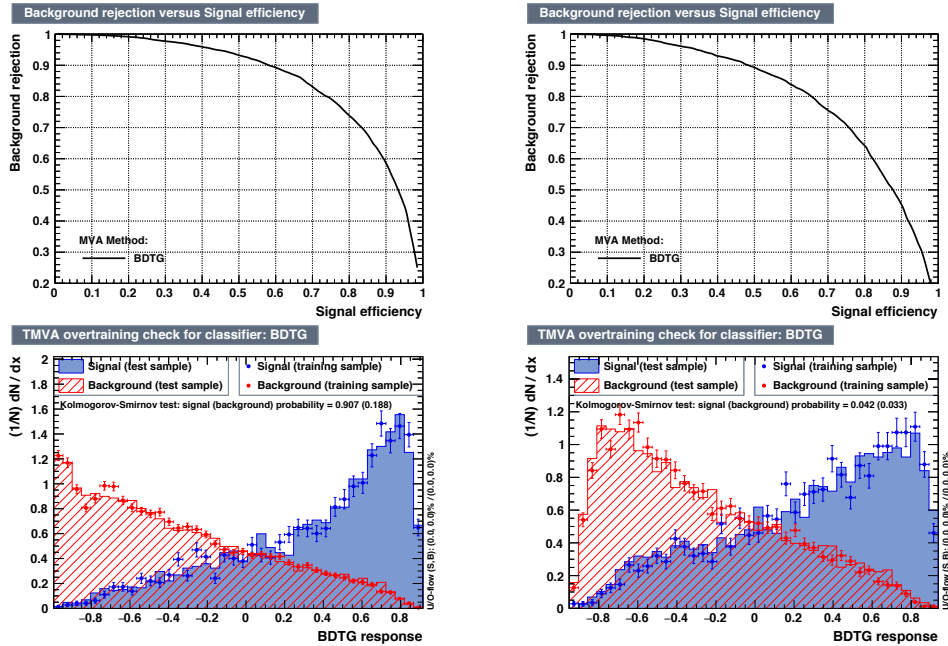


Figure A.16: Top: ROC curves for BDTG classifiers (top) in the 3ℓ channel against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Bottom: classifier output distributions for training against $t\bar{t}V$ (left) and against $t\bar{t}$ (right).

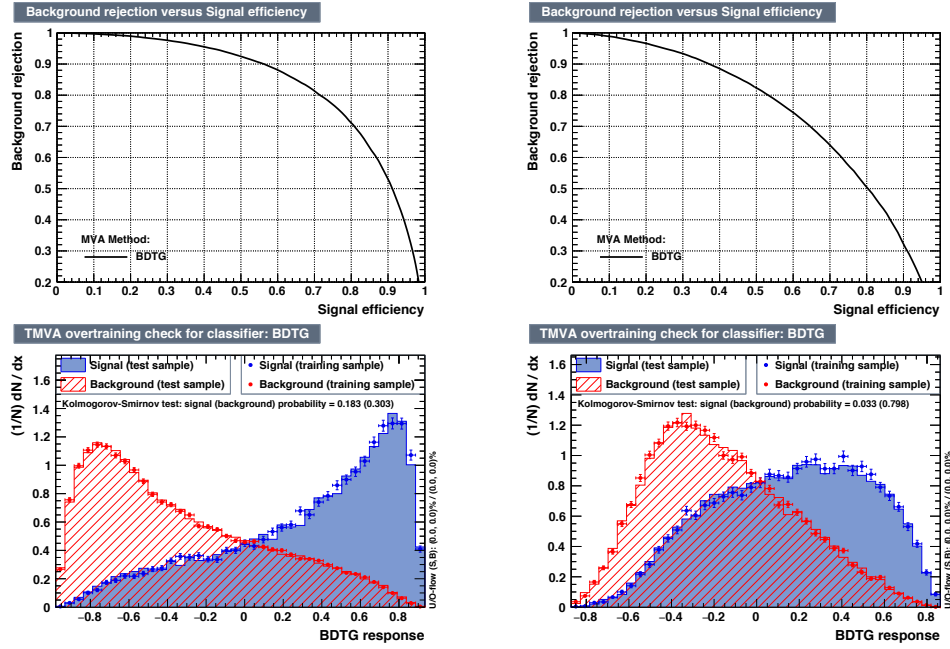


Figure A.17: Top: ROC curve in the $2\ell ss$ channel for BDTG discriminator, against $t\bar{t}V$ (left) and $t\bar{t}$ (right). Bottom: classifier output distribution, for training against $t\bar{t}V$ (left) and against $t\bar{t}$ (right).

The ranking of input variables according to importance in the discrimination is listed in Tables [A.6](#) and [A.7](#).

Table A.6: TMVA input variables ranking for BDTG method in the 3ℓ channel.

	$t\bar{t}$ training		$t\bar{t}V$ training	
Rank	Variable	Importance	Variable	Importance
1	minDRll	1.434×10^{-1}	maxEtaJet25	1.301×10^{-1}
2	maxEtaJet25	1.108×10^{-1}	Lep3Pt	1.149×10^{-1}
3	fwdJetPt25	1.032×10^{-1}	nJet25	1.135×10^{-1}
4	dPhiHighestPtSSPair	1.001×10^{-1}	fwdJetPt25	1.078×10^{-1}
5	dEtaFwdJetClosestLep	9.736×10^{-2}	minDRll	1.063×10^{-1}
6	Lep3Pt	9.369×10^{-2}	dPhiHighestPtSSPair	8.804×10^{-2}
7	dEtaFwdJet2BJet	8.341×10^{-2}	dEtaFwdJetBJet	8.734×10^{-2}
8	dEtaFwdJetBJet	8.093×10^{-2}	dEtaFwdJetClosestLep	8.593×10^{-2}
9	nJet25	7.303×10^{-2}	dEtaFwdJet2BJet	8.326×10^{-2}
10	lepCharge	6.200×10^{-2}	lepCharge	4.372×10^{-2}
11	nJetEta1	5.209×10^{-2}	nJetEta1	3.919×10^{-2}

Table A.7: TMVA input variables ranking for BDTG method in the $2\ell ss$ channel.

	$t\bar{t}$ training		$t\bar{t}V$ training	
Rank	Variable	Importance	Variable	Importance
1	dEtaFwdJetClosestLep	1.362×10^{-1}	maxEtaJet25	1.100×10^{-1}
2	minDRll	1.228×10^{-1}	dEtaFwdJetBJet	1.043×10^{-1}
3	Lep2Pt	1.063×10^{-1}	fwdJetPt25	1.034×10^{-1}
4	dEtaFwdJetBJet	1.013×10^{-1}	sumJetsPt	9.917×10^{-2}
5	maxEtaJet25	9.704×10^{-2}	minDRll	9.693×10^{-2}
6	dPhiHighestPtSSPair	9.432×10^{-2}	Lep2Pt	9.447×10^{-2}
7	dEtaFwdJet2BJet	8.179×10^{-2}	dEtaFwdJet2BJet	8.999×10^{-2}
8	fwdJetPt25	7.825×10^{-2}	dPhiHighestPtSSPair	8.262×10^{-2}
9	lepCharge	5.922×10^{-2}	dEtaFwdJetClosestLep	7.234×10^{-2}
10	sumJetsPt	5.072×10^{-2}	BDThttTT	5.286×10^{-2}
11	nJetEta1	3.302×10^{-2}	nJet25	4.053×10^{-2}
12	nJet25	3.299×10^{-2}	nJetEta1	2.933×10^{-2}
13	BDThttTT	6.078×10^{-3}	lepCharge	2.411×10^{-2}

Appendix B

Control region distributions

The distributions of the input variables used to compute the BDT discriminators are inspected in the control regions of the selection. Control regions are defined with selections similar to the signal regions, but with some cuts inverted or modified in order to enhance the contributions from specific background processes.

B.1 Two leptons same-sign

In the case of $2\ell ss$ channel, the main background is $t\bar{t}$. The background contribution from $t\bar{t}$ is enhanced in a dedicated region by requiring no un-tagged jet beyond $|\eta| > 2.0$. In addition, to blind the control region to the $t\bar{t}H$ process, events with four or more central ($|\eta| < 2.4$) jets are rejected. Figs. [B.1](#) and Figs. [B.2](#) show some of the kinematic distributions in this $t\bar{t}$ enriched selection for the $2\ell ss \mu^\pm\mu^\pm$ channel. Similarly, distributions in the $t\bar{t}$ enriched selection for the $2\ell ss e^\pm\mu^\pm$ region are shown in Figs. [B.1](#) and [B.2](#).

B.2 Three leptons

The three leptons signal selection is enriched with Z events from the $t\bar{t}Z$ and WZ processes by requiring $|m_{\ell\ell} - 91.2| < 10$ GeV. Furthermore, no un-tagged jet beyond $|\eta| > 2.0$ should be present. Figures [B.5](#) and [B.6](#) show some of the kinematic distributions in this selection for 2016 and 2017 analyses respectively.

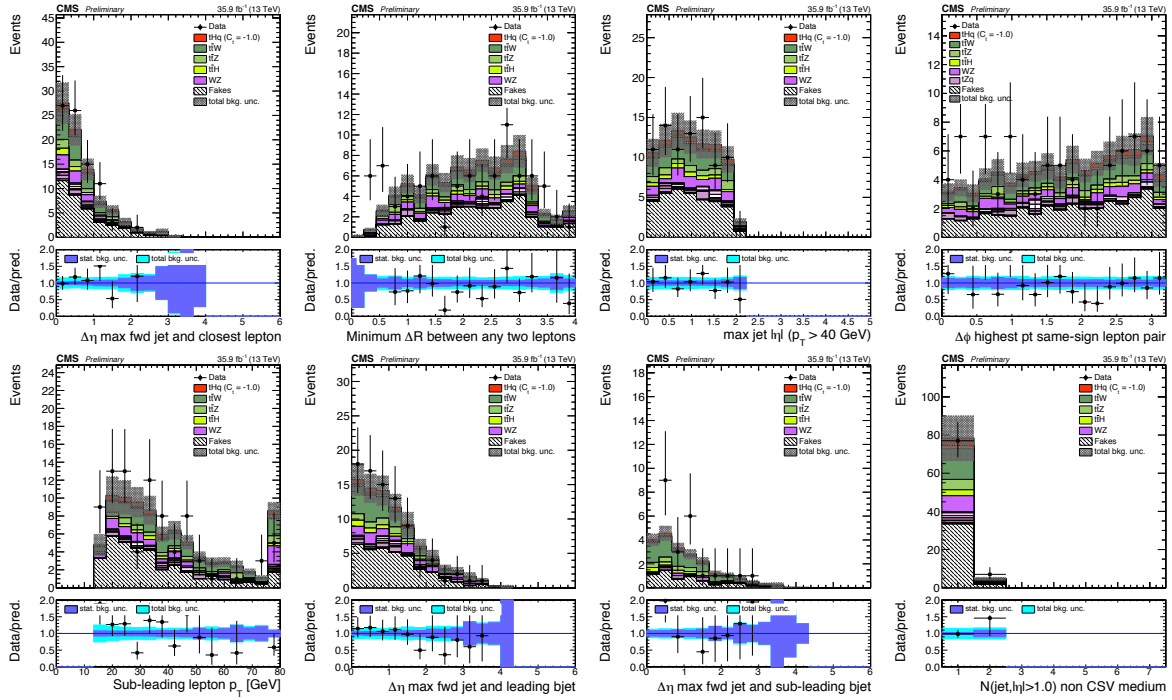


Figure B.1: Contributions from signal and background in $t\bar{t}$ control region for $\mu^{\pm}\mu^{\pm}$ channel in the 2016 analysis.

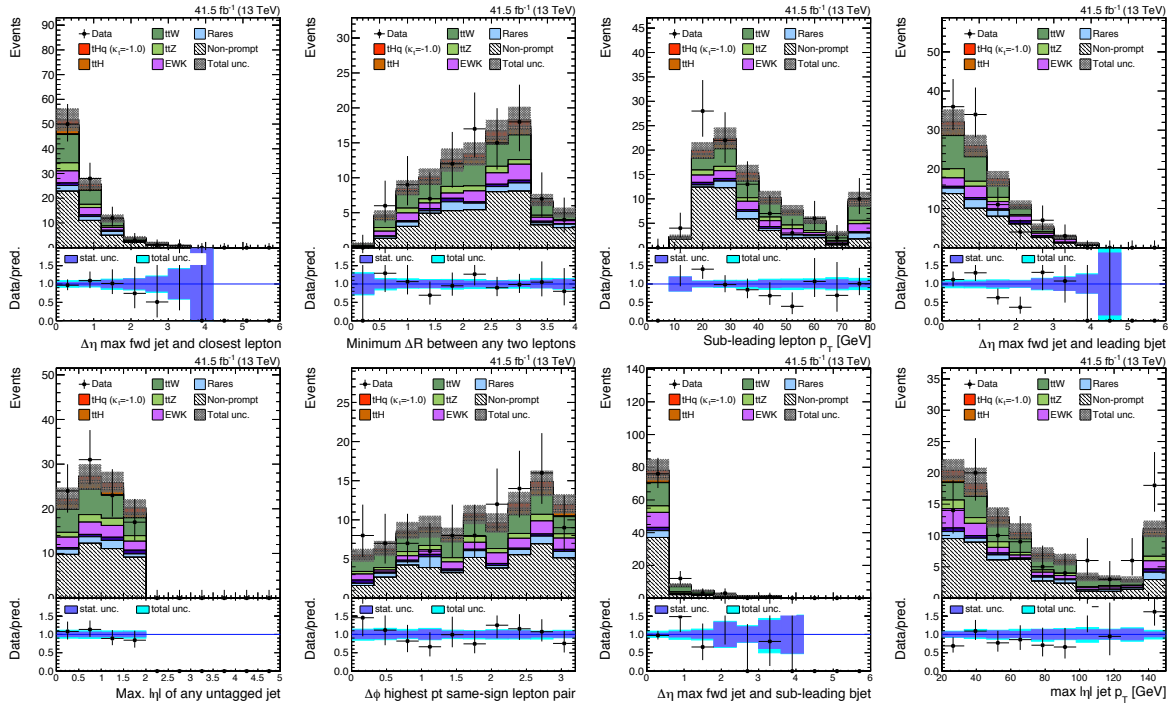


Figure B.2: Contributions from signal and background in $t\bar{t}$ control region for $\mu^{\pm}\mu^{\pm}$ channel in the 2017 analysis.

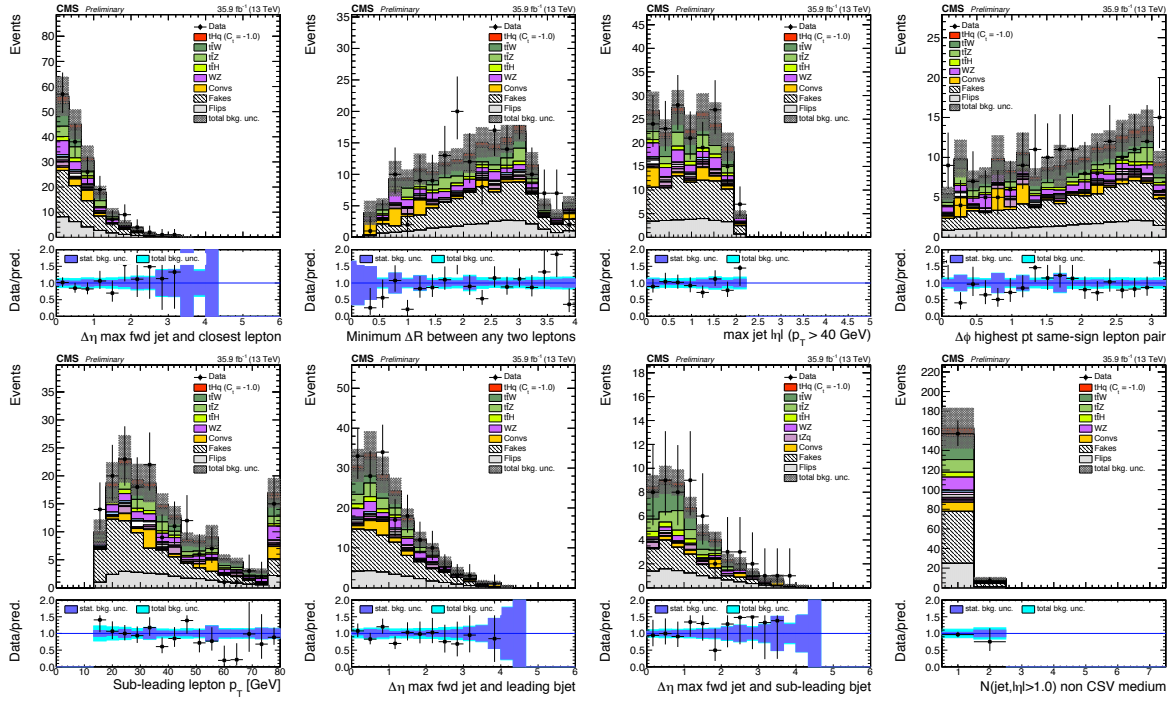


Figure B.3: Contributions from signal and background in $t\bar{t}$ control region for $e^{\pm}\mu^{\pm}$ channel in the 2016 analysis.

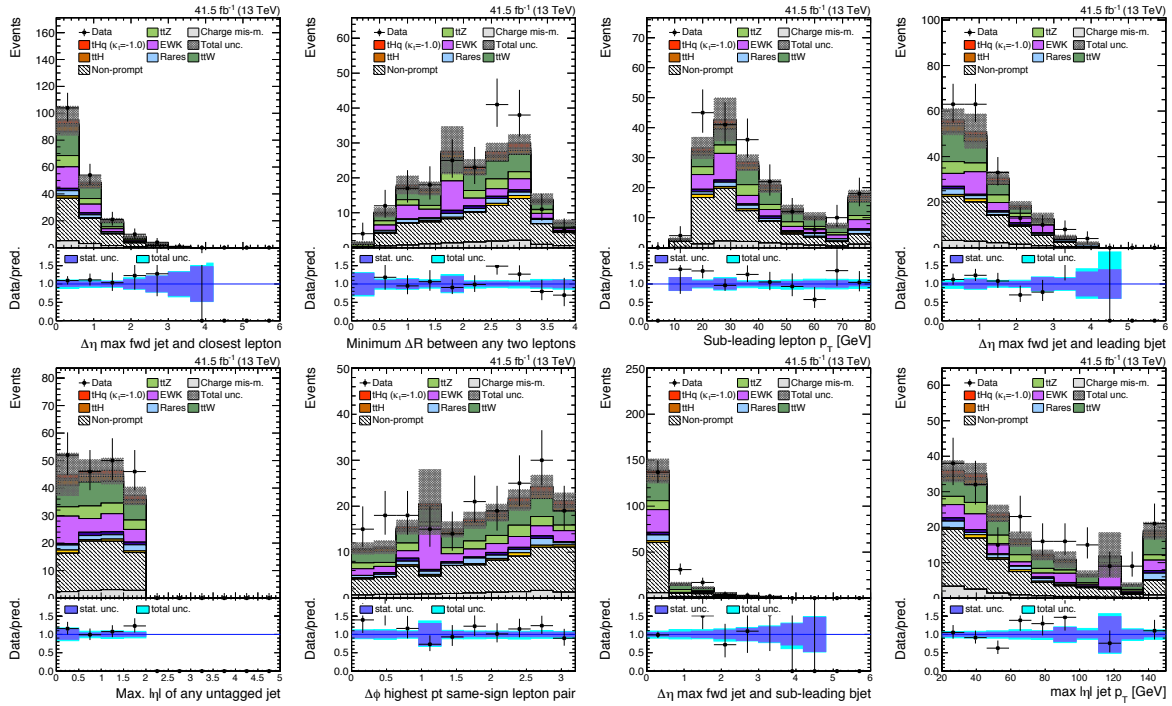


Figure B.4: Contributions from signal and background in $t\bar{t}$ control region for $e^{\pm}\mu^{\pm}$ channel in the 2017 analysis.

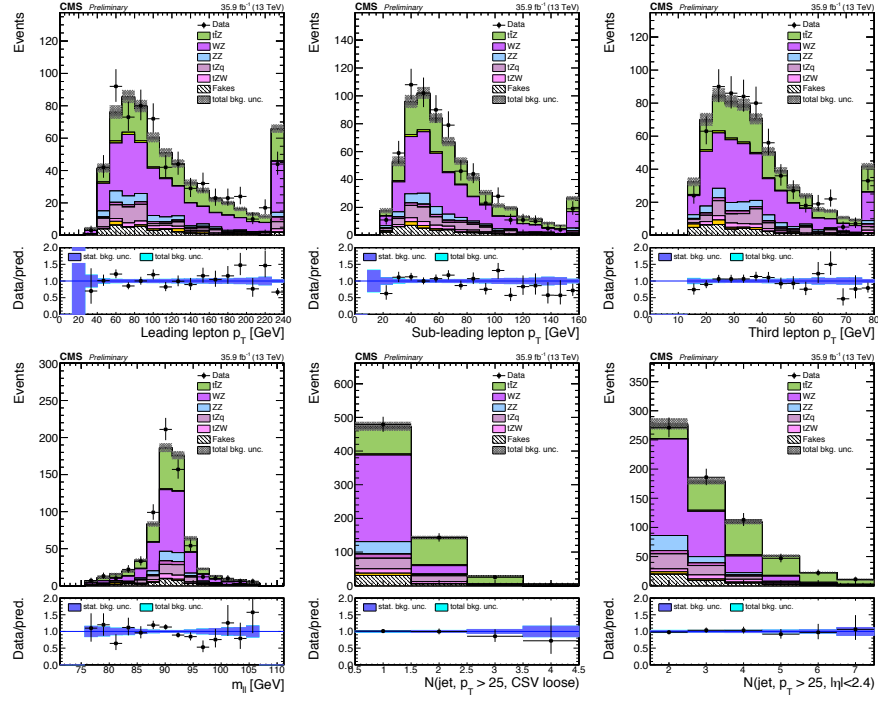


Figure B.5: Distributions of input variables to the BDT for signal discrimination, in Z enriched 3ℓ selection in the 2016 analysis.

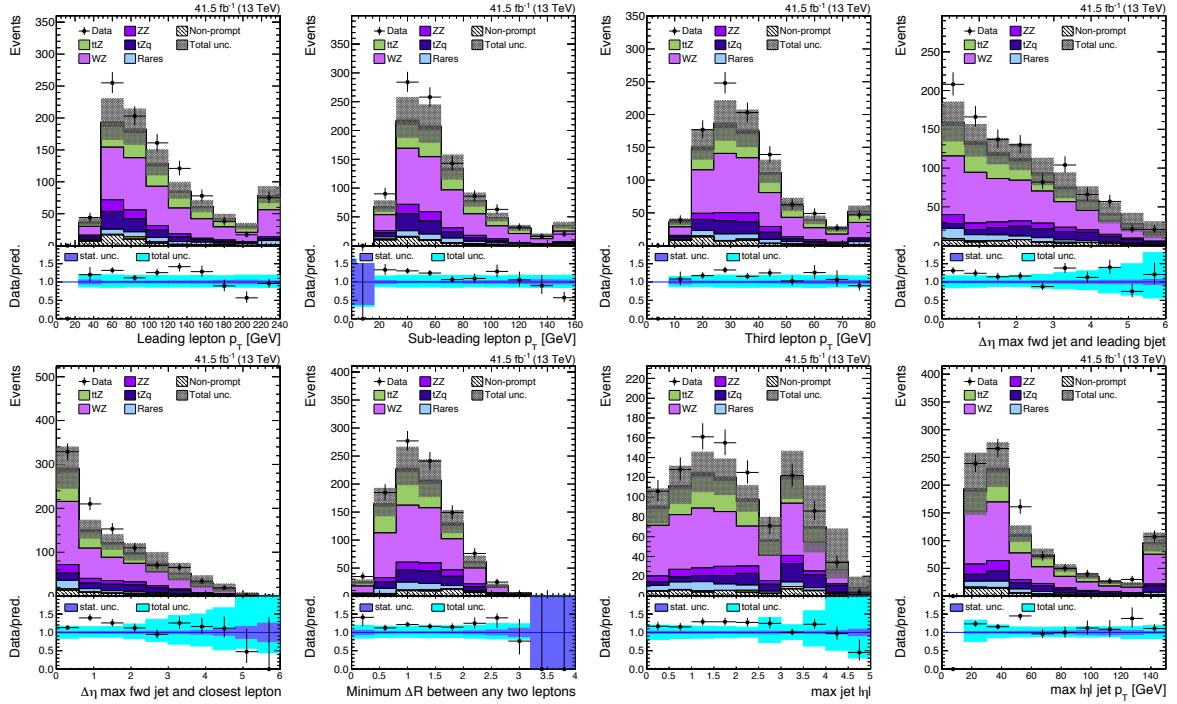


Figure B.6: Distributions of input variables to the BDT for signal discrimination, in Z enriched 3ℓ selection in the 2017 analysis.

B.3 Two leptons opposite-sign

The two leptons opposite-sign control region is identical to the $e^\pm\mu^\pm$ signal region, except for the requirement of oppositely-charged leptons. Kinematic distributions related to the most forward light jet are studied are compared for the observed data and the $t\bar{t}$ MC predictions as shown in Fig. B.7. The agreement of the η distribution of forward jets for a p_T cut of 25 GeV is poor, especially at higher values of $|\eta|$. The (central) jet multiplicity is observed to be poorly modelled by the MC sample. Note that the $t\bar{t}$ background in this analysis is estimated with a data-driven method and these disagreements do not directly affect the $t\bar{t}$ contribution in the analysis. They do however reflect the expected agreement in these distributions for the irreducible backgrounds and the signal.

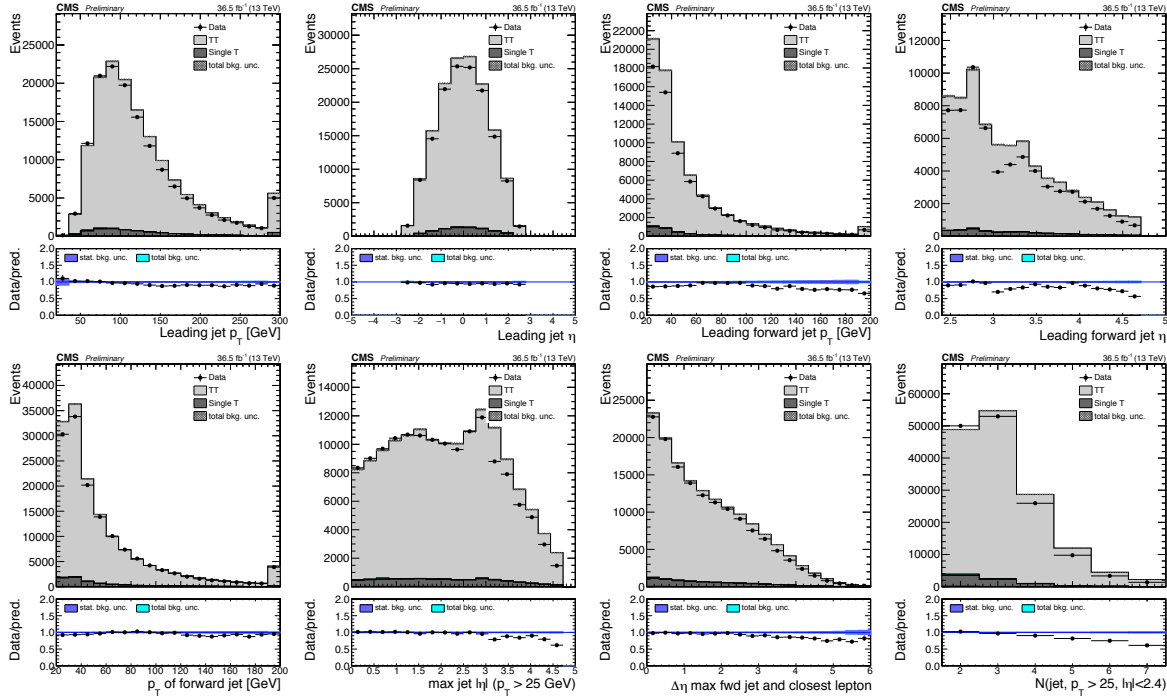


Figure B.7: Kinematic distributions in the $t\bar{t}$ -enriched opposite-sign $e\mu$ selection in 2016 analysis. Top row, left to right: p_T and $|\eta|$ of leading central ($\eta < 2.4$) jet; p_T and $|\eta|$ of leading forward ($\eta > 2.4$) jet. Bottom row: p_T and $|\eta|$ of most forward un-tagged jet (“light forward jet”); $\Delta\eta$ between light forward jet and closest lepton; number of central jets.

The data/MC agreement in the forward jet η distribution improves significantly at higher jet p_T s. The effect of higher p_T cuts on the forward jet has been studied for three values:

25, 30 and 40 GeV. In order to take into account the data/MC disagreement in the high η regions, the events are weighted accordingly to the data/MC ratio of the unity normalized control plots shown in Fig. B.8. Table B.1 shows the scale factors obtained for the three

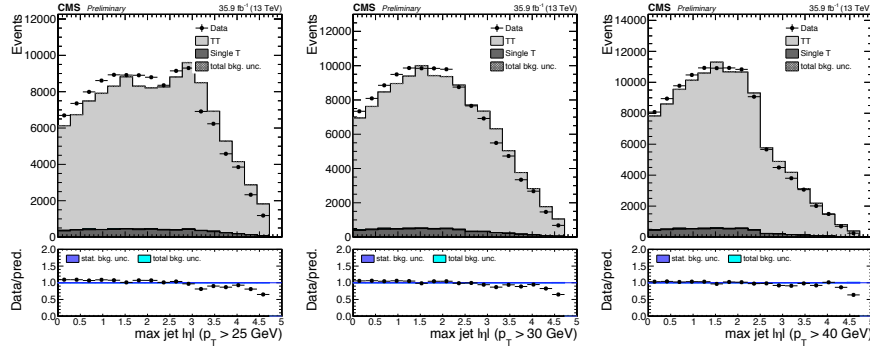


Figure B.8: Pseudorapidity of the light forward jet in the $t\bar{t}$ -enriched opposite-sign $e\mu$ selection for the three p_T cut values in 2016 analysis.

p_T values. With higher cuts in p_T , the expected limit on signal strength in the three lepton channel is seen to improve from 25 GeV to 40 GeV. The impact of the data/MC disagreement for forward jet η is observed to reduce with higher p_T cuts.

In the ECAL-HCAL transition region known to be “noisy”, a large data/MC disagreement is observed in the region $2.7 < |\eta| < 3.0$ in the 2017 data, which is not properly accounted for by the JEC factors. In order to mitigate this issue, several options were investigated :

- Option 0 : All selected jets with $p_T > 25$ GeV are used (default),
- Option 1 : Jets in the region $|\eta| > 2.4$ must satisfy an additional $p_T > 40$ GeV cut (as in the 2016 analysis),
- Option 2 : Jets in the region $2.7 < |\eta| < 3.0$ are rejected,
- Option 3 : Jets in the region $2.7 < |\eta| < 3.0$ must satisfy an additional $p_T > 60$ GeV cut (as done e.g. in Ref. [160]).

The distribution of the pseudorapidity of the most forward light jet obtained for each option is shown in Fig. B.9. As the option 3 mitigates the issue while retaining most signal events, and considering the fact that this solution was already adopted by an analysis

Table B.1: Data/MC scale factors for η distribution of light forward jet with three different p_T cuts.

$ \eta $ range	$p_T > 25$ GeV	$p_T > 30$ GeV	$p_T > 40$ GeV
0 – 0.278	1.0925	1.0566	1.0326
0.278 – 0.556	1.0920	1.0617	1.0407
0.556 – 0.833	1.0675	1.0459	1.0244
0.833 – 1.111	1.0888	1.0593	1.0340
1.111 – 1.389	1.0759	1.0508	1.0322
1.389 – 1.667	1.0109	0.9847	0.9661
1.667 – 1.944	1.0727	1.0448	1.0239
1.944 – 2.222	1.0715	1.0457	1.0169
2.222 – 2.500	1.0112	0.9871	0.9746
2.500 – 2.778	1.0387	0.9942	0.9816
2.778 – 3.056	0.9687	0.9427	0.9200
3.056 – 3.333	0.8137	0.8695	0.9092
3.333 – 3.611	0.9010	0.9387	0.9807
3.611 – 3.889	0.8685	0.8887	0.9213
3.889 – 4.167	0.9277	0.9466	1.0135
4.167 – 4.444	0.8111	0.8278	0.8637
4.444 – 4.722	0.6497	0.6485	0.6367
4.722 – 5.000	1.0000	1.0000	1.0000

targeting a similar final state [160] and using the same dataset, it was adopted for this analysis. The expected upper limits on the signal strength were also checked to be better in case of option 3 compared to other three options. The data/MC scale factors derived based on the distributions shown in Fig. B.9 are listed in Table B.2.

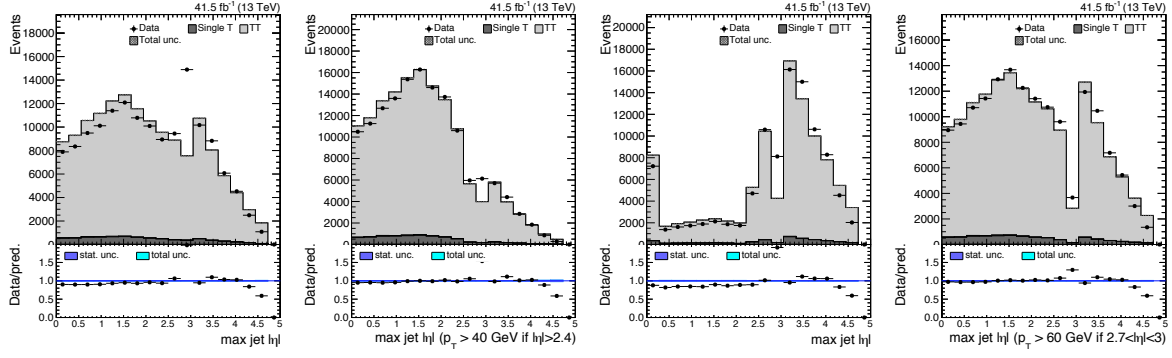


Figure B.9: Pseudorapidity of the light forward jet in $t\bar{t}$ enriched opposite sign $e\mu$ selection in 2017 analysis. From left to right: Option 0, Option 1, Option 2 and Option 3 respectively.

Table B.2: Data/MC scale factors for different requirements on the light forward jet.

$ \eta $ range	Option 0	Option 1	Option 2	Option 3
0 – 0.278	0.8876	0.9504	0.8773	0.9715
0.278 – 0.556	0.8845	0.9541	0.8185	0.9639
0.556 – 0.833	0.8854	0.9482	0.8447	0.9647
0.833 – 1.111	0.8922	0.9578	0.8459	0.9721
1.111 – 1.389	0.9186	0.9906	0.8408	1.0045
1.389 – 1.667	0.9347	1.0003	0.8972	1.0176
1.667 – 1.944	0.9189	0.9896	0.8670	1.0046
1.944 – 2.222	0.9451	1.0188	0.8876	1.0235
2.222 – 2.500	0.9260	0.9843	0.8941	1.0121
2.500 – 2.778	1.0786	1.0564	1.0149	1.0744
2.778 – 3.056	2.0302	1.5379	1.9081	1.2978
3.056 – 3.333	0.9231	0.9840	0.9533	0.9398
3.333 – 3.611	1.0778	1.1136	1.1167	1.0973
3.611 – 3.889	1.0270	1.0121	1.0622	1.0456
3.889 – 4.167	1.0097	1.0243	1.0601	1.0279
4.167 – 4.444	0.8161	0.8865	0.8323	0.8308
4.444 – 4.722	0.5824	0.5877	0.5956	0.5930
4.722 – 5.000	1.0000	1.0000	1.0000	1.0000