

Florida State University Libraries

2016

Measurement of the Triple Differential Cross Section for Photon+Jet Events with the CMS Detector at a Center of Mass Energy of 8 TeV

Ajeeta Khatiwada

CERN-THESIS-2016-405



FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

MEASUREMENT OF THE TRIPLE DIFFERENTIAL CROSS SECTION FOR PHOTON+JET
EVENTS WITH THE CMS DETECTOR AT A CENTER OF MASS ENERGY OF 8 TEV

By

AJEETA KHATIWADA

A Dissertation submitted to the
Department of Physics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2016

Ajeeta Khatiwada defended this dissertation on June 3, 2016.
The members of the supervisory committee were:

Andrew Askew
Professor Co-Directing Dissertation

Susan Blessing
Professor Co-Directing Dissertation

Qing-Xiang Sang
University Representative

Todd Adams
Committee Member

Joseph Owens
Committee Member

Jorge Piekarewicz
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

To my parents

ACKNOWLEDGMENTS

This work is not mine alone, but is a culmination of direct and indirect contribution from many of my teachers, colleagues and my family. First and foremost, I am indebted to my advisor Prof. Andrew Askew for being available at any moment when I needed his support. He was always on top of everything, whether it be writing recommendation letters, informing me about career opportunities or giving helpful advice. As I leave FSU, I will miss our general discussions and his ever enthusiastic viewpoints on various topics in physics. At the same time, I am equally confident that I will continue to benefit from the lessons he has imparted through his personal achievements and dedication to the field. Similarly, I also owe my gratitude to Prof. Susan Blessing for providing me inputs in several occasions including during the finalization of this dissertation.

I am grateful to the all of the FSU Physics faculty members, some of whom I would like to specifically mention. This work would not have been possible without the support and vast expertise of Prof. Joseph Owens, for which I am deeply appreciative. I also thank Prof. Todd Adams and Prof. Harrison Prosper for countless discussions during FSU-CMS meetings and support in more than one ways. I thank Prof. Jorge Piekarewicz, who is the best teacher I ever had, and a guardian figure. I have deeply enjoyed the classes I took with Prof. Laura Reina, Prof. Efstratios Manousakis, Prof. Takemichi Okui and Prof. Alexander Volya. I also greatly appreciate Prof. Qing-Xiang Sang for taking her invaluable time to be in my committee at the last minute.

I thank the entire CMS collaboration (many of whom I do not know personally) for laying foundation for this work. I am appreciative of the many Standard Model Physics conveners past and present, with a special gratitude to Konstantinos Theofilatos, Simon De Visscher, Emanuela Barberis, and Philippe Gras for their invaluable feedbacks and involvement. I have also benefited from the initial involvement of Giulia D’Imperio, Paulo Meridiani and Chiara Rovelli in this research. I thank my fellow graduate students from the FSU-HEP group, particularly Arka, Sam, and Kenji for discussions on physics, programming and sometimes helping me lighten up with conversations on random subjects. I am also appreciative of Kathy for providing diversion from the work through her pleasant personality and for making multiple travel arrangements for me. I would be remiss if

I did not mention the computational support Joe and Kurtis have provided countless number of times during the course of graduate school.

I have deep respect for some teachers and mentors I have had prior to graduate school, without whose guidance I would not have reached to my present state. Joelle Murray (Linfield College), Kostas Kalogerakis (SRI International), and Sahar Allam (Fermilab) come to my mind instantly. I also thank members of DØ collaboration, particularly Douglas Orbaker, Andreas Jung and Dmitry Bandurin from whom I learned many things at the beginning of my graduate research work.

Perhaps the most important contributions come from my mom, Bishnu, and my dad, Ghanashyam, in planting the importance of education in my head since very early on. They went through a lot of economic and social hardship to earn higher education degree. They devoted their lives to seeing my sister and I become independent and accomplished individuals, for which I can never truly pay them back. They also gave me my best friend in the form of my elder sister, who has been a role model to me. I continue to get awe-inspired by her determination and strength. Rakshya, I cannot thank you enough for being there with me during my formative years to watch the night sky, formulate childish hypothesis on some of the greatest mysteries in physics, laugh together, cry together and most importantly keep my priorities in focus. Through my sister I have found a great friend in my brother-in-law, Tim. He has been a massive support in helping me understand the theory concepts in high energy physics. I thank him for constantly inspiring, and encouraging me. I am also thankful to Tim's parents, Dan and Eileen Hobbs, for being a constant source of moral support. My life partner, colleague, and soul mate, Sean Kuvin, has helped me in this journey in more ways than is possible to enumerate. Whether discussing course work during the first year of graduate school, or discussing physics concepts and programming during later years, I have always benefited in having him by my side. Watching him work endlessly in his own career also inspires me to be a better version of myself. If it was not for the stimulating political, social and philosophical conversations along with some infrequent vacations with him, graduate school would have been less than pleasurable. Lastly, my in-laws, Barbara and Michael Kuvin, have always been caring and supportive. They have been my second sets of parents and I am forever grateful to them for their love.

TABLE OF CONTENTS

List of Tables	viii
List of Figures	xiii
Abstract	xxiv
1 Introduction	1
1.1 Short History of Particle Physics	1
1.2 The Standard Model (SM)	3
1.2.1 Fundamental Particles	4
1.2.2 Interactions	5
2 Theory	8
2.1 Quantum Chromodynamics	8
2.1.1 Perturbative QCD (pQCD)	9
2.1.2 Non-perturbative QCD	11
2.2 Photon+jet physics	14
2.3 Observable	17
3 Experiment	19
3.1 Accelerators	19
3.1.1 The Large Hadron Collider	20
3.1.2 The LHC Luminosity	21
3.2 The Compact Muon Solenoid Detector	23
3.2.1 The Tracker	24
3.2.2 The Electromagnetic Calorimeter	27
3.2.3 The Hadronic Calorimeter	32
3.2.4 The Solenoid	34
3.2.5 Muon Detectors	35
3.3 Triggers	37
3.3.1 Level 1 Trigger	38
3.3.2 The High Level Trigger	40
3.3.3 CMS Luminosity Monitoring	40
4 Event Reconstruction and Identification	42
4.1 Track Reconstruction	42
4.2 Primary Vertex Reconstruction	44
4.3 Photon Reconstruction	45
4.3.1 Photon Identification	47
4.3.2 Multivariate Photon Identification	49
4.4 Particle Flow Algorithm	52
4.4.1 Calorimeter Energy	53
4.4.2 Linking	53

4.5	Jet Reconstruction	54
4.5.1	Jet Algorithm	54
4.5.2	Jet Energy Corrections	56
4.5.3	Jet Identification	57
4.6	Electron Reconstruction	57
4.7	Muon Reconstruction	58
5	Analysis	60
5.1	Dataset	61
5.2	Triggers	62
5.2.1	Triggers Used on Single Photon Dataset	62
5.2.2	Triggers Used on Single Electron Dataset	67
5.3	Monte Carlo Samples	67
5.4	Event Selection	69
5.4.1	Photon Selection	69
5.4.2	Generator Level Photon Selection	70
5.4.3	Jet Selection	71
5.5	Purity Estimate	72
5.5.1	Background Template Construction	76
5.5.2	Fitting Procedure	83
5.5.3	Systematic Uncertainties in the Purity Measurement	86
5.6	Efficiency	95
5.6.1	Tag and Probe	110
5.6.2	Trigger Efficiency	111
5.6.3	Cluster Efficiency	115
5.6.4	Selection Efficiency	115
5.6.5	Conversion Safe Electron Veto Efficiency	116
5.6.6	MVA Efficiency	119
5.7	Unfolding	122
5.8	Results and Theoretical Comparison	127
6	Conclusion	142
Appendices		
A	MVA-Isolation Correlation Study	144
B	Bias Planes	159
C	Purity Fits	167
D	f_{in} Vs. f_{out}	196
	Bibliography	204
	Biographical Sketch	207

LIST OF TABLES

1.1	Bosons and their masses, the interactions they are responsible for, and the range of their interactions.	4
1.2	Fermions and their properties such as their masses, charges, and the interactions they experience [1].	5
4.1	Parameters related to the ECAL clustering and the associated threshold values. . . .	46
4.2	Effective area for ρ correction.	48
4.3	Linear corrections applied to MC for some variables before entering the MVA. The corrected value, $x_{corr}=a \times x + b$	52
5.1	Datasets used in the photon+jet triple differential cross section measurement.	62
5.2	HLT trigger paths used in the photon+jet triple differential cross section measurement along with the L1 triggers, the effective integrated luminosity, the offline p_T^γ range and the prescale values corresponding to the paths.	63
5.3	MC samples used in the photon+jet measurement for signal template construction, background sideband optimization, and the study of systematics.	68
5.4	Summary of the preselection cuts.	70
5.5	Summary of the jet selection cuts.	72
5.6	Number of candidate events along with statistical uncertainty for different p_T^γ and $ \eta^{jet} $ bins at $ \eta^\gamma < 0.8$	73
5.7	Number of candidate events along with statistical uncertainty for different p_T^γ and $ \eta^{jet} $ bins at $0.8 < \eta^\gamma < 1.44$	74
5.8	Number of candidate events along with statistical uncertainty for different p_T^γ and $ \eta^{jet} $ bins at $1.56 < \eta^\gamma < 2.1$	74
5.9	Number of candidate events along with statistical uncertainty for different p_T^γ and $ \eta^{jet} $ bins at $2.1 < \eta^\gamma < 2.5$	75
5.10	Photon isolation cuts applied to data for background templates in each bin of p_T^γ and $ \eta^\gamma $	83
5.11	Uncorrected values of purity and the associated fitting uncertainty for different p_T^γ and $ \eta^{jet} $ bins at $ \eta^\gamma < 0.8$	84

5.12	Uncorrected values of purity and the associated fitting uncertainty for different p_T^γ and $ \eta^{jet} $ bins at $0.8 < \eta^\gamma < 1.44$	85
5.13	Uncorrected values of purity and the associated fitting uncertainty for different p_T^γ and $ \eta^{jet} $ bins at $1.56 < \eta^\gamma < 2.1$	85
5.14	Uncorrected values of purity and the associated fitting uncertainty for different p_T^γ and $ \eta^{jet} $ bins at $2.1 < \eta^\gamma < 2.5$	86
5.15	The output signal fraction is corrected as $f_{out}^c = p0 + p1 \times f_{out} + p2 \times f_{out}^2$	87
5.16	The output signal fraction is corrected as $f_{out}^c = p0 + p1 \times f_{out} + p2 \times f_{out}^2$	87
5.17	Bias corrected values of purity for different p_T^γ bins in the barrel region. JetBin1, JetBin2, JetBin3 and JetBin4 correspond to $0.8 < \eta^{jet} < 1.5$, $1.5 < \eta^{jet} < 2.1$, and $2.1 < \eta^{jet} < 2.5$ respectively.	88
5.18	Bias corrected values of purity for different p_T^γ bins in the endcaps. JetBin1, JetBin2, JetBin3 and JetBin4 correspond to $0.8 < \eta^{jet} < 1.5$, $1.5 < \eta^{jet} < 2.1$, and $2.1 < \eta^{jet} < 2.5$ respectively.	88
5.19	Total uncertainties on the purity along with the breakdown of the uncertainties into different categories for $ \eta^\gamma < 0.8$, $ \eta^{jet} < 0.8$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	96
5.20	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $ \eta^\gamma < 0.8$, $0.8 < \eta^{jet} < 1.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	97
5.21	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $ \eta^\gamma < 0.8$, $1.5 < \eta^{jet} < 2.1$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	98
5.22	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $ \eta^\gamma < 0.8$, $2.1 < \eta^{jet} < 2.5$. Here, ‘Stat.’ refers to statistical uncer-	

	tainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp. refers to uncertainty due to signal template, and ‘Bg Temp. refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	99
5.23	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $0.8 < \eta^\gamma < 1.5$, $ \eta^{jet} < 0.8$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp. refers to uncertainty due to signal template, and ‘Bg Temp. refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	100
5.24	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $0.8 < \eta^\gamma < 1.5$, $0.8 < \eta^{jet} < 1.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp. refers to uncertainty due to signal template, and ‘Bg Temp. refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	101
5.25	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $0.8 < \eta^\gamma < 1.5$, $1.5 < \eta^{jet} < 2.1$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp. refers to uncertainty due to signal template, and ‘Bg Temp. refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	102
5.26	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $0.8 < \eta^\gamma < 1.5$, $2.1 < \eta^{jet} < 2.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp. refers to uncertainty due to signal template, and ‘Bg Temp. refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	103
5.27	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $1.5 < \eta^\gamma < 2.1$, $ \eta^{jet} < 0.8$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp. refers to uncertainty due to signal template, and ‘Bg Temp. refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	104

5.28	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $1.5 < \eta^\gamma < 2.1$, $0.8 < \eta^{jet} < 1.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	104
5.29	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $1.5 < \eta^\gamma < 2.1$, $1.5 < \eta^{jet} < 2.1$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	105
5.30	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $1.5 < \eta^\gamma < 2.1$, $2.1 < \eta^{jet} < 2.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	105
5.31	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $2.1 < \eta^\gamma < 2.5$, $ \eta^{jet} < 0.8$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	106
5.32	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $2.1 < \eta^\gamma < 2.5$, $0.8 < \eta^{jet} < 1.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	106
5.33	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $2.1 < \eta^\gamma < 2.5$, $1.5 < \eta^{jet} < 2.1$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to	

	the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	107
5.34	Total uncertainty on purity along with breakdown of the uncertainties into different categories for $2.1 < \eta^\gamma < 2.5$, $2.1 < \eta^{jet} < 2.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.	107
5.35	Trigger efficiency values for different p_T^γ bins in the barrel ($ \eta^\gamma < 1.44$) ($1.56 < \eta^\gamma < 2.5$) region.	114
5.36	Selection efficiencies for $ \eta^\gamma < 0.8$	117
5.37	Selection Efficiency for $0.8 < \eta^\gamma < 1.5$	118
5.38	Selection efficiency for $1.5 < \eta^\gamma < 2.1$	118
5.39	Selection efficiency for $2.1 < \eta^\gamma < 2.5$	118
5.40	MVA cut efficiency for $ \eta^\gamma < 0.8$	120
5.41	MVA cut efficiency for $0.8 < \eta^\gamma < 1.5$	121
5.42	MVA cut efficiency for $1.5 < \eta^\gamma < 2.1$	121
5.43	MVA cut efficiency for $2.1 < \eta^\gamma < 2.5$	121
5.44	Triple differential cross section values with the statistical and systematic (lower and higher) uncertainties for various p_T^γ regions and $ \eta^{jet} $ bins for $ \eta^\gamma < 0.8$	138
5.45	Triple differential cross section values with the statistical and systematic (lower and higher) uncertainties for various p_T^γ regions and $ \eta^{jet} $ bins for $0.8 < \eta^\gamma < 1.44$	139
5.46	Triple differential cross section values with the statistical and systematic (lower and higher) uncertainties for various p_T^γ regions and $ \eta^{jet} $ bins for $1.56 < \eta^\gamma < 2.1$	140
5.47	Triple differential cross section values with the statistical and systematic (lower and higher) uncertainties for various p_T^γ regions for four bins of $ \eta^{jet} $. These values are for $2.1 < \eta^\gamma < 2.5$	141

LIST OF FIGURES

2.1	PDFs ($f(x, Q^2)$) scaled by the momentum transfer (x) for the sea quarks, valence quarks, and gluons from MSTW 2008 at NNLO. The thickness of the bands shows the 1σ uncertainty [2].	13
2.2	Schematic representation of the role of non-perturbative QCD in a theory-data comparison.	14
2.3	Processes contributing to the photon+jet cross section.	16
2.4	Improvement in the NNPDF2.1 NLO for gluons without (green) and with (dashed blue) the inclusion of 7 TeV isolated direct photon data. The left plot shows the ratio of the old PDF to the new PDF while the right figure shows the reduction in the gluon PDF uncertainty as a function of the gluon momentum fraction x at a typical LHC scale of $Q=100$ GeV [3].	17
3.1	The CERN accelerator complex [4].	20
3.2	Schematics of the LHC. Here the two beams going counter-clockwise and clockwise are shown in blue and red respectively. The interaction points where the detectors are placed are shown by stars. More details are provided in [4].	22
3.3	Diagram of CMS detector showing various sub-detectors and their components [5]. . .	25
3.4	Layout of the CMS pixel detector. The end disks are shown in color, while the black surface represents the outermost layer of the cylindrical layer [6].	27
3.5	Schematics showing a cross section of the CMS tracker. Each line represents a detector module [6].	28
3.6	Schematic showing the arrangement of crystal in modules, supermodules and super-crystals in the endcaps of the ECAL. The preshower detector placed before the EE is also labeled. Details of the ECAL subcomponents are provided in [7].	29
3.7	Feynman diagram for bremsstrahlung process (left) and pair production mechanism (right).	29
3.8	Lead tungstate crystals with photodetectors attached. The left panel shows a barrel crystal with APD, while the right panel shows an endcap crystal with VPT[6].	30
3.9	Longitudinal view of the CMS HCAL along with the various components (HB, HE, HF, and HO). Pseudorapidity values are overlayed on the diagram to show the coverage of each component [7].	33

3.10	Schematic of one of the quadrants of CMS showing the four DT stations in the barrel (MB1-MB4, green), four CSC stations in the endcap (ME1-ME4, blue), and the RPC stations (red) [8].	36
3.11	The L1 trigger schematics [9].	38
4.1	Schematic of a decision tree [10].	50
4.2	Diagram showing jet clustering using the anti- k_t algorithm. The different colors correspond to areas corresponding to different jets [11].	55
5.1	HLT efficiency curves for paths HLT_Photon50_CaloIdVL_v*, HLT_Photon75_CaloIdVL_v*, HLT_Photon90_CaloIdVL_v*, HLT_Photon135_v*, and HLT_Photon150_v* (top to bottom and left to right) in barrel as a function of the offline p_T^γ	65
5.2	HLT efficiency curves for paths HLT_Photon50_CaloIdVL_v*, HLT_Photon75_CaloIdVL_v*, HLT_Photon90_CaloIdVL_v*, HLT_Photon135_v*, and HLT_Photon150_v* (top to bottom and left to right) in endcaps as a function of the offline p_T^γ	66
5.3	An example of the MVA response distribution for truth matched signal and background samples from MC simulation. Both signal and background areas are normalized to unity.	76
5.4	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD fakes in the barrel region using MC simulation.	78
5.5	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD fakes in the endcaps region using MC simulation.	79
5.6	Sideband variation plots. The z axis shows bias values obtained for MC toy experiments done with tentative background templates with the sideband window corresponding to the photon isolation values represented by x and y axes. Black dots represent the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin are used to construct the background template.	81
5.7	The purity fits before the bias correction is applied. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of the purity when fitting the candidate data with Gaussian variations in signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	89
5.8	The purity fits before the bias correction is applied. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of the purity when fitting the candidate data with Gaussian variations in signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	90

5.9	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins in $ \eta^\gamma $	91
5.10	Fits of the $Z\gamma$ data (red circles) with the sum of Voigtian function and Gaussian (red line). The blue lines correspond to the Voigtian function describing the signal and green lines correspond to the Gaussian related to the background.	93
5.11	The MVA response distributions for QCD background scaled to background fraction from Figure 5.10 in the mass region of $80 < M_{\gamma\mu\mu} < 100$ GeV is shown in cyan, $Z\gamma$ MC FSR distribution scaled to the signal fraction in the same three-body mass region is shown in pink, and the $Z\gamma$ data distribution is represented by the black squares with error bars.	94
5.12	Bias corrected purity values as a function of p_T^γ for various $ \eta^\gamma $, and $ \eta^{jet} $ combinations, where the photon is in the barrel.	108
5.13	Bias corrected purity values as a function of p_T^γ for various $ \eta^\gamma $, and $ \eta^{jet} $ combinations, where the photon is in the endcaps.	109
5.14	Trigger efficiency as a function of reconstructed photon p_T for the HLT_Photon135_v* path in the barrel (left) and endcap (right) regions. The relative efficiencies are found using the ratio/bootstrapping method and the absolute efficiencies are estimated using the Tag and Probe method.	112
5.15	Trigger efficiency as a function of reconstructed photon p_T for the HLT_Photon135_v* path in the barrel (left) and endcap (right) region. The relative efficiencies are multiplied by a scale factor estimated by taking the ratio between the two methods (Tag and Probe to estimate absolute efficiencies and ratio/bootstrapping to calculate relative efficiency). The shaded area shows the uncertainty on the relative efficiency scaled by the SF, whereas the blue error bar shows the statistical uncertainty associated with the Tag and Probe method.	113
5.16	Trigger efficiency as a function of reconstructed photon p_T for barrel (left) and endcap (right) regions. The values correspond to the relative efficiency measurements for HLT_Photon50_CaloIdVL_v*, HLT_Photon75_CaloIdV_v*, HLT_Photon90_CaloIdVL_v*, HLT_Photon135_v*, and HLT_Photon150_v* with and without the application of the Scale factor at the p_T^γ ranges where these paths are used.	113
5.17	Clustering efficiency as a function of generated photon p_T in four different generated $ \eta^\gamma $ regions.	115
5.18	Selection efficiencies as a function of generated p_T^γ in four different generated $ \eta^\gamma $ regions using four different methods.	117
5.19	MVA cut efficiencies as a function of generated p_T^γ in four different generated $ \eta^\gamma $ regions using four different methods.	120

5.20	Response matrices obtained for different choices of $ \eta^\gamma $. $\gamma + \text{jet}$ MC events passing all the event selection criteria are used for obtaining the matrices. The migration from over-flow and under-flow bins (not shown) is also considered.	124
5.21	Ratio of the $\gamma + \text{jet}$ MC testing sample unfolded using the response matrices from Figure 5.20 to the truth distribution. The closure test is done using three different algorithms provided in the RooUnfold package.	125
5.22	Original data distribution, the unfolded distribution, and the distribution unfolded using the dataweighted response matrix for uncertainty measurement for various bins with p_T^γ in the barrel region. The unfolded distributions do not differ much from the original distributions indicating a very small effect due to unfolding.	126
5.23	Original data distribution, the unfolded distribution, and the distribution unfolded using the dataweighted response matrix for uncertainty measurement for various bins with p_T^γ in the barrel region. The unfolded distributions do not differ much from the original distributions indicating a very small effect due to unfolding.	128
5.24	Measured triple differential cross section values along with the total uncertainty for different bins in p_T^γ , $ \eta^\gamma $, and $ \eta^{jet} $. Notice that the distributions are multiplied by a factor of 100, 10,000 and 1,000,000 for $0.8 < \eta^{jet} < 1.5$, $1.5 < \eta^{jet} < 2.1$, and $2.1 < \eta^{jet} < 2.5$, respectively. The systematic uncertainties are represented by colored bands around the central value, whereas the statistical uncertainties are represented by error bars that are not visible due to the extremely large vertical axis range.	129
5.25	Breakdown of the systematic uncertainties on the cross section for different bins in p_T^γ , $ \eta^\gamma $, and $ \eta^{jet} $. The different colors/markers correspond to contributions from the purity measurement, efficiency measurement, luminosity measurement, unfolding, and the total uncertainty.	130
5.26	Breakdown of the systematic uncertainties on the cross section for different bins in p_T^γ , $ \eta^\gamma $, and $ \eta^{jet} $. The different colors/markers correspond to contributions from the purity measurement, efficiency measurement, luminosity measurement, unfolding, and the total uncertainty.	131
5.27	Theory prediction of $\gamma + \text{jet}$ triple differential cross section using GamJet package at NLO.	134
5.28	Theory prediction of $\gamma + \text{jet}$ triple differential cross section using GamJet package at NLO.	135
5.29	Ratios of measured triple differential cross section to the theory prediction using GamJet at NLO.	136
5.30	Ratios of measured triple differential cross section to the theory prediction using GamJet at NLO.	137

A.1	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	144
A.2	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	145
A.3	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	146
A.4	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	147
A.5	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	148
A.6	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	149
A.7	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	150
A.8	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	151
A.9	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	152
A.10	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	153
A.11	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	154
A.12	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	155
A.13	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	156
A.14	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	157
A.15	Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.	158
B.1	Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the	

	bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.	159
B.2	Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.	160
B.3	Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.	161
B.4	Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.	162
B.5	Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.	163
B.6	Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.	164
B.7	Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.	165
B.8	Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.	166
C.1	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show	

	the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	167
C.2	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	168
C.3	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	169
C.4	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	170
C.5	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	171
C.6	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	172
C.7	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	173

C.8	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	174
C.9	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	175
C.10	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	176
C.11	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	177
C.12	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	178
C.13	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	179
C.14	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting	

	the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	180
C.15	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	181
C.16	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	182
C.17	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	183
C.18	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	184
C.19	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	185
C.20	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	186
C.21	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show	

	the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	187
C.22	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	188
C.23	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	189
C.24	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	190
C.25	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	191
C.26	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	192
C.27	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	193

C.28	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	194
C.29	The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.	195
D.1	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.	196
D.2	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.	197
D.3	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.	198
D.4	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.	199
D.5	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.	200
D.6	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.	201
D.7	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.	202
D.8	MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.	203

ABSTRACT

The production of direct photons in association with jets can be used to probe gluon parton distribution functions as well as to test perturbative Quantum Chromodynamics. This thesis presents the measurement of the triple differential cross section for photon plus jet events as a function of photon transverse momentum (p_T^γ), photon pseudorapidity (η^γ), and jet pseudorapidity (η^{jet}). The measurement is made using 19.7 fb^{-1} of data collected by the Compact Muon Solenoid detector in proton-proton collisions at the center-of-mass energy of 8 TeV. The kinematic regions probed are $40 < p_T^\gamma < 1000 \text{ GeV}$, $|\eta^\gamma| < 2.5$ and $|\eta^{\text{jet}}| < 2.5$. These results are compared to next-to-leading order theoretical predictions.

CHAPTER 1

INTRODUCTION

“The Cosmos is all that is or was or ever will be. Our feeblest contemplations of the Cosmos stir us – there is a tingling in the spine, a catch in the voice, a faint sensation, as if a distant memory, of falling from a height. We know we are approaching the greatest of mysteries.” - Carl Sagan

Since very early in the course of human civilization, curiosity has led some of us to derive extreme pleasure in understanding the workings of nature. It is perhaps this quest for knowledge that has been the single most important factor in propelling the advancement of physics. With the progress in classical mechanics and thermodynamics, it has been said that in early 1900s physicists felt an uncharacteristic complacency. Lord Kelvin has even been quoted to say that there was nothing new left in physics to be discovered. In hindsight, he could not have been more wrong since many of the greatest achievements of physics were soon to follow including the theories of general and special relativity in the early 1900s and the development of an entirely new sub-field called particle physics.

1.1 Short History of Particle Physics

Though particle physics is a development of late 19th and early 20th century, it builds on the tradition of atomism that dates as far back as to the ancient Greeks. Ancient thinkers like Democritus and Leucippus first came up with the concepts of atoms in the 5th century B.C. However, their beliefs of atoms as indivisible units is now known to be wrong. Modern particle physics began with the discovery of the electron, one of the building blocks of atoms, by J.J Thompson. Soon after, Einstein’s photoelectric theory paved the way for the discovery of the second elementary particle: photons, which were recognized only as waves until then. These discoveries were followed

by Rutherford's gold foil experiment, which is unarguably the predecessor for modern high energy experiments that attempt to understand the structure of the subatomic world by bombarding matter with highly energetic projectiles and studying the properties of the outgoing particles. The picture of atomic nuclei was more or less completed with Chadwick's discovery of the neutron in 1932. The same year the very first antiparticle, the positron, was discovered by Anderson in a study of cosmic radiation using cloud chambers. Other particles such as pions, muons and kaons were detected soon after. It was decades after their discovery that these particles were known to be composite particles made of fundamental particles known as quarks and gluons. These particles are categorized as mesons [12].

Parallel to the discoveries of new particles, Paul Dirac developed his 'theory of radiation' that laid the foundation for Quantum Electrodynamics (QED), which combines electromagnetism with quantum theory. QED describes the interaction of electrically charged particles (such as electrons) through emission and absorption of photons. This theory was later refined by physicists such as Feynman, Schwinger, Tomonaga, and Dyson. Two of the major contributions made by them include unification of quantum mechanics with special relativity and the solution to the infinite loop problem present in the methods of calculation based on perturbation theory. A method called renormalization was implemented, which combined the infinite loops together into corrections for physical quantities such as the mass and charge of the electron. The results obtained using this procedure were profoundly successful, encouraging theoretical physicists to extend the gauge invariant formulation of QED to describe other interactions of nature. Here, gauge invariance refers to the property of observables to remain constant even when the underlying field is changed through a process commonly known as gauge transformation [12].

In 1954, Yang and Mills attempted to generalize the gauge theory of QED to strong and weak interactions. The carrier particle of the Yang-Mills theory is required to be massless in order to satisfy gauge invariance, but it needed to be massive for weak interactions because of the short range nature of the interactions. Weinberg, Glashow and Salam later combined electromagnetism with the weak interaction by incorporating the Higgs mechanism into their model, which solved the problem of the massless bosons. Here, the Higgs mechanism refers to a way to generate mass for the

gauge bosons without breaking the gauge theory, and was proposed by Anderson, Brout, Englert, Higgs, Guralnik, Hagen and Kibble. The four massless force carriers before the addition of Higgs mechanism to the model become three massive force carriers and with one remaining massless in the modified model. Around the same time, the earlier experimental discovery of a zoo of particles called hadrons (later known to be composite particles made of fundamental particles held together by the strong force) had led scientists like Gell-Mann and Zweig to propose an underlying flavor ‘symmetry’ generated by quarks - fundamental constituents of hadrons. Here, symmetry refers to a property of a system that remains unchanged under some transformation. According to this model, the resonances (peak around certain energy associated to subatomic particles and their excitations) would be organized as singlets, doublets or multiplets having common spin and parity (intrinsic properties of particles) for each member of the group. This quark model came to be incorporated into the framework of Quantum Chromodynamics (QCD) in 1973. This theory predicted eight massless force carriers called gluons.

By the 1970s, theoretical advancement in particle physics had superseded experimental discoveries. One of the first breakthroughs in experimental particle physics after this period came through the discovery of a new particle called the ‘ J/ψ ’ simultaneously at Stanford Linear Accelerator Center (SLAC) and Brookhaven National Laboratory. J/ψ is a meson consisting of a charm quark and a charm antiquark, and its existence was theoretically predicted based on the quark model. The J/ψ discovery confirmed the presence of the charm quark. One after another, discoveries of various fundamental particles occurred such as the tau lepton at SLAC, the W and Z bosons at CERN, top quark at the Tevatron. More information on the fundamental particles and their properties will be provided in remainder of this chapter. In this section, it is sufficient to note that all these discoveries along with the latest discovery of the Higgs boson by the CMS and ATLAS collaborations provided evidence for the theoretical model that has now come to be known as the Standard Model.

1.2 The Standard Model (SM)

In simple words, the Standard Model is a theory that describes the fundamental particles and their interactions in terms of the exchange of virtual particles. It is a well-tested and consistent

Table 1.1: Bosons and their masses, the interactions they are responsible for, and the range of their interactions.

Boson	Mass[GeV/ c^2]	Interaction	Interaction Range[m]
Photon(γ)	0	EM	Infinite
$W^\pm$	80.4	Weak	10^{-18}
Z	91.2	Weak	10^{-18}
gluon(g)	0	Strong	10^{-15}

theory that has undergone only minor adjustments such as in the inclusion of neutrino mass that was not described by the original theory. In a gauge description, the SM is a quantum field theory with an underlying $SU(3)_C \otimes SU(2)_W \otimes U(1)_Y$ gauge symmetry. Gauge symmetry demands that the Lagrangian of a system is invariant under the continuous groups of local transformations. In this description, C refers to the color charge, W stands for weak isospin, and Y refers to the weak hypercharge. These concepts will be detailed in the rest of this chapter.

1.2.1 Fundamental Particles

The fundamental particles in the SM fall into two categories: bosons and fermions. Bosons are the force carriers with integer spins, whereas fermions have half-integer spins. Table 1.1 lists the fundamental Standard Model bosons and some of their properties. Referring to the table, one can see that the gauge bosons mediating the weak force, i.e. $W^\pm$ and Z , are massive while the ones mediating the electromagnetic and the strong forces, i.e. γ and g , are massless. Fermions are categorized into leptons and quarks. The electron (e^-), muon (μ^-), tau (τ^-) and their corresponding neutrinos, i.e. electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ), constitute the six leptons. Similarly, there are six different quarks: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). All of these elementary particles have corresponding antiparticles with identical mass and opposite electric charge.

Both leptons and quarks are also categorized into three generations and are listed in Table 1.2. The primary difference between particles in the different generation is their mass. The first generation particles are the lightest and the more stable than their counterparts from second and third generations. At temperatures relevant to human life, they make up all the stable and many unstable

Table 1.2: Fermions and their properties such as their masses, charges, and the interactions they experience [1].

Generation	Fermion	Mass[MeV/ c^2]	Charge[e]	Interaction
Leptons				
I	e	≈ 0.511	-1	EM, W
	ν_e	$< 2 \times 10^{-6}$	0	W
II	μ	≈ 105.66	-1	EM, W
	ν_μ	< 0.19	0	W
II	τ	≈ 1776.82	-1	EM, W
	ν_τ	< 18.2	0	W
Quarks				
I	u	≈ 2.3	$+2/3$	EM, W, S
	d	≈ 4.8	$-1/3$	
II	c	$\approx 1.275 \times 10^3$	$+2/3$	
	s	≈ 95	$-1/3$	
II	t	$\approx 173.21 \times 10^3$	$+2/3$	
	b	$\approx 4.18 \times 10^3$	$-1/3$	

matter we know to exist. Not surprisingly, the most familiar examples of hadrons such as protons and neutrons are made of up and down quarks, and the electron is also a first generation lepton.

1.2.2 Interactions

The SU(3) symmetry group in the SM originates from the strong sector. The strong force binds quarks together to form hadrons and at a larger scale protons and neutrons are bound with strong force to form atomic nuclei. The Pauli Exclusion Principle requires that the quarks present in a hadron possess different quantum numbers because no two fermions can occupy the same state. This quantum number for the strong force is known as color charge. There are three colors — red, blue, green. A baryon, which is a bound state of three different quarks, possesses three different color quarks. A meson, on the other hand, possesses a quark and an antiquark with corresponding anticolor. The hadrons themselves are color neutral. The strong force is experienced only by those particles that possess color charge and is mediated through the exchange of eight different gluons. Particles other than quarks and gluons do not possess color charge, hence, they are not able to interact through the strong force. However, since gluon possesses color charge, it is able to interact with other gluons.

The weak force is experienced by almost all fundamental particles. The weak interaction can either be propagated by the exchange of the electrically neutral Z boson or by the exchange of the charged W^+/W^- bosons. All fundamental particles with the exception of the gluon and photon interact through the exchange of Z bosons. The W boson interact even with photons. The weak interaction is the force responsible for the creation of elements and plays an important role in the powering of stars.

At the quantum scale, the weak force is unified with the electromagnetic force to make up what is known as electroweak force. The associated symmetry is known as electroweak symmetry. The subsequent breaking of the unified theory is known as Electroweak Symmetry Breaking (EWSB). The unified electroweak (EW) symmetry does not manifest in physical gauge bosons. The weak bosons and photon that we observe in nature are a result of the Higgs mechanism, which is responsible for mixing the four gauge bosons of the unified theory to give a different set of four bosons in the spontaneously broken theory. The Higgs mechanism also gives rise to a scalar boson, known as the Higgs boson. The SM Higgs boson interacts with anything that has mass. The latest fundamental particle to be discovered is a SM Higgs boson candidate having a mass of $125.09 \pm 0.21 \pm 0.11$ GeV [13]. The exact properties of this new boson are in the process of being determined.

With the exception of neutrinos, all leptons, quarks, and W bosons can interact via the exchange of a photon, thereby experiencing the electromagnetic (EM) force. It is responsible for holding atoms and molecules together. The most prominent technological advances that humankind has seen in recent times such as electronics and optics have been through the understanding of the EM interaction. The $U(1)$ symmetry group in the description of the SM describes this interaction. The quantum number associated with EM force is called electric charge, which is a non-trivial combination of the weak isospin (a property of the EW theory that defines a particle's transformation under the $SU(2)$ symmetry) and the hypercharge (property of the EW theory that defines a transformation under the $U(1)$ group). Since, the photon itself does not carry electric charge, it is not able to interact with itself. The difference in this property between gluons and photons gives rise to some of the unique properties of QCD. Since, the measurement that is covered in this

dissertation consists of a photon in the final state with QCD as the dominant interaction, these properties will be discussed in detail in the next chapter.

CHAPTER 2

THEORY

“When you ask what are electrons and protons I ought to answer that this question is not a profitable one to ask and does not really have a meaning. The important thing about electrons and protons is not what they are but how they behave, how they move. I can describe the situation by comparing it to the game of chess. In chess, we have various chessmen, kings, knights, pawns and so on. If you ask what chessman is, the answer would be that it is a piece of wood, or a piece of ivory, or perhaps just a sign written on paper, or anything whatever. It does not matter. Each chessman has a characteristic way of moving and this is all that matters about it. The whole game of chess follows from this way of moving the various chessmen.”- Paul A.M. Dirac

2.1 Quantum Chromodynamics

Quantum Electrodynamics (QED) forms the theoretical foundation for Quantum Chromodynamics (QCD) not just in a gauge description, but it is useful to understand the non-relativistic static potential associated with the strong interaction. In QED, the interaction between two charges is mediated by one or more photons and the associated non-relativistic potential is given by the Coloumb potential $V_{QED}(r)$ as

$$V_{QED}(r) = K_e \frac{q_1 q_2}{r}. \quad (2.1)$$

Here, q_1 and q_2 are the charges of electromagnetic particles, and K_e is the proportionality constant called Coulomb’s constant. The electromagnetic force between q_1 and q_2 is inversely proportional to the square of the distance r between them. In QCD, the interaction between quarks is mediated by the exchange of gluons, which also interact to other gluons. A Coloumb like potential, therefore, is not sufficient to describe the interaction between the quarks. The static potential V_{QCD} between the quarks is of the form

$$V_{QCD}(r) = -\frac{A(r)}{r} + Kr. \quad (2.2)$$

When $r \rightarrow 0$, the first term dominates and the strong force is described by a Coloumb-like interaction. The strength of the interaction depends on the distance r . This behavior gives rise to a special

characteristic of the strong interaction called asymptotic freedom. Asymptotic freedom means that the interaction between particles becomes asymptotically weaker as the interaction energy increases or the distance decreases, allowing predictions to be made from the model at small distances or high energies.

On the other hand, when $r \rightarrow \infty$, gluon self-interaction becomes important and the second term dominates [14]. The strong interaction at large distances is non-linear and is difficult to calculate mathematically. This term in the interaction is responsible for another important property of the strong interaction called confinement. At large distances, the force between the quarks does not decrease as they are pulled apart. The energy of the gluon field becomes larger than the energy required to produce a quark-antiquark pair. It becomes energetically favorable for gluons to produce a quark antiquark pair. The pair that is created from the gluons combine with the original quarks to produce color neutral hadrons. This property of the strong force is known as color confinement, since color states are confined inside hadrons and cannot exist freely. This is in contrast with the behavior seen in QED. In QED, when two electrons are pulled apart, the strength of the interaction is reduced, allowing the electrons to exist as free states.

2.1.1 Perturbative QCD (pQCD)

Often when a mathematical problem cannot be solved exactly, perturbation theory can be applied to approximate its solution. The basic idea is to define the problem in terms of a small parameter ϵ such that its solution at $\epsilon = 0$ is solvable. Perturbation theory gives a systematic expansion in terms that can be added to the ‘unperturbed’ parts to approximate the solution to the original problem. The solution converges only as long as the solutions to the perturbed part are small compared to the solution to the unperturbed part. One of the best examples of the application of this theory is that of QED. In QED, the coupling constant α is approximately $\frac{1}{137}$ at large distances. The coupling constant is a parameter that determines the strength of the interaction. Perturbation theory allows us to expand the solution to a QED physical process in terms of an expansion in powers of α . Because of the sufficiently small value of α , i.e. weakness of the interaction, QED converges rapidly. For example, consider an electron-positron annihilation process producing a muon-antimuon pair ($e^+e^- \rightarrow \mu^+\mu^-$). The total probability amplitude M_{total} ,

whose modulus squared gives the probability for such an interaction to occur, can be expressed as:

$$M_{total} = M_{LO} + M_{NLO} + M_{NNLO} + \dots, \quad (2.3)$$

where M_{LO} is the probability amplitude at leading order, M_{NLO} is the amplitude at next-to leading order and M_{NNLO} is the next-to-next to leading order term. Each of the contributions is proportional to a different power of the coupling constant. Hence, $M_{LO} \propto \alpha$, $M_{NLO} \propto \alpha^2$, and so on. The cross section is proportional to M_{total}^2 . Therefore, the total cross section σ_{total} is given by the sum of cross section terms at different orders;

$$\sigma_{total} = \sigma_{LO} + \sigma_{NLO} + \sigma_{NNLO} + \dots \quad (2.4)$$

The major contribution for this process is via a single intermediate virtual photon, whose contribution to the cross section is of the order α^2 . The electron-positron pair that forms the virtual photon of momentum transfer Q equal to the center-of-mass energy $\sqrt{s}$, can also fluctuate into a quark antiquark pair. If Q is large enough, the production rate for such short-distance process can be predicted by perturbation theory. Though the process by which the quarks and gluons form hadrons will modify the final state, it occurs at a much later time so that it does not modify the original probability of the event to occur. Therefore, the rules of QED can be applied to quarks in order to calculate the original probability for the electron-positron annihilation process to produce a quark-antiquark pair in the final state ($e^+e^- \rightarrow \gamma \rightarrow q\bar{q}$). This process can, in fact, be used to demonstrate the existence of quarks, though they only exist as bound states. At leading order, the total cross section is

$$\sigma_{tot}(e^+e^- \rightarrow q\bar{q}) = \sum_q \sigma(e^+e^- \rightarrow q\bar{q}) = 3 \sum_q e_q^2 \sigma(e^+e^- \rightarrow \mu^+\mu^-), \quad (2.5)$$

where e_q is the charge of quark q . The factor 3 comes from three color degrees of freedom. Since the cross section $\sigma_{tot}(e^+e^- \rightarrow q\bar{q})$ has been well studied, it can be used to test the prediction for the existence of quarks and their colors, and predictions have confirmed the existence of quarks in three colors.

Any calculation beyond tree level involves loop integrals that appear when integrating over internal momenta in the Feynman diagrams (mathematical representations of interaction between particles). When these integrals become infinite they are called divergences. Divergences that occur due

to contributions from physical phenomena at very short distances or high energies are known as ultraviolet divergences. They are rendered finite through a process called regularization and subtracted through a process of renormalization. This systematic approach introduces the dependence of the QCD coupling constant on the renormalization scale. Therefore, the coupling constant is not actually constant but a running coupling that evolves as a function of the renormalization scale:

$$\alpha_s(\mu) = \frac{12\pi}{(33 - 2f) \ln(\frac{\mu^2}{\Lambda^2})}. \quad (2.6)$$

Here, f is the number of quark flavors, μ is the renormalization scale, and Λ is a dimensionless QCD scale parameter that is approximately 200 MeV or 1 fm [15]. The renormalization scale μ is often chosen to be the characteristic momentum transfer Q for the interaction. In the electron-positron annihilation process to produce, $Q = \sqrt{s}$, where $\sqrt{s}$ is the collision energy. By the uncertainty principle this interaction will happen on a distance scale of the order $1/Q$. Furthermore, based on the above equation it is easy to see that Λ is an interaction scale at which the QCD interaction becomes strong. At high energies or short distance scales the QCD coupling decreases. Equivalently, it increases at low energies or short distance scales. This behavior of the coupling is related to the QCD properties of asymptotic freedom and confinement.

2.1.2 Non-perturbative QCD

Though solutions to QCD processes at the parton (the constituents of a hadron) level can be found via perturbative methods, it is not sufficient to fully understand how the strong force behaves. As such, it is equally important to learn about the composition and formation of hadrons. These concepts can be explained through the use of valence quarks and sea quarks. Valence quarks in a hadron are the quarks that contribute to the quantum numbers of the hadron. Apart from the valence quarks, hadrons consist of many virtual quark-antiquark pairs formed from gluons which are called sea quarks. Similarly, the quark-antiquark pairs can annihilate to produce gluons. For example, though a proton is said to consist of two up quarks and one down quark (the valence quarks), there are also a large number of gluons and quark antiquark pairs of different flavors (the sea quarks) present inside the proton. In this way, a hadron can be thought of as a soup of gluons and sea quarks in which valence quarks are embedded.

In high-energy collisions involving hadrons in the initial state, it is necessary to know the distribution of the constituents as it is the quarks and gluons that interact to produce the final state particles. Similarly, if the final state particles are quarks and gluons, they cannot exist freely. After the collision, the quarks will immediately radiate gluons. The gluons produce quark-antiquark pairs, which then combine to produce hadrons. This process is known as hadronization and is one of the long-distance or low-energy features of QCD. Long distance phenomena like these are not predictable by pQCD since they have ‘infrared singularities’ in perturbation theory.

Infrared singularities may come from either soft divergences or collinear divergences. Soft divergences occur when the radiated massless particle’s energy is close to zero, and collinear divergences occur when the angle between the radiated particle and the final state parton approaches zero. These divergences can be absorbed into non-perturbative functions known as Parton Distribution Functions (PDFs) and Fragmentation Functions (FFs). PDF quantify the probability density to find parton of momentum fraction x at a squared energy scale Q^2 . Similarly, FFs parameterize the formation of hadrons from partons of a momentum fraction x at a squared energy scale Q^2 . Calculation of the parton momentum distributions inside of hadrons requires knowledge of hadron wave functions, which perturbation theory cannot predict [16]. Therefore, these quantities must be estimated empirically. Large collections of data collected from diverse measurements undergo a procedure known as global PDF analysis to determine these non-perturbative quantities.

The process for the determination of a PDF involves selection of a starting Q_0 at which perturbation theory is valid. Based on a particular scheme and an order of perturbation theory, quark and gluon distribution are parameterized at Q_0 as a function of x . Sets of QCD evolution equations called DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) [17] are solved to obtain the PDFs at any value of $Q > Q_0$. Though the PDFs are non-perturbative, the DGLAP equations themselves can be solved using perturbation theory. Data available from many experiments involving hadrons in the initial and final states are chosen and the evolved equations are used to calculate χ^2 between the theory and data. The parameterizations of the input distributions are adjusted to minimize the χ^2 . Finally, an analytical function to describe the PDF as a function of x and Q is obtained

MSTW 2008 NNLO PDFs (68% C.L.)

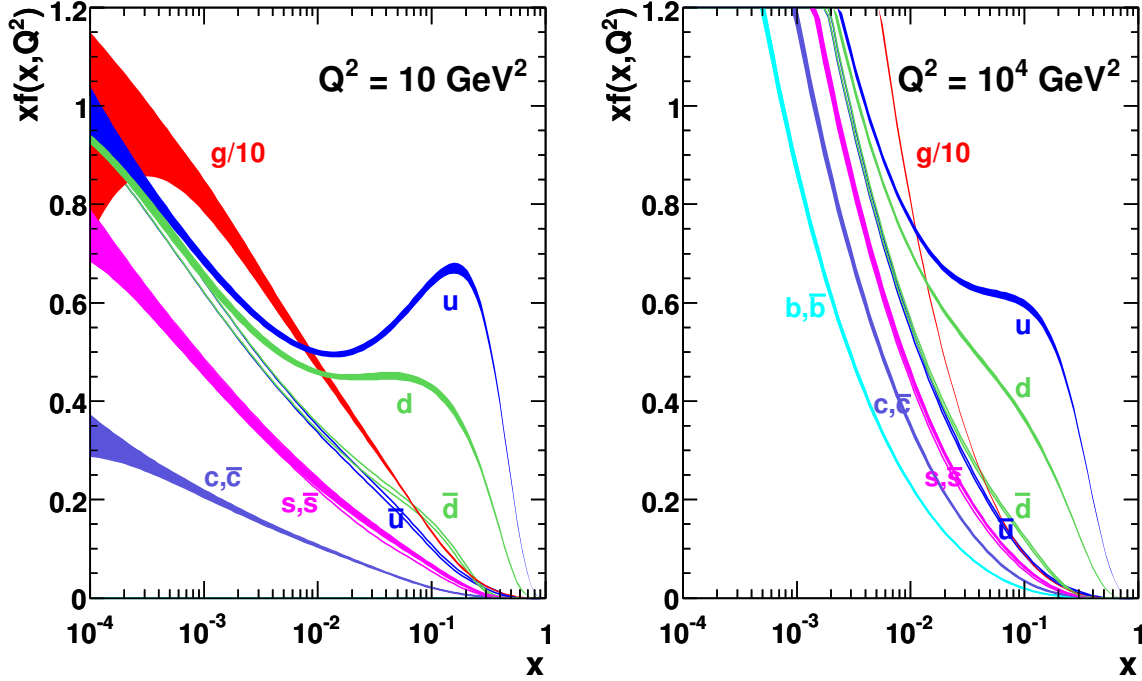


Figure 2.1: PDFs ($f(x, Q^2)$) scaled by the momentum transfer (x) for the sea quarks, valence quarks, and gluons from MSTW 2008 at NNLO. The thickness of the bands shows the 1σ uncertainty [2].

[15]. An example of PDFs for the valence and sea quarks as well as the gluons determined from the MSTW 2008 PDF set at NNLO can be seen in Figure 2.1. It can be seen that since there needs to be equal number of sea quarks as antiquarks, the PDFs for charm, strange, and bottom quarks and antiquarks are the same inside of a proton. Also, the PDFs have large uncertainties at some combinations of x and Q^2 . These PDF uncertainties can constitute a large percentage of the total uncertainty for various searches at the Large Hadron Collider (LHC). The only way to constrain the PDFs is by the inclusion of more data capable of probing a wider range of the Q^2-x kinematic plane.

Just as experiments are required to estimate and constrain the PDFs and FFs, comparisons of most measurements from the LHC with a theory prediction requires knowledge of these quantities. A schematic of a typical LHC process comprised of the various non-perturbative phenomena and a

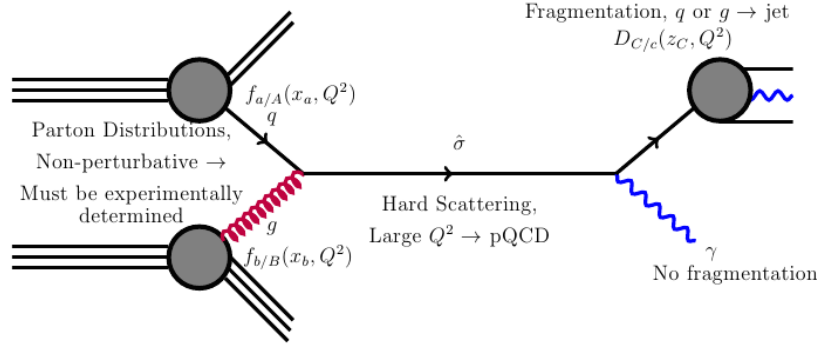


Figure 2.2: Schematic representation of the role of non-perturbative QCD in a theory-data comparison.

perturbatively calculable hard scattering cross section is presented in Figure 2.2. This shows that the partonic cross section has to be convoluted with the non-perturbative functions to compare the cross section theory prediction with a measurement at hadron colliders. In this way, the role of these non-perturbative functions in a hadron collider experiment is cyclical.

2.2 Photon+jet physics

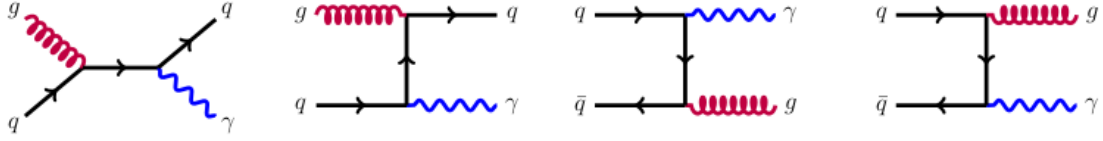
Photons that are produced from the hard scattering of partons are known as direct photons. Direct photons do not include photons produced from the secondary electromagnetic decay of hadrons. At leading order (LO) in pQCD, photon+jet events are produced mainly through quark-gluon Compton scattering ($qg \rightarrow q\gamma$) and quark antiquark annihilation ($q\bar{q} \rightarrow g\gamma$), where the final state photon is a direct photon. At higher orders, significant contributions to the process with photon+jet in the final state come from fragmentation photons, where the final state partons fragment to produce photons that are part of hadronic showers. The background to photon + jet events mostly comes from meson decay to two photons ($\pi \rightarrow \gamma\gamma$, $\eta \rightarrow \gamma\gamma$), where the photons are surrounded by energy contribution from the hadron constituents. These events are mostly suppressed by isolation requirements that are often used in photon + jet measurements at hadron colliders. Isolation cuts simply require that the total energy in a region around a photon candidate be lower than some cutoff value. The cross section corresponding to such background is many orders of

magnitude larger than the photon+jet signal; therefore, isolation cuts are often used to minimize such backgrounds despite the disadvantage of removing a large amount of fragmentation photons. Some of the processes contributing to the photon+jet cross section along with the ones described in this paragraph are shown in Figure 2.3.

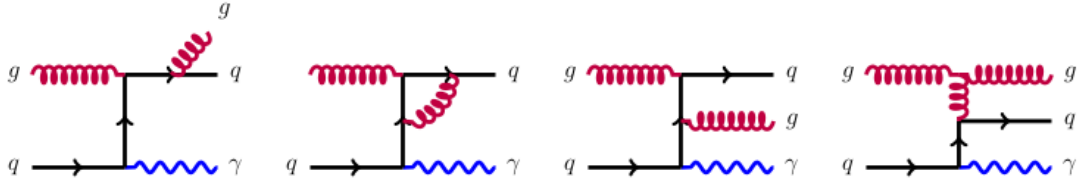
At the LHC, there is an abundance of low momentum fraction (x) gluons, which makes the Compton scattering mechanism the most likely. Because of the gluon in the initial state of this process, events with a photon and jet(s) in the final state can provide constraints on the gluon PDFs and also provide precision tests of pQCD. Historically, direct photon production in hadron collisions have been used until a decade and half ago to constrain the gluon PDF in global PDF fits as well as in determining the strong coupling constant. Measurements carried out by the E706 collaboration at center-of-mass energies of 20–40 GeV showed disagreement in the photon cross section compared to the pQCD prediction at NLO. This theory-experiment disagreement and the availability of jet measurements led theorists to remove photon measurements from the global fits. However, with the inclusion of a large amount of LHC data at the highest center-of-mass energy ever reached in hadron colliders, gluon PDF measurements have been shown to benefit with the inclusion of direct photon data [3]. Figure 2.4 shows an example of constraining the gluon PDF for NNPDF2.1 at NLO using isolated photon data from the LHC at a center-of-mass energy of 7 TeV. There is also the added benefit of an experimentally ‘cleaner’ signal when using final state photons instead of jets to study QCD at a hadron collider. The experimental details on what makes the photon a good probe of QCD will be covered in the later chapters.

Because of the availability of a very large dataset (19.7 fb^{-1}) at a wide range of photon and jet kinematics, the cross section measurement of photon+jet events at 8 TeV is capable of probing QCD at a wide range of Q^2 and x . With the kinematic cuts that are provided in Chapter 5, this analysis can probe $Q^2 = (E_T^\gamma)^2 = [1.6 \times 10^3 - 10^6] \text{ GeV}^2$. Furthermore, the quantity $x_T = 2E_T^\gamma/\sqrt{s}$, which is an approximation of the momentum fraction ($x = E_T^\gamma(e^{\pm\eta^{jet}} + e^{\pm\eta^\gamma})/\sqrt{s}$) when both photon and jet are centrally produced, can be probed at a range of [0.01–0.25]. The lowest order in perturbation theory makes several predictions about correlations between the direct photon and the associated jet. The correlation at different combinations of photon-jet kinematics can be exploited

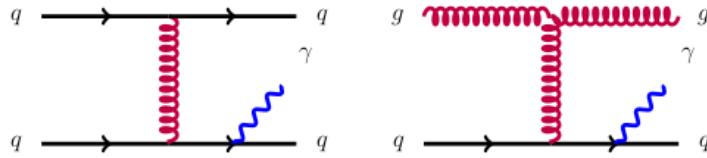
1 LO



2 NLO



3 Bremsstrahlung



4 Fragmentation

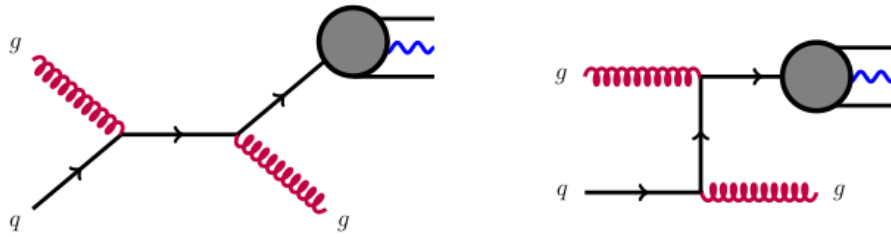


Figure 2.3: Processes contributing to the photon+jet cross section.

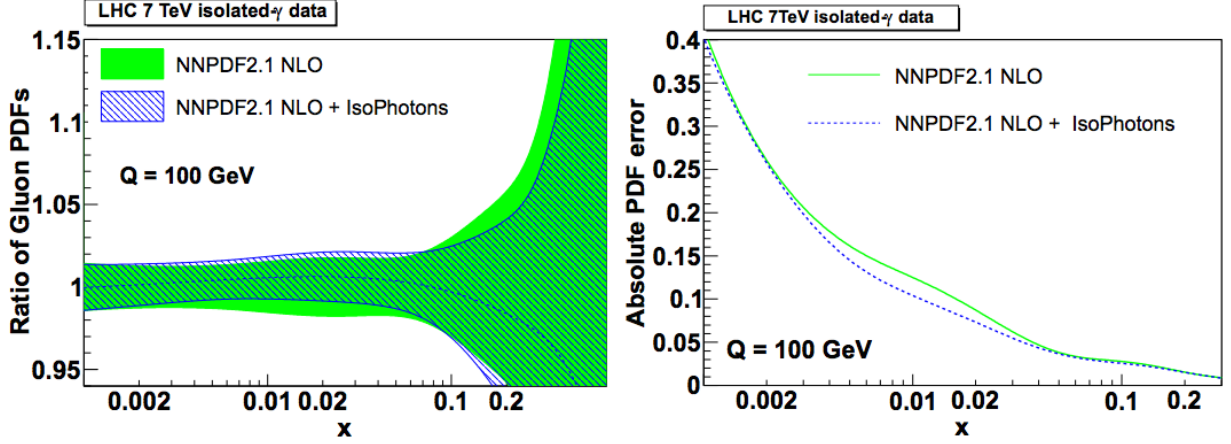


Figure 2.4: Improvement in the NNPDF2.1 NLO for gluons without (green) and with (dashed blue) the inclusion of 7 TeV isolated direct photon data. The left plot shows the ratio of the old PDF to the new PDF while the right figure shows the reduction in the gluon PDF uncertainty as a function of the gluon momentum fraction x at a typical LHC scale of $Q=100$ GeV [3].

to provide tighter constraints on the underlying parton-level kinematics at very small values of x , where a large uncertainty on the gluon PDF currently exists. Therefore, this analysis can help in the understanding of QCD at ranges not possible with past experiments.

Though the primary motivation for this measurement is to provide data that can be used to understand QCD interactions, photon+jet events can also be useful for understanding other physics processes. They are a major source of background for many standard model and beyond the standard model processes including $H \rightarrow \gamma\gamma$ and extra dimensions. Furthermore, they are also used for jet energy corrections [18].

2.3 Observable

The triple differential cross section that is measured in this analysis for photon+jet process ($AB \rightarrow \gamma + X$) can be theoretically expressed in terms of the convolution of non-perturbative functions and the parton level hard-scattering cross section,

$$\frac{d^3\sigma}{dp^3}(AB \rightarrow \gamma + X) = \frac{1}{E^\gamma} \sum_{abcd} \int dx_a dx_b dz_c f_{a/A}(x_a) f_{b/B}(x_b) D_{\gamma/c}(z_c) \frac{\hat{s}}{z_c^2 \pi} \frac{d\sigma}{d\hat{t}}(ab \rightarrow cd) \times \delta(\hat{s} + \hat{t} + \hat{u}) \quad (2.7)$$

Here, p is the four-momentum of the photon+jet system; E^γ is the energy of the photon; A, B, X are hadrons; and γ is a photon. a, b, c and d are partons. $D_{\gamma/c}(z_c)$ is the FF and characterizes the probability for the final state parton c to fragment into a final state particle γ with momentum fraction z_c evaluated at energy scale Q^2 . $\frac{d\hat{\sigma}}{d\hat{t}}(ab \rightarrow cd)$ is the differential cross section for partons a and b to produce partons c and d . Similarly, $f_{a/A}(x_a)$ and $f_{b/B}(x_b)$ are the PDFs and characterize the probability density for finding a parton $a(b)$ within hadron $A(B)$ with momentum fraction $x_a(x_b)$ evaluated at energy scale Q^2 . $\hat{s}$, $\hat{t}$, and $\hat{u}$, called Mandelstam variables, are Lorentz-invariant quantities of the partonic process. $\hat{s}$ is the square of the center-of-mass energy of the partons while, $\hat{t}$ and $\hat{u}$ are the squares of the four momentum transfers from partons a and b to parton c [15].

CHAPTER 3

EXPERIMENT

“The overwhelming presence of machines and instrumentation must be one of the most salient features of the modern scientific laboratory...The development of science depends at least as much on new machinery as it does on new ideas.” - Ronald Giere

The experimental setup for this measurement is located at the European Organization for Nuclear Research (CERN) on the border of France and Switzerland. There are two major components that come together to make this measurement possible. A series of accelerators including the Large Hadron Collider (LHC) are responsible for creating the proton beams that collide to produce the photon+jet events that we are measuring in this analysis. Another equally important piece of equipment is the Compact Muon Solenoid (CMS) detector that identifies and measures the properties of the particles produced in LHC collisions.

3.1 Accelerators

The CERN accelerator complex consists of a series of accelerators that work together to produce and accelerate the proton beams to the desired energy. The CERN complex consisting of Linacs, Booster, Proton Synchrotron (PS), Super Proton Synchrotron (SPS), and the LHC is shown in Figure 3.1. At one end of Linac 2 (one of the linear accelerators in the CERN complex), protons are produced by passing hydrogen gas through an electric field and stripping electrons off of the atoms. Linac 2 uses radiofrequency (RF) cavities to accelerate the protons to an energy of 50 MeV. The protons are then passed onto the second accelerator in the sequence, which is the PSB. It is composed of four superimposed synchrotron rings. The PSB accelerates the protons to 1.4 GeV and injects them into the PS. The protons are accelerated by the PS to 25 GeV and injected to the SPS. The beams are bent around the 7 kilometer circumference ring by 744 dipole magnets and accelerated to 450 GeV in the SPS before being injected to the two rings of the LHC [4].

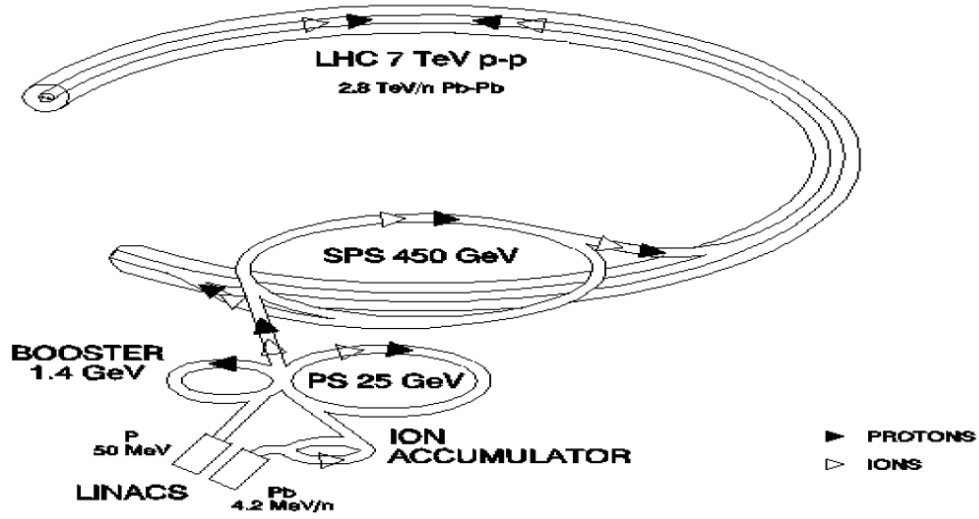


Figure 3.1: The CERN accelerator complex [4].

3.1.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring, superconducting hadron accelerator and collider. It is the world's largest particle collider, housed in a tunnel that was previously used by Large Electron-Positron Collider (LEP) from 1989 to 2000. LEP was used to deliver data that enabled the precise measurement of the W and Z boson masses. It is roughly 27 kilometers in circumference and 45 to 170 meters underground. At the LHC, collisions take place between two highly energetic proton beams traveling in opposite directions. The center of mass energy during the first collisions in 2010 was 7 TeV. It was raised to 8 TeV from 2012 to 2013 before the LHC was shut down for a planned upgrade. In mid 2015, LHC resumed with a center of mass energy of 13 TeV. The phase after the planned upgrade is known as Run II while the data taking period before the upgrade is referred to as Run I. The measurement presented in this thesis is done with the 8 TeV data collected during Run I.

The proton beams collide at four points where the four experiments are located: ALICE, ATLAS, CMS, and LHCb as can be seen in Figure 3.2. Each of the beams has its own beam pipe with magnets and vacuum chambers. The two beam pipes have common sections only where collisions

occur. There are 1232 dipole magnets that bend the beams, and 392 quadrupole magnets that focus the beams. These superconducting magnets that need to be kept cool by circulating 96 tons of helium. Each beam is composed of distinct bundles of protons called “bunches”. A proton beam at full intensity consists of 2808 bunches with a nominal bunch spacing of 25 ns. During the 8 TeV run, the bunch spacing was 50 ns – equivalent to a collision frequency of 20 MHz. The beams are kept in focus with 392 quadrupole magnets to increase the chances of interaction. They are still a few centimeters long and $16\ \mu\text{m}$ wide during collision. Each bunch is composed of approximately 1.5×10^{11} protons and each bunch crossing produces about 25 proton-proton interactions. Out of all the proton-proton interactions observed per bunch crossing, only a few are hard scattering (interactions that result in large momentum transfer). The additional proton-proton interactions other than the hard scatters are called ‘pile-up’, and are distinct from ‘underlying event’ that describe additional interactions originating from the same proton-proton collision.

3.1.2 The LHC Luminosity

One of the most important parameters in a particle physics experiment is the amount of energy that is available for producing events of interest. Another important parameter is the number of interactions that is associated with the interaction cross section, where interaction cross section refers to an effective area that quantifies the probability of the particular interaction taking place when particles collide. The instantaneous luminosity, $\mathcal{L}(t)$, of an experiment gives the number of particles per unit time (dN/dt) through unit cross section (σ):

$$\mathcal{L}(t) = \frac{1}{\sigma} \frac{dN}{dt}. \quad (3.1)$$

A quantity that is often used to characterize the amount of data delivered by a collider is an integrated luminosity. Integrated luminosity ($\mathcal{L}$) of a collider simply measures the total luminosity delivered in a given amount of time and can be obtained by integrating the instantaneous luminosity:

$$\mathcal{L} = \int \mathcal{L}(t) dt, \quad (3.2)$$

and

$$\mathcal{L} = \frac{N}{\sigma}. \quad (3.3)$$

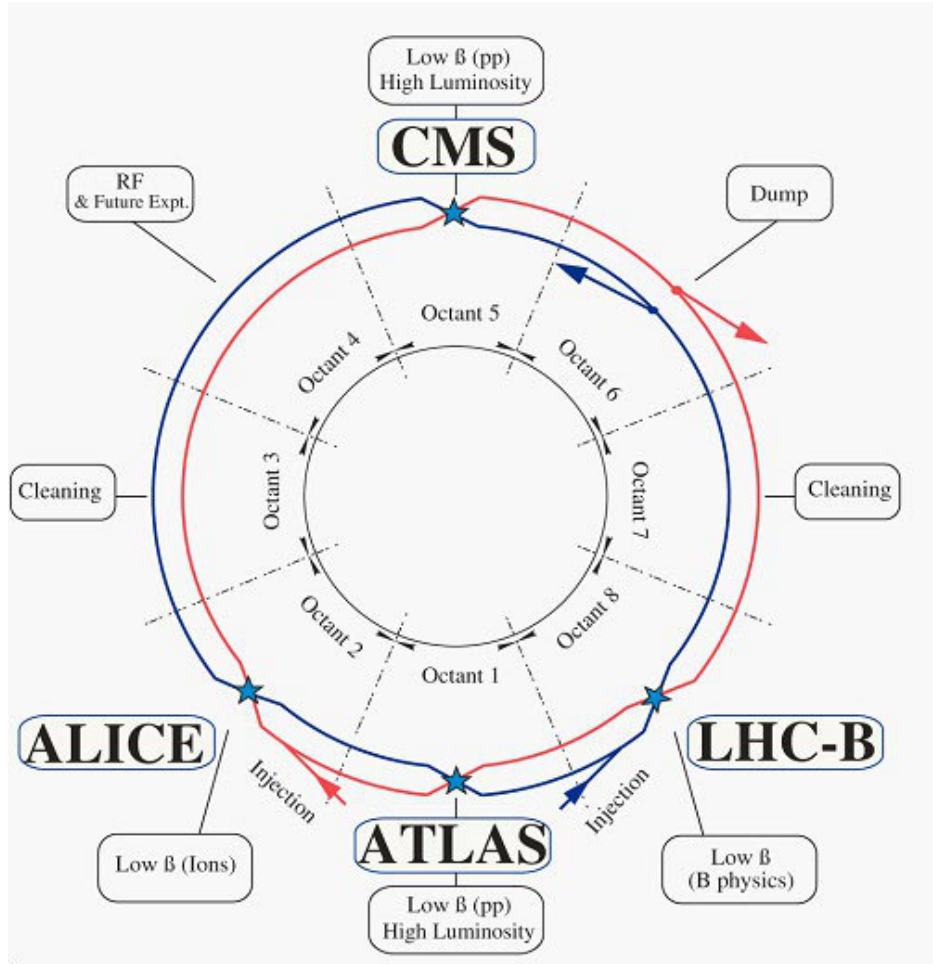


Figure 3.2: Schematics of the LHC. Here the two beams going counter-clockwise and clockwise are shown in blue and red respectively. The interaction points where the detectors are placed are shown by stars. More details are provided in [4].

For a production of a particular process, the cross section is constant at a fixed value of center-of-mass energy. Therefore, the number of events can be increased by increasing the integrated luminosity, which can be achieved by either increasing the instantaneous luminosity or by collecting data over a longer period of time. This is particularly beneficial for rare events with lower cross sections. The collected data is measured in inverse units of area. It is a common practice to use units of barn to measure the cross section and integrated luminosity in the units of inverse barn. One barn is equal to 10^{-28} m^2 , which is approximately the cross sectional area of a uranium nucleus.

Luminosity at the LHC can also be expressed in terms of the beam parameters:

$$\mathcal{L} = \frac{fn^2}{4\pi\sigma_x\sigma_y}. \quad (3.4)$$

Here, n is the number of protons in each crossing bunch, f is frequency of bunch crossing, σ_x and σ_y are the width of the bunches in the x and y direction, respectively. One can see from Equation 3.4 that the luminosity can be improved by increasing the number of particles in the bunches or frequency of bunch crossing or by better focusing the beams at the collision point. The choice of frequency is affected by the time resolution of the detector. If the bunch spacing is smaller than the time resolution, then detector suffers from pile-up contributions from successive bunch crossings referred to as out-of-time pile-up. On the other hand, increasing the number of bunches or number of protons per bunch results in increased number of pile-up events within the same bunch crossing.

3.2 The Compact Muon Solenoid Detector

The Compact Muon Solenoid (CMS) detector is one of the two all purpose detectors at the CERN complex. It is roughly cylindrical in shape with 28.7 m in length and 15 m in diameter. It is situated in a cavern 100 m underground and weighs 14,000 tons. The detector is designed to provide high granularity and hermetic coverage in units of pseudorapidity.

CMS uses a right-handed coordinate system, with the origin at the nominal interaction point. The x -axis points towards the center of the LHC, the y -axis points up, and the z -axis points in the counter-clockwise beam direction. A modified spherical coordinate system with r , η , and ϕ

are used. Here, the radial coordinate $r = \sqrt{x^2 + y^2}$ is the distance from the center of the beam pipe, $\phi = \tan^{-1}(\frac{y}{x})$ is the azimuthal angle measured in the x/y plane and η is a quantity called pseudorapidity. η is related to the polar angle θ as:

$$\eta = -\ln \tan(\theta/2). \quad (3.5)$$

For relativistic particles, pseudorapidity is a approximation for rapidity. This quantity is preferred at the colliders because difference in rapidity is Lorentz invariant under boosts along the longitudinal axis. It is defined in terms of the energy E and the longitudinal component of momentum p_z of the particle,

$$y = \frac{1}{2} \log\left(\frac{E + p_z}{E - p_z}\right). \quad (3.6)$$

Figure 3.3 shows the layout of CMS, which is very similar to most general purpose detectors. The innermost detectors surrounding the beam line are the tracker. The electromagnetic calorimeter (ECAL) surrounds the tracker. The Hadronic calorimeter (HCAL) constitutes the third detector and is immediately outside of the ECAL. A 3.8 Tesla superconducting solenoidal magnet encloses the calorimeters and the tracker. The outermost and largest detector is the muon system.

3.2.1 The Tracker

The silicon tracker is the innermost detector component at CMS. The purpose of the tracker is to measure tracks of charged particles and determine the position of primary and secondary vertices. This is an important step towards the full event reconstruction. When charged particles travel through the tracker they ionize the material. The ionized electrons within an applied electric field produces a weak current which is amplified to produce “hits”. Since the tracker lies within the CMS solenoid with a magnetic field in the z direction, charged particles follow a curved path in the transverse plane inside the tracker. Mapping a particle’s position at various points in the detector allows for measurement of the particle’s transverse momentum. Based on the direction of the curve, the particle’s charge can be measured as well.

CMS DETECTOR

Total weight : 14,000 tonnes
 Overall diameter : 15.0 m
 Overall length : 28.7 m
 Magnetic field : 3.8 T

STEEL RETURN YOKE
 12,500 tonnes

SILICON TRACKERS
 Pixel ($100 \times 150 \mu\text{m}$) $\sim 16\text{m}^2 \sim 66\text{M}$ channels
 Microstrips ($80 \times 180 \mu\text{m}$) $\sim 200\text{m}^2 \sim 9.6\text{M}$ channels

SUPERCONDUCTING SOLENOID
 Niobium titanium coil carrying $\sim 18,000\text{A}$

MUON CHAMBERS
 Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
 Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers

PRESHOWER
 Silicon strips $\sim 16\text{m}^2 \sim 137,000$ channels

FORWARD CALORIMETER
 Steel + Quartz fibres $\sim 2,000$ Channels

CRYSTAL
 ELECTROMAGNETIC
 CALORIMETER (ECAL)
 $\sim 76,000$ scintillating PbWO_4 crystals

HADRON CALORIMETER (HCAL)
 Brass + Plastic scintillator $\sim 7,000$ channels

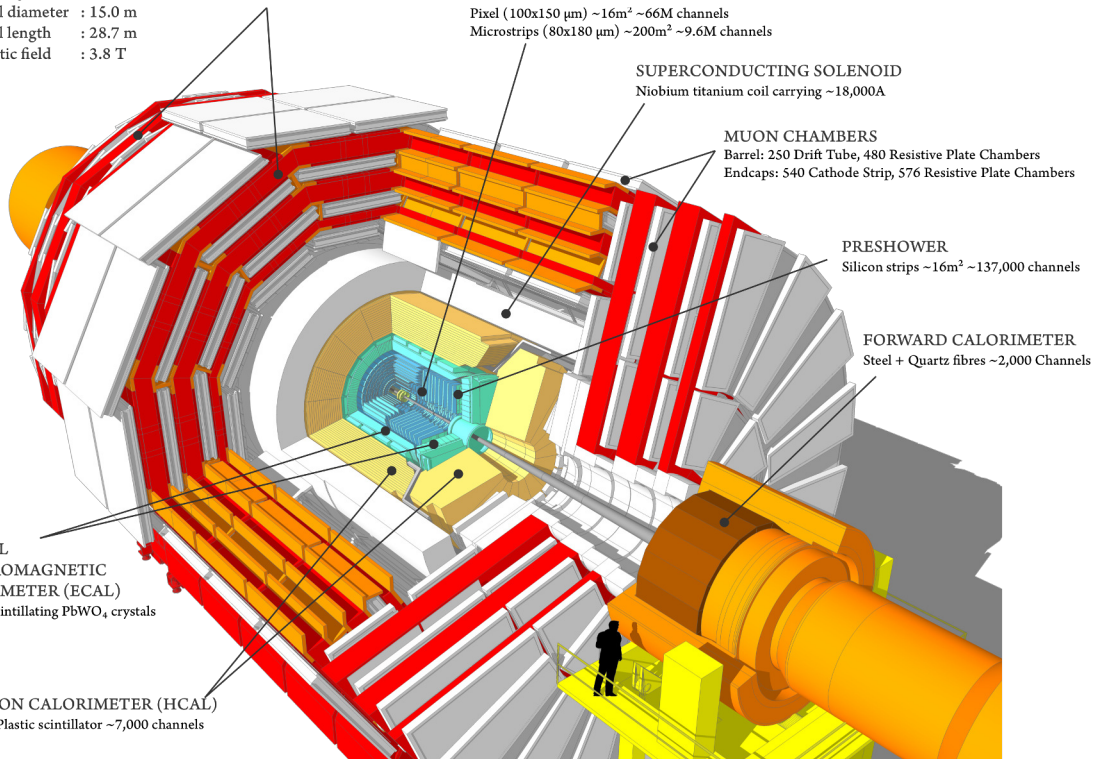


Figure 3.3: Diagram of CMS detector showing various sub-detectors and their components [5].

The silicon pixel and silicon strip tracker (SST) together constitute 200 m² of active silicon area making CMS tracker the largest silicon tracker ever built [7]. It provides coverage in the pseudo-rapidity range of $|\eta| < 2.5$. Some of the major constraints that were taken into consideration in the design of the CMS tracker were survival in a high radiation environment, low mass to reduce the loss of particle energy before entering the calorimeters, high mechanical stability, size of the system for easy assembly, and cost effectiveness [19]. Every bunch crossing at the designed LHC luminosity creates approximately 1000 particles. Since this leads to high hit rate per surface area closest to the interaction point, in order to keep the occupancy at or below 1%, pixel detectors are used closest to the interaction point.

At the core of the CMS tracker are three 53 cm long cylindrical layers of pixel detectors situated at radii of 4.4, 7.3, and 10.2 cm. In the endcaps, there are two layers of end disks on each side. The disks are 6 to 15 cm in radius placed at $z = \pm 34.5$ and ± 46.5 cm as shown in Figure 3.4. These three layers in the barrel and two layers in the endcaps contain 66 million pixels mounted on cooling tubes and each with a dimension 100 $\mu\text{m} \times 150 \mu\text{m}$ (approximately the width of two hair strands). When charged particles pass through the detectors, they lose energy that ionizes the silicon atoms. The electrons ejected from the atoms are collected and the signal is amplified by a silicon chip attached to each pixel. Knowing the position of the pixel that was touched helps in the 3-dimensional reconstruction of the particle's trajectory. The high granularity of the pixel detector allows for better charged track pattern recognition. The pixel detector has a position resolution of $\approx 15 \mu\text{m}$, which is particularly beneficial for reconstructing secondary vertices from b and τ decays [20].

The SST, which is based on micro-strip silicon devices forms the first four barrel layers and radially surround the pixel detector. They are referred to as Tracker Inner Barrel (TIB) and have a typical cell size of 10 cm $\times$ 80 μm . In the endcaps, there are three inner disks known as Tracker Inner Disks (TID) at each end with the strips radially arranged on the disks. The TIB/TID provide up to 4 $r-\phi$ measurements along a particle's trajectory with the 320 μm thick silicon sensors. The SST layers may be either single or double sided (back-to-back sensors). Because of the reduced particle flux, the cell size in the outer six layers can be bigger; hence, they have dimensions of 25 cm $\times$ 180 μm in

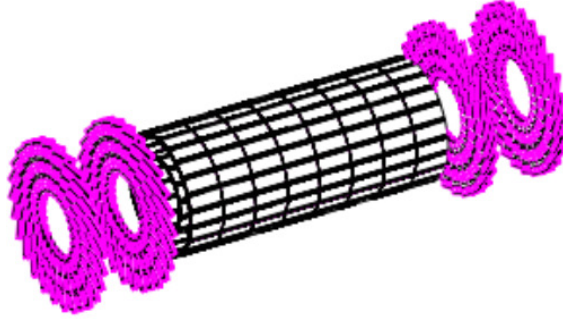


Figure 3.4: Layout of the CMS pixel detector. The end disks are shown in color, while the black surface represents the outermost layer of the cylindrical layer [6].

the barrel. The outer barrel layers are known as Tracker Outer Barrel (TOB) and at a thickness of $500\text{ }\mu\text{m}$ are thicker than the TIB. The TECs (Tracker EndCaps) cover $124\text{ cm} < |z| < 282\text{ cm}$ and $22.5\text{ cm} < |z| < 113.5\text{ cm}$ and are composed of nine disks. The thickness for the first four rings is $320\text{ }\mu\text{m}$ and increases to $500\text{ }\mu\text{m}$ for the remaining five rings. This results in total of 15,148 silicon microstrip detector modules with 10 million silicon strips read out by 80,000 micro-electronic chips. As in the case with the pixel detector, the SST also produces a signal through ionization of the silicon which moves through the external electric field to generate a small amount of current. A cross sectional layout of the CMS tracker with its various subcomponents is provided in Figure 3.5.

3.2.2 The Electromagnetic Calorimeter

CMS Electromagnetic Calorimeter (ECAL) is a hermetic homogeneous calorimeter made of lead tungstate (PbWO_4) crystals. It is situated between the tracker and the Hadronic Calorimeter. Based on the pseudorapidity values, it is divided into a barrel section (EB) and two endcaps (EE). EB covers the pseudorapidity range $|\eta| < 1.479$ and consists of 61,200 crystals organized into 36 supermodules each with about 17,000 crystals. The EB crystals are tapered in shape, and their front faces are approximately 0.0174×0.0174 in $\eta - \phi$ coordinates, i.e. $22 \times 22\text{ mm}^2$ in the front and $26 \times 26\text{ mm}^2$ in the back. EEs are made up of about 15,000 crystals each and cover $1.556 < |\eta| < 3.0$. The EE crystals' granularity varies in $\eta - \phi$ coordinates from 0.0175×0.0175 to 0.05×0.05 , i.e. $28.62 \times 28.62\text{ mm}^2$ in the front and $30 \times 30\text{ mm}^2$ in the back. In the forward and backward pseudorapidity region, there are additional ECAL preshower detectors (ESs) providing

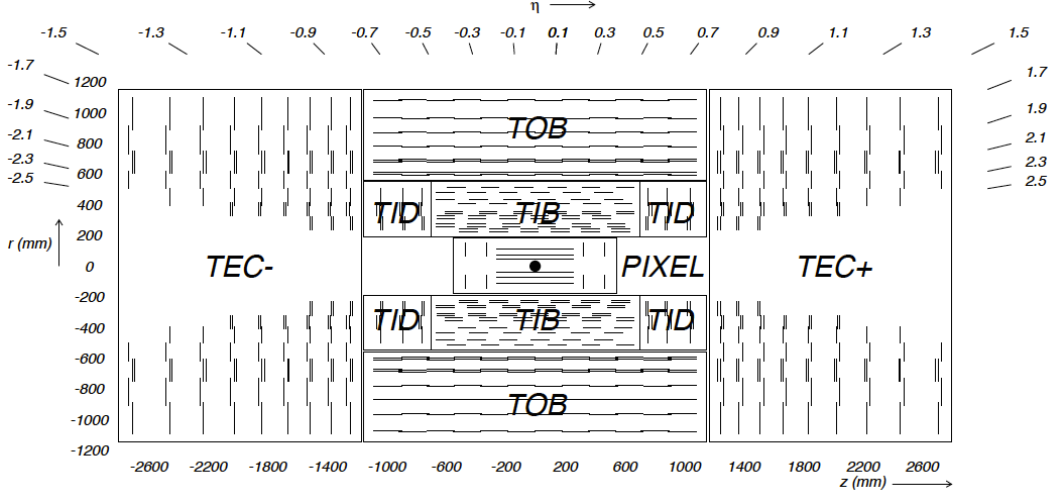


Figure 3.5: Schematics showing a cross section of the CMS tracker. Each line represents a detector module [6].

extra spatial resolution to the EEs. They cover a fiducial range $1.654 < |\eta| < 2.6$ and consist of two layers of silicon strips with an orthogonal strip arrangement and a lead absorber. The subcomponents of the ECAL detector along with the arrangement of crystals is shown in Figure 3.6.

The dominant interaction mechanism by which electrons and positrons interact in the ECAL crystals is through bremsstrahlung. Bremsstrahlung is a process by which charged high energy particles radiate photons when in the presence of the Coulomb field of an atomic nuclei. A Feynman diagram for the bremsstrahlung process is shown in the left hand side of Figure 3.7. The energy loss $-\frac{dE}{dx}$ of an electron through bremsstrahlung can be expressed as

$$-\frac{dE}{dx} = \frac{1}{X_0} E, \quad (3.7)$$

where X_0 is the radiation length and refers to the property of matter that quantifies the mean distance traveled by a high-energy electron before losing all but $1/e$ of its energy through interaction. PbWO_4 crystals have a radiation length of 0.89 cm, which enables electrons and photons to deposit most of their energy in the ECAL. X_0 is directly related to the fourth power of the particle's charge and inversely related to the square of the mass of the interacting particle. Therefore, the analogous quantity for a muon is suppressed by the square of its mass compared that of electron. It

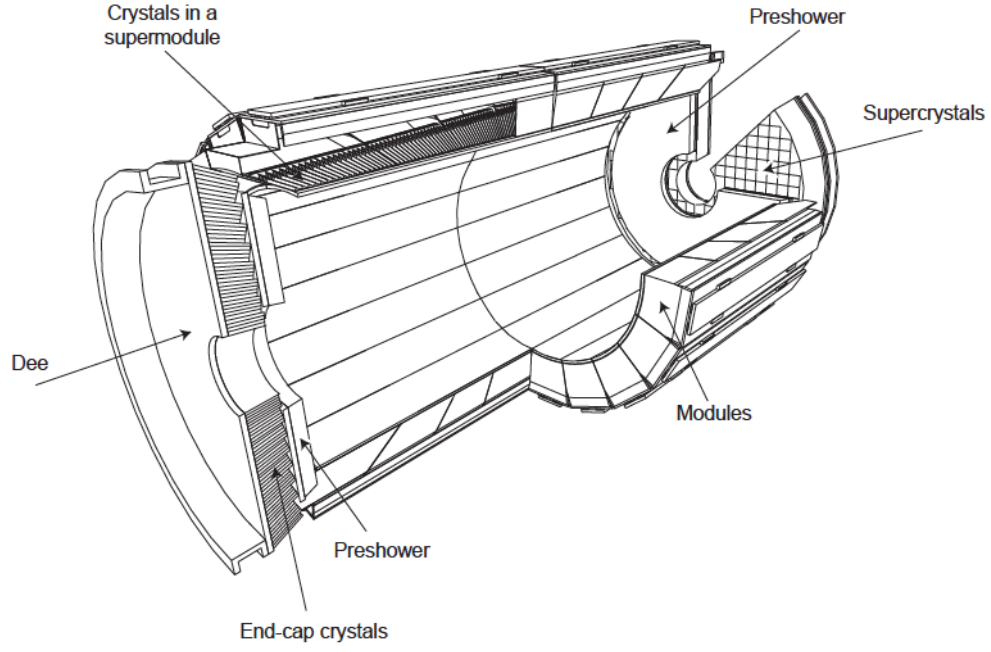


Figure 3.6: Schematic showing the arrangement of crystal in modules, supermodules and supercrystals in the endcaps of the ECAL. The preshower detector placed before the EE is also labeled. Details of the ECAL subcomponents are provided in [7].

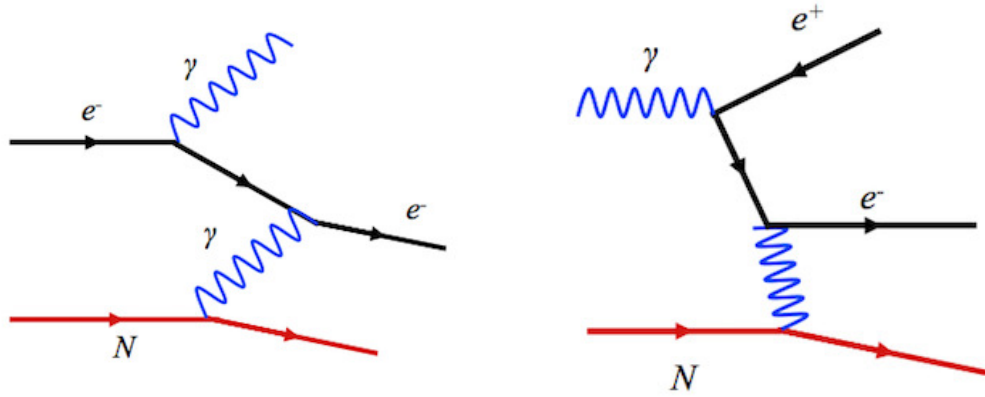


Figure 3.7: Feynman diagram for bremsstrahlung process (left) and pair production mechanism (right).

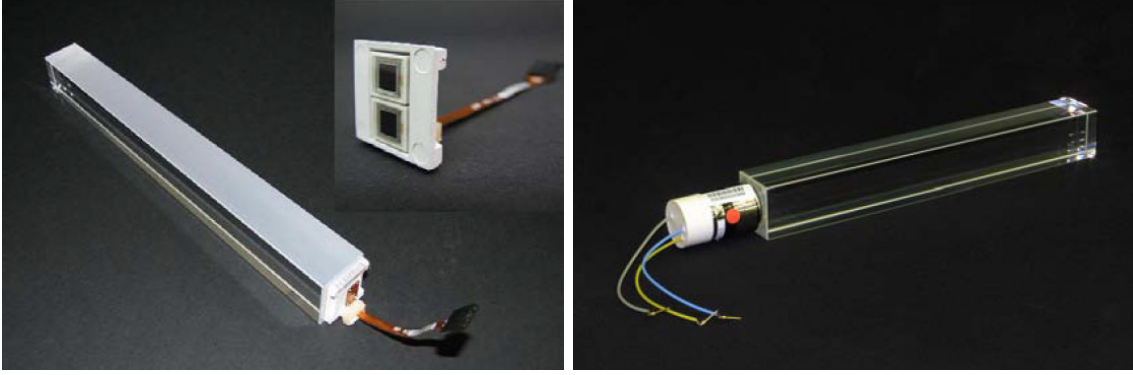


Figure 3.8: Lead tungstate crystals with photodetectors attached. The left panel shows a barrel crystal with APD, while the right panel shows an endcap crystal with VPT[6].

is because of this property that muons are not generally stopped within the calorimeter and require the outer muon system. Since photons have no electric charge, they need to transfer their energies to charged particles in order to be detected. At the CMS, photons interact with a heavy nucleus by a mechanism known as pair production, shown in the right panel of Figure 3.7, to produce an electron-positron pair. This gives an additional factor of 9/7 to the radiation length of photons as compared to the electrons.

The main purpose of any calorimeter is to measure the energy of an incoming single particle by converting its energy into a cascade of lower energy particles. At a depth of one radiation length, an electron radiates one photon, and hence there are two particles originating from one electron. After X radiation lengths, the electron will produce 2^X particles each with approximately $(1/2)^X$ of the original energy. When this energy becomes equal to the critical energy E_c , the shower stops. The critical energy is the energy of the particles in the shower below which the particles do not interact with the ECAL. The longitudinal shower depth X_c scales logarithmically with the particle's incident energy E_0 :

$$X_c = \frac{\ln(E_0/E_c)}{\ln 2}. \quad (3.8)$$

The transverse dimension of the fully contained electromagnetic shower is characterized by a constant called the Moliere radius. It is the radius of a cylinder containing approximately 90% of an electron shower's energy deposition. PbWO_4 has a Moliere radius of 2.2 cm for electrons and

photons, which results in well-collimated electromagnetic showers. It also has a high density of 8.3 g/cm³ and the oxygen in the crystal makes it highly transparent as can be seen in Figure 3.8. The high density, short radiation length and small Moliere radius make lead tungstate a good material choice for the ECAL.

Lead tungstate crystals emit photons whose energy is proportional to the interacting particle's energy via scintillation light. When the ECAL crystals absorb energy, electrons are excited to higher energy levels and subsequently radiate photons while returning to lower energy states. Almost 80% of the scintillation light is emitted within the 25 ns LHC bunch spacing. This property enables incoming particles to interact with the material producing short and rapid photon bursts, without losing performance over time. The scintillation light coming is collected with radiation-hard photodetectors mounted on each crystal. Avalanche Photo Diodes (APDs) are used with the barrel crystals; whereas Vacuum Phototriodes (VPTs) are used with the endcaps crystals shown in Figure 3.8. The signals are amplified and digitized by 40-MHz analog-to-digital converters (ADCs) before being pushed to the trigger system. The light yield of the crystals is dependent on the temperature, and therefore the crystals are kept at 16°C–18°C, with a temperature stability of 0.1°C.

Since CMS was designed to search for the decay of the Higgs boson via the two photon channel, the ECAL was designed to have good energy resolution to improve the diphoton invariant mass resolution. The energy resolution of a generic calorimeter can be expressed as

$$\frac{\sigma_E}{E} = \frac{a}{E} \oplus \frac{b}{\sqrt{E}} \oplus c. \quad (3.9)$$

Here, a is the noise term, b is the stochastic term, and c is a constant term. The noise term includes contributions from electronic noise, digitization noise, pile-up etc. The stochastic term with contributions coming from lateral shower containment and, fluctuations in the energy deposition, whereas the constant term is related to the non-uniformity of the longitudinal light collection, intercalibration uncertainty, leakage of showers, etc. The first two terms reduce with the energy; therefore the constant term is the most important contribution to the ECAL resolution for electromagnetic objects. The noise, stochastic and constant terms obtained for electrons in a 3×3 barrel crystal matrix during a 2004 test beam are 0.12%, 2.8% and 0.3% respectively [7].

Constant exposure to radiation changes the transparency of the ECAL crystals. The damage causes the constant term to increase, reducing the resolution. The radiation damage is continuously monitored in-situ using a laser monitoring system. The system checks the relative loss of transmittance of the crystals using pulsed laser light of a known wavelength. Any change in the observed response is accounted for in the data during reconstruction. Non uniformities in the ϕ direction are also corrected for by inter-calibration of η rings using energy deposited by pile-up, underlying events, and decay of precisely known resonances such W and Z [21].

3.2.3 The Hadronic Calorimeter

The CMS hadronic calorimeter (HCAL) is situated behind the ECAL. It is designed mostly for the energy measurement of hadronic particles and missing transverse energy. Based on the pseudorapidity region, the HCAL is divided into the barrel region (HB), endcap region (HE), and forward detector (HF). HB covers $|\eta| < 1.3$ at a radius of $1.806 \text{ m} \leq r \leq 2.950 \text{ m}$ and is divided into 2 halves each consisting of 18 identical azimuthal wedges. Each wedge is divided into four sectors in ϕ . HE covers $1.3 < |\eta| < 3.0$, and HF covers $3.0 < |\eta| < 5.2$. HCAL also consists of extra scintillators, referred to as the outer calorimeter (HO), located outside the solenoid in the pseudorapidity region $|\eta| < 1.3$. A longitudinal view of the CMS HCAL along with its subcomponents is provided in Figure 3.9.

HB and HE are sampling calorimeters with an interleaving pattern of brass absorber plates and plastic scintillators. When a particle hits the absorber layer, the interaction produces secondary particles. The secondary particles comprise a wide range of particles with protons, neutrons, and pions as the major contributions. The secondary particles interact with the successive layer of absorber producing a cascade of particles making a hadron shower. The particle multiplicity depends on energy and type of the particle produced. Neutral pions and eta mesons produced in the hadronic shower decay to a two-photon final state giving rise to an electromagnetic component of the hadronic shower. Due to the important role the strong interaction plays in the interaction of hadrons, hadronic showers are usually more complex than electromagnetic showers. The associated nuclear interaction length (λ), which characterizes the mean path length required by relativistic

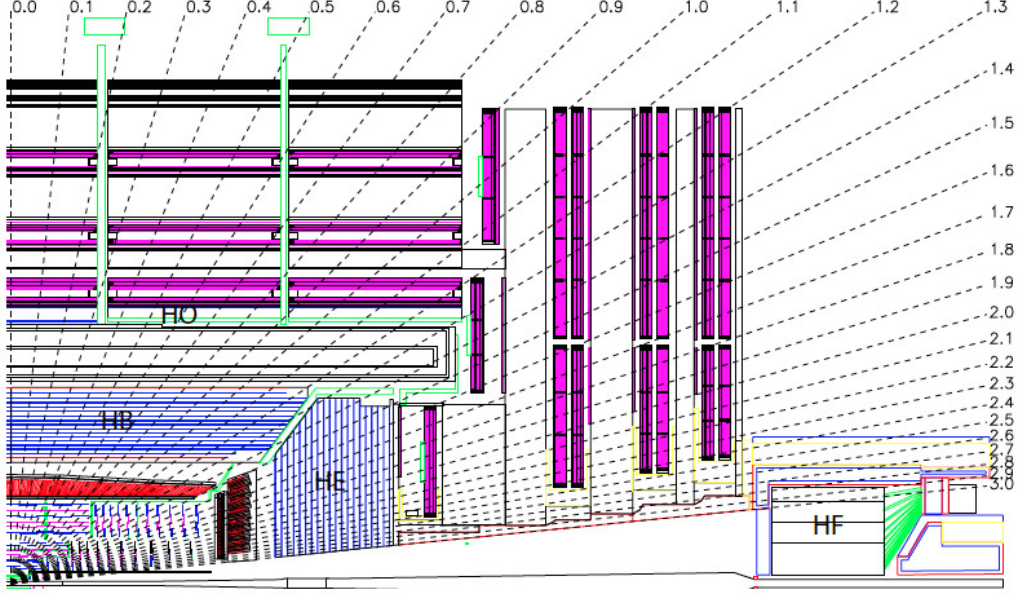


Figure 3.9: Longitudinal view of the CMS HCAL along with the various components (HB, HE, HF, and HO). Pseudorapidity values are overlaid on the diagram to show the coverage of each component [7].

charged particles to traverse before losing all but $1/e$ of their initial energy, for brass absorber is 16.42 cm. Though the absorber plates have a high density of 8.53 g/cm^3 , full shower containment is achieved only at roughly 6λ at $\eta=0$, and 10λ in the HE. This results in the large size of the HCAL, which in turn is a limiting factor in the choice of the HCAL material.

The secondary particles, which are part of the hadron shower, when passing through the scintillating layer emit violet light. The wavelength of the emitted light is shifted using wavelength shifting (WLS) fibres into green light. The WLS fibers are connected to optical cables that transfer the light to readout boxes arranged in strategic positions within the HCAL volume to avoid incident particle interaction. Multipixel hybrid photodiodes (HPDs) in the HB/HE/HO and radiation hard photomultiplier tubes (PMTs) in the HF convert the light signals to electric signals, which are then digitized with ADCs for readout [7].

Since the HF receives a larger flux of particles, it is designed with materials that can withstand

higher radiation doses. The HF is made up of an interleaving pattern of steel absorber and quartz fibers as scintillators. In the central rapidity region, hadronic showers may not be effectively stopped by the EB and HB. Therefore, the HO can catch late-starting showers by using the solenoid coil as an additional absorber and scintillator tiles as an active layer. With the four sub-detector components placed without projective gaps, CMS HCAL along with the ECAL is very hermetic and has $> 10\lambda$ coverage for all values of pseudorapidity.

Similar to the ECAL, the HCAL is also calibrated both on-site and off-site using various techniques. The ex-situ techniques involve measurement of light yield of each tile in a megatile (grouping of multiple scintillating tiles) before inserting it into the HCAL. A Cesium-137 gamma source is moved with computer control over each megatile, and the output signal is recorded. The energy responses are adjusted to match the incident energy. After inserting the tiles into the HCAL, each megatile is also scanned with a moving wire source. Another ex-situ technique includes using hadrons and muons of varying known energies in test beam experiments to determine absolute calibration between beam energy and energy response. Similar to that in ECAL, HCAL is also calibrated using ϕ symmetry of energy deposition from minimum bias events, single track hadrons from taus, jet balancing and dijet resonances. Another in-situ technique uses tunable UV laser light monitoring system to monitor any radiation damage, and uses the light output to adjust the HCAL response [7].

3.2.4 The Solenoid

The fourth layer of CMS is a solenoid that fits the tracker and calorimeter detectors inside its magnetic coil, while the muon detectors are interleaved with the iron “return yoke” that surrounds the magnet coils. It was designed to reach a 4 T magnetic field, and currently generates a 3.8 T field. For comparison, that is approximately 100,000 times the magnetic field of the Earth, and it is the largest superconducting magnet ever built. Superconducting solenoids are magnets made of coils that conduct electricity with zero resistance at low temperature and generate a magnetic field when current flows through them. They must be kept at low temperature with a cryogenic system. The CMS solenoid is cooled to a temperature of 4.5 K. With a 6.3 m diameter and 12.5 m length, it weighs 12 kilotons and can store 2.6 GJ of energy. Dump circuits are present to dissipate the energy in case there is quenching of the magnet.

The main function of the solenoid is to provide a large uniform magnetic field in which charged particles can bend. A large magnetic field causes a charged particle to have larger curvature. The curvature of its trajectory enables measurement of its momentum using hits in the tracker and the muon system. The higher the transverse energy of the particles, the less they will bend; therefore making it difficult to measure their momentum precisely. The solenoid's return yoke has a 14 m diameter and is also useful in filtering particles other than muons from reaching the muon system. Any charged particle reaching through the muon system will curve in the opposite direction outside of the magnet making a roughly 'S'-shaped path.

3.2.5 Muon Detectors

Muons and neutrinos produced in CMS are not stopped by the calorimeters. The main goal of muon system is to complement the tracker for the precise measurement of muon momentum. Specifically, it plays a role in the triggering, identification and reconstruction of the muon. The muon system comprise many subsystems and three gaseous detectors in a return yoke of the superconducting magnet. It is divided into a central part covering a pseudorapidity range $|\eta| < 1.2$, and a forward region of $1.2 < |\eta| < 2.4$. A longitudinal view of the muon system consisting of 250 Drift Tubes (DTs), 540 Cathode Strip Chambers (CSCs), and 610 Resistive Plate Chambers (RPCs) can be seen in Figure 3.10. The barrel region consists of DTs and CSCs organized in 4 concentric circles and 5 rings along the z axis. The muon flux as well as the neutron-induced background is small in the region where DTs are used. CSCs and RPCs are used in the endcaps, where the flux is high, as well as where the magnetic field large and non-uniform.

The DTs consist of 4 cm wide tubes with a wire within a volume of gas (85% Ar + 15% CO₂). When a charged particle passes through a DT, it knocks electrons off the atoms of the gas. The electrons follow an external electric field ending up at the positively charged wire. The hit position in the wire and the particle's distance away from the wire give two coordinates for the particle's position. DTs are utilized in both the measurement of muon momenta and the triggering.

The CSCs consist of an orthogonal arrangement of positively charged wires and negatively charged

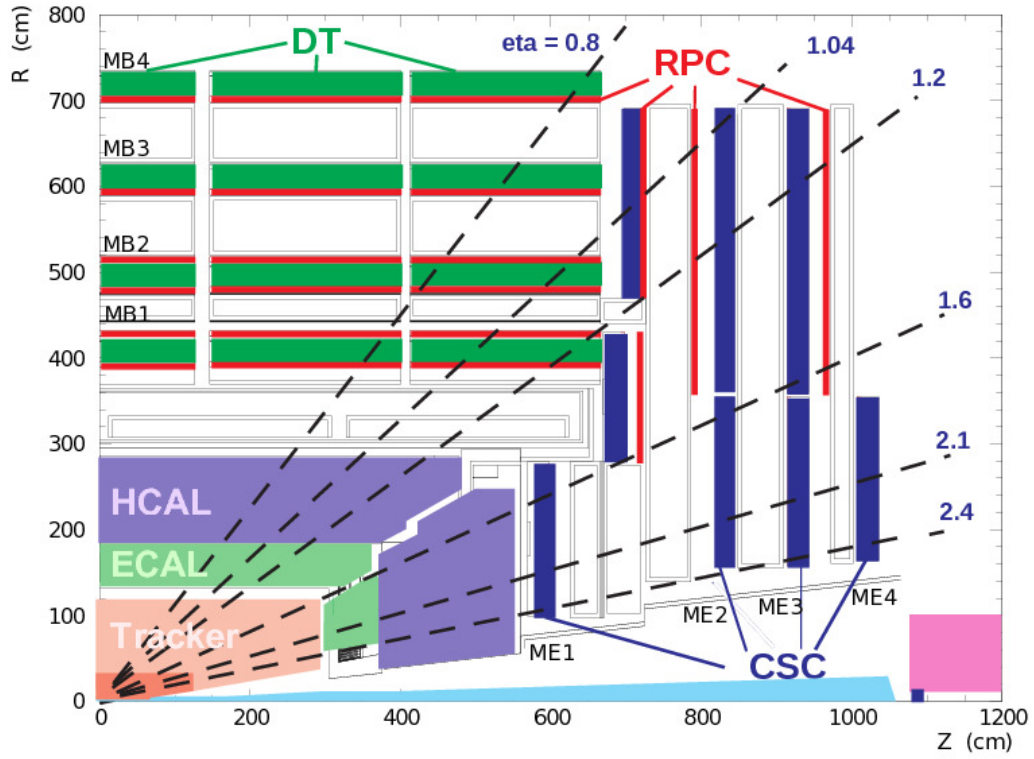


Figure 3.10: Schematic of one of the quadrants of CMS showing the four DT stations in the barrel (MB1-MB4, green), four CSC stations in the endcap (ME1-ME4, blue), and the RPC stations (red) [8].

strips in a volume of gas (40%Ar+50%CO₂+10%CF₄). The basic mechanism by which the position of a charged particle is registered is similar to that in DT. A particle passing through the gaseous volume knocks off electrons from the gas atoms, which move to the positively charged anode wires, allowing for the positively charged ions to flock to the negatively charged cathode strips. The cathode strips that are arranged radially outward and provide a precision measurement in the $r - \phi$ plane, whereas the anode wires provide a z measurement. CSCs are utilized in the triggering because of their ability to perform fast pattern recognition for rejecting non-muon backgrounds and efficiency in matching the hits to other muon system components and the CMS tracker.

The RPCs are a gas filled detector that consist of two parallel plates made of phenolic resin with a bulk resistivity of $10^{10} - 10^{11} \Omega\text{cm}$ separated by a 2 mm gap. The plates are coated with conductive graphite paint to form high voltage and ground electrodes. Aluminum strips separated from the graphite coating by an insulator film performs the readout. RPCs have excellent time resolution of 1 ns much smaller than the LHC nominal bunch crossing of 25 ns. Triggering utilizing the RPCs is therefore useful in identifying the correct bunch crossing associated with the muon tracks.

3.3 Triggers

During 8 TeV data collection, the bunch crossing frequency was 20 MHz. Each bunch crossing results in about 25 simultaneous pp collisions. However, not all of these collisions result in hard scattering. In a hard scatter there is a large momentum transfer between initial and final state partons, giving rise to interesting physics scenarios. So it is advantageous to implement a trigger system that preferentially selects hard scatter events. Furthermore, a typical raw data event size at CMS is of the order of 1 MB, which means every second CMS would need 20-40 TB of storage to store all events. It would be very challenging, if not entirely impossible, to sort and reconstruct all the events given the current technological constraints. Therefore, the trigger system at CMS is implemented to reduce the event rate to a more realistic value of about 400 Hz.

Triggering at CMS is done by implementing minimum identification based on calorimeter deposits or hits in the muon systems while keeping the full resolution data in pipelined memory. This is

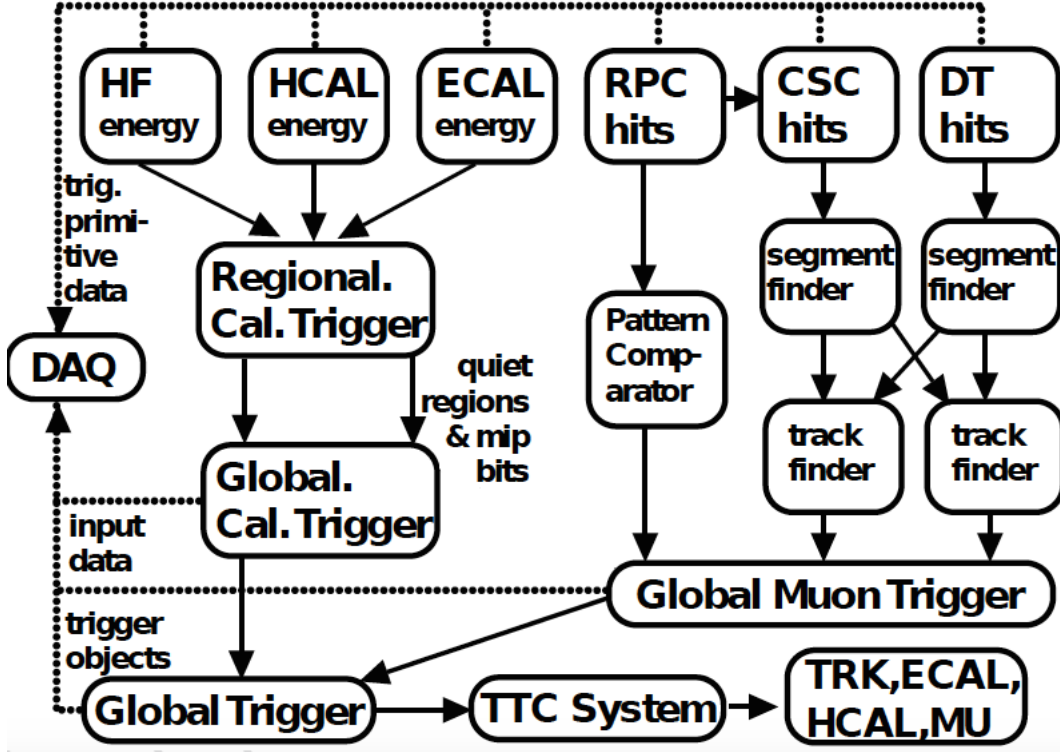


Figure 3.11: The L1 trigger schematics [9].

followed by online reconstruction of specifically interesting events. Thresholds are applied so that the efficiency for selecting interesting events is not depleted. The trigger system at CMS achieves its goal through a two level structure. The Level 1(L1) trigger reduces the rate from 40 MHz to 100 kHz, which is then further reduced to 400 Hz by the High-Level Trigger (HLT).

3.3.1 Level 1 Trigger

The L1 trigger is a completely automatic hardware-based trigger system. It selects 1 out of approximately 100,000 events. Figure 3.11 shows the schematic for the L1 trigger. There are multiple stages to L1. Information from the calorimeters and several components of the muon detector go through local and regional stages before the final Global Trigger decision is made.

The Regional Calorimeter Trigger (RCT) gets information on the quality of the energy deposition

and the transverse energy from the 8000 ECAL and HCAL towers (layers of HCAL cells) in $|\eta| < 5.0$. The RCT processes the information to determine e/γ candidates and sums of E_T . The Global Calorimeter Trigger (GCT) uses this information to find jets, classifies them as forward, central, or tau using the transverse energy sums. It also calculates quantities such as missing E_T and sends information for isolated e/γ candidates, non-isolated e/γ candidates, central jets, forward jets, and taus.

The front end trigger electronics in the DTs and CSCs of the muon system identify track segments from hits in the muon system. They also identify the bunch crossing corresponding to the event from which each hit originates. They are transported via optical fiber to regional track-finders. The regional track-finders use pattern recognition algorithms to join segments and complete the tracks. Muon candidates are identified and their momenta calculated based on their curvature in the magnetic field within muon system. There is an overlap between the DT and CSC in pseudorapidity. Information is shared between the DT track-finder (DTTF) and CSC track-finder (CSCTF) in this region for efficient coverage. The hit information from RPC goes directly from the front end electronics to the pattern comparator logic board. The information from DTs, CSCs, and RPCs is then combined by the Global Muon Trigger to get better momentum resolution and efficiency than with the individual sub systems.

The Global Trigger (GT) is the final step of the CMS L1 Trigger and implements a trigger menu. A trigger menu is a set of selection requirements applied to the final list of objects (i.e. electrons/photons, muons, jets, missing E_T or taus). The menu includes triggers ranging from simple single object selections with transverse energy above a set threshold to complex topological requirements on muon and calorimeter objects at the same time. The GT is responsible for generating a L1 trigger accept, which triggers the readout of the detector and is required by the High Level Trigger algorithms. The L1 accept signal is delivered via the Timing Trigger and Control (TTC) system, after which the data is pushed from the front-end buffers in to the data acquisition (DAQ) system [9].

3.3.2 The High Level Trigger

Data have to be accepted by HLT in order to be recorded for offline physics analysis. The HLT processes all events accepted by L1 and consists of a combination of software modules that perform well-defined tasks. Together the modules work to process the events by quickly reconstructing parts of physics objects that can be used for selection. Algorithms similar to those applied offline are applied online for high reconstruction efficiency and better object identification. The trigger path consists of reconstruction and filtering sequences where faster algorithms are run first. Events that fail the faster algorithms are thrown away first so that fewer and fewer events remain as the complexity of the algorithms goes up. This helps in reducing the total processing time.

HLT paths range from simple paths with cuts on object kinematics to complicated paths involving filters on reconstructed quantities for various objects. For example, one of the HLT paths that will be used in the measurement presented in this thesis is `HLT_Photon30_CaloIdVL_v*`. This particular path applies a transverse momentum threshold of 30 GeV on photon objects and a ‘very loose’ threshold on shower width associated with the photon reconstructed online. The detailed information on what the ‘very loose’ criteria entail will be presented in Chapter 5. Furthermore, often HLT paths corresponding to physics scenarios with large cross section are ‘prescaled’ where prescaling refers to storing only a fraction of events that would otherwise pass the trigger criteria. Prescaling is implemented to optimize CMS’s ability to perform interesting analyses in various different processes and channels ranging from very rare to comparatively common processes.

3.3.3 CMS Luminosity Monitoring

The luminosity delivered by the LHC is not the same as the luminosity recorded by CMS. CMS keeps track of both of these luminosities separately. However, for offline analyses the relevant luminosity is the recorded luminosity. CMS uses the HF to monitor instantaneous luminosity via a ‘zero counting’ method [22]. The occupancies of HF towers are monitored to get information on the number of interactions per bunch crossing. This is performed very rapidly, which allows for precise luminosity measurement every few seconds. Information from the linear relationship between the sum of the transverse energy and the luminosity is also exploited. The luminosity monitoring done

with the HF suffers from inaccuracies due to calibration drifts because of gain changes in the HF photo multiplier tubes and non-linear detector response due to pileup.

There is a complementary offline method of Pixel Cluster Counting (PCC) that is implemented to monitor the luminosity. This method counts the average number of pixel clusters occurring in a zero-bias event (an event triggered by requiring only that two bunches cross at the CMS interaction point). The instantaneous luminosity is proportional to the average number of clusters per event ($\langle n \rangle$):

$$\mathcal{L} = \nu \frac{\langle n \rangle}{\sigma_{vis}}. \quad (3.10)$$

Here, σ_{vis} is the visible cross section, and ν is the beam revolution frequency. The absolute calibration scale is determined by working with other LHC experiments to perform the Van der Meer scans, which involves scanning the LHC beams to determine the size of the beams at the collision points. The relationship between luminosity and beam quantities was presented in Equation 3.4. The absolute luminosity determined with this method is used to calibrate the HF luminometer [23].

CHAPTER 4

EVENT RECONSTRUCTION AND IDENTIFICATION

“All things are subject to interpretation, whichever interpretation prevails at a given time is a function of power and not truth.” - Friedrich Nietzsche

Once an event passes the trigger system, it is stored to be used for data analysis. To that end, the raw information from CMS detector components such as hits from the tracker and muon system as well as energy depositions from the calorimeters needs to be reconstructed into physics objects such as electrons, muons, photons and jets. The objects that are of importance to the analysis presented in this thesis are photons and jets. Electrons and muons are also used in the estimation of the systematic uncertainties. The reconstruction algorithm for each of these objects is unique. However, only the algorithms that are relevant to this analysis are described below.

4.1 Track Reconstruction

Though tracks by themselves are not physics objects, they form the backbone of the reconstruction of almost all the physics objects at CMS. Track reconstruction occurs in five phases. The first step of the process is the reconstruction of hits in the pixel and strip detector. This is achieved through a process known as ‘local reconstruction’ in which the raw signal outputs from the tracker electronics are clustered together into hits. The geometrical position of the hit and its uncertainty are also incorporated in this step. In the remaining four steps, together known as ‘global reconstruction’ phase, the hits obtained from local reconstruction are used to estimate the trajectory of the charged particles. Reconstruction of charged particles is important for vertex (origin of particles in a collision) reconstruction, which impacts the reconstruction of every object. Conversely, a rough estimate of the primary vertex is necessary for track reconstruction process itself. Charged particle track reconstruction is performed using an algorithm called the Combinatorial Track Finder

(CTF), which is an adaptation of the Combinatorial Kalman Filter - an extension of Kalman filtering [24]. The Kalman filter is a recursive estimation algorithm that forms the basis of most of the CMS fitting techniques. It is used to estimate the states of a dynamical system. It works by updating the trajectory estimate by adding one reconstructed hit at a time, updating the trajectory parameters and then using the previously estimated parameters to search for new hits to add. The predictions are updated and repeated multiple times to arrive at a more accurate estimate [25].

The global track reconstruction occurs in four stages: seed generation, trajectory building (seeded pattern recognition), trajectory cleaning (resolution of ambiguities), and trajectory smoothing (final fit). Seeds are first estimated either from triplets of hits or from pairs of hits with additional constraint from the ‘beam spot’ - the luminous region formed by the proton-proton collisions. Once the seeds are selected, they are propagated outward, adding newfound hits in each layer. The track parameters are updated at each layer and the track trajectory is built. The trajectory is dependent on the energy and charge of the particle. Gaussian energy loss and multiple-scattering between layers are also taken into account while searching for new hits. Different scenarios are considered while constructing the trajectory, and if there are multiple possible hits, all the trajectories are formed in parallel to avoid any bias. After no new hits are found in the tracker, the hit collections are re-fit to obtain the best estimate for the track parameters in the last layer. The algorithm is initialized at the innermost hit location. The fit works in an iterative way going through all the hits from inside to out updating the trajectory estimate. In the next step, the trajectories are cleaned by resolving the ambiguities (single seed resulting in multiple trajectories). It is done by removing the incompatible hits to reduce inefficiencies. The compatibility of a hit is measured in relation to the rest of the candidates for the track.

When the trajectory is built and ambiguities are resolved, parameters for the surviving trajectories are known at the inclusion of the last hit only. The track parameters that are used are d_0 , z_0 , ϕ , $\cot\theta$, and the p_T of the track. d_0 and z_0 are the coordinates of the impact point (point of closest approach of the track to the beam axis) in the radial and z directions, respectively. ϕ and θ are the azimuthal and polar angles of the momentum vector associated with the track. In order to obtain the optimal estimates of the measurements at each point in the track and better

rejection of the outliers, a final fit is performed. It is initialized using the results of the previous fit and fit inward to the beam-line starting from outermost hit. This phase is known as smoothing. The track parameters that are ultimately considered at each layer are the weighted averages of the parameters associated with the two (inside-out and outside-in) fits [26].

The tracking procedure is iterated many times, each time working on hits that have not yet been associated to the tracks. The main distinction between iterations is in the seed generation. Iteration 0 of the CTF algorithm is designed for tracks with $p_T > 0.8$ GeV and having three hits in the pixel detector. This constitutes most of the reconstructed tracks. Iteration 1 is designed for events with only two hits in the pixel detector. Iteration 2 is for lower p_T tracks. Iterations 3 through 5 are performed to find tracks that were not obtained with previous iterations and those that originate outside of the beam spot.

4.2 Primary Vertex Reconstruction

A vertex is the proton-proton interaction vertex from which the tracks originate. The reconstruction of the primary vertex (PV) is carried out in three steps: selection of the reconstructed tracks, clustering of the tracks that appear to have originated from the same vertex, and fitting of the tracks to estimate the position of the vertex. The tracks are first selected from the collection of reconstructed tracks based on their compatibility with the beam spot, numbers of pixel and strip hits, and the normalized χ^2 from a trajectory fit. The selected tracks are then extrapolated back to the nominal interaction point in the beamline. They are clustered on the basis of their z_0 values. Those candidates that have at least two tracks are then fit with a 3 dimensional vertex fit under the assumption that the tracks originate from the same interaction point. The fitting algorithm is based on a standard Kalman filter and is referred to as an ‘adaptive vertex fit’ algorithm [27]. It does not reject outlying tracks, but it gives the tracks a small weight during fitting. The position of the primary vertex is determined from this fit. If the tracks do not originate from the nominal interaction region or if a track splits into multiple tracks along its trajectory, the vertex is known as a secondary vertex [26]. For a particular event, a vertex associated with the prompt tracks (tracks corresponding to the hard interaction) is selected by choosing a PV that has the largest sum of

the p_T^2 of the tracks, since the tracks with small p_T are likely to have originated from pile-up events.

4.3 Photon Reconstruction

As mentioned in Chapter 3, high-energy photons deposit most of their energy in the ECAL through electromagnetic shower generation. However, because of the presence of material in front of the ECAL, the particles have non negligible probability of interacting with the material in front. As a result, the charged particles present in the electromagnetic shower are often spread in the ϕ direction due to the magnetic field. In order to obtain an accurate measurement of the photon's energy, it is essential to collect these bremsstrahlung photons and electron-positron pairs. At CMS, the spread of energy is reconstructed by clustering algorithms that group energy depositions in ECAL crystals of a fixed array, which are then grouped together to form superclusters (SCs). Prior to the clustering, the signal amplitudes from ECAL electronics are digitized and intercalibrated. CMS implements two different methods of clustering an electromagnetic shower: hybrid superclustering in the barrel region and multi5 $\times$ 5 superclustering in the endcaps.

The hybrid superclustering algorithm's goal is to exploit the showers spread in the ϕ direction to group energy depositions in a small η window while allowing for the cluster to be wide in the ϕ direction. It starts with the selection of a seed crystal that has the local maximum energy and is greater than the transverse energy threshold $E_{T,hybseed}$. Starting with the highest energy seed, a rectangle of 3×1 dimension in the $\eta - \phi$ plane is formed around the seed. If the central crystal of the array has energy greater than a threshold E_{wing} , a 5×1 rectangle is formed with the seed crystal at the center. The algorithm takes steps within a configurable N_{steps} in the $\pm\phi$ direction about the crystal adding the neighboring rectangles if their energy is above a threshold E_{thresh} . Since the shower does not spread in η , the η -width of the cluster remains as five crystals. Adjacent arrays are then grouped in ϕ to form local maxima if each seed array has a threshold higher than E_{seed} [28]. The local maxima are termed as “basic clusters”.

The multi5 $\times$ 5 method has fewer parameters than hybrid clustering does. The idea behind the clustering is similar in that both group the energy depositions into small clusters, which make up

Table 4.1: Parameters related to the ECAL clustering and the associated threshold values.

Barrel		Endcaps	
Parameter	Threshold	Parameter	Threshold
$E_{T,hybseed}$	1 GeV	$E_{T,EEseed}$	0.18 GeV
E_{wing}	0	$E_{T,EEcluster}$	1 GeV
N_{steps}	17	η_{range}	0.07
E_{thresh}	0.1 GeV	ϕ_{range}	0.3 radian
E_{seed}	0.35 GeV		

the SCs. However, the algorithm works in an opposite manner in that hybrid algorithm first forms a SC that is decomposed into basic clusters; whereas multi5 $\times$ 5 first makes the basic clusters that are grouped into SC. Multi5 $\times$ 5 begins with seed crystals which are the local maxima with the highest energy deposition compared to the neighboring crystals. They are also required to have an E_T higher than threshold $E_{T,EEseed}$. A 5 $\times$ 5 array of crystals is formed around each seed starting from the highest energy seed to form clusters. The clusters are allowed to partially overlap, without sharing the crystals adjacent to the seed. To group the clusters into a SC, the individual clusters are required to have an E_T greater than a threshold of $E_{T,EEcluster}$ and fall within a predefined $\eta_{range} \times \phi_{range}$ window. In the endcaps, the preshower (ES) energy needs to be added to the SC energies. To cluster the energies in the ES, the energy weighted position of all the clusters belonging to a SC are extrapolated to the ES surface with the most energetic cluster as a reference point. The ES energies within $\eta = \pm 0.15$ and $\phi = \pm 0.15$ from the reference point are added into the SC energy. The preshower energy is measured from the clustered energy in each layer of the double layered preshower detector. The threshold values used in the 8 TeV reconstruction are listed in Table 4.1 [28].

After the superclusters are reconstructed, the position of a SC is measured from the log energy weighted mean of the basic clusters. The energy of a SC is corrected as a function of η , E_T , R_9 , and the spread of energy in ϕ . The variables are described in detail in Section 4.3.1. The correction is required to improve the energy resolution. Several factors related to the ECAL geometry, shower containment in the ECAL, and the energy containment in the superclusters necessitate the energy correction. The correction is obtained by observing energy losses in simulated $Z \rightarrow ee$ events and

are validated using data [28]. The reconstructed SC energy can be obtained as

$$E_{e,\gamma} = F_{e,\gamma} \sum_i G(\text{ADC}/\text{GeV}) \times S_i(T, t) \times c_i \times A_i + E_{ES}, \quad (4.1)$$

where A_i is the raw energy deposited in each ECAL crystal digitized by a sampling ADC, c_i is the inter-calibration coefficient, $S_i(T, t)$ is a correction factor assigned to account for radiation induced response changes as a function of time, G is a factor to convert the energy from units of ADC counts to GeV, $F_{e,\gamma}$ is the energy correction applied to the energy deposits, and E_{ES} corresponds to the energy in the preshower, which is relevant only for endcap clusters [29].

4.3.1 Photon Identification

The analyses that have signal photons must remove the large background from the diphoton decay of neutral mesons. These photons are collimated and reconstructed as a single photon. However, unlike the signal photons, the photons from meson decay are non-isolated since they are typically part of a hadron shower. The reduction of background is often accomplished by placing isolation requirements on the photon candidates. Isolation measures the amount of energy present in a cone of fixed radius ($\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$) around the center of the shower. The photon identification prescription used in the photon+jet cross section measurement consists of a cut on a particular type of isolation called particle flow (PF) isolation. There are three kinds of PF isolation: charged hadron isolation, neutral hadron isolation, and photon isolation. These isolations are obtained by summing up the energy of particle flow objects (described in Section 4.4) in a cone of $\Delta R = 0.3$ around the photon candidate. As the name suggests, only PF charged hadrons, PF neutral hadrons, and PF photons are summed in the isolation cone to obtain the charged hadron (CH), neutral hadron (NH), and photon isolation, respectively. Before being used in the analysis, these isolations are corrected for pile-up using the following expression:

$$Iso_{\text{corrected}} = Iso - \rho \times A_{\text{eff}}. \quad (4.2)$$

Here, ρ is the estimate of the energy density per unit area associated to the pile-up interactions and underlying events. For each event, it is measured by taking the median jet energy in a jet area clustered using anti- k_t jet algorithm (additional information provided in Section 4.4) [30]. Events

Table 4.2: Effective area for ρ correction.

η range	Charged Hadrons	Neutral Hadrons	Photons
$ \eta \leq 1.0$	0.012	0.030	0.148
$1.0 \leq \eta < 1.479$	0.010	0.057	0.130
$1.479 \leq \eta < 2.0$	0.014	0.039	0.112
$2.0 \leq \eta < 2.2$	0.012	0.015	0.216
$2.2 \leq \eta < 2.3$	0.016	0.024	0.262
$2.3 \leq \eta < 2.4$	0.020	0.039	0.260
$ \eta > 2.4$	0.012	0.072	0.266

with large underlying activities and pile-up will have large values of event ρ . The energy from pile-up inside the isolation cone is calculated on an event-by-event basis by multiplying ρ with an effective area (A_{eff}) for isolation. The effective area gives information about the susceptibility to soft contamination and is chosen such that the isolation is flat with respect to the number of pile-up interactions. In doing so, A_{eff} acquires an η dependence. Table 4.2 shows A_{eff} for different pseudorapidity regions. Once the pile-up energy inside the isolation cone is estimated, it is subtracted from the isolation energy inside the cone on an event-by-event basis.

Apart from the isolation, another variable that is used to identify hadrons is H/E . It is the ratio of the energy measured in a single tower of HCAL behind the supercluster to the energy of ECAL supercluster. Since photons and electrons deposit most of their energy in the ECAL, this quantity is smaller for photons and electrons in comparison to that for jets. By applying a cut on the upper value of H/E , background from hadrons is reduced.

Furthermore, due to the amount of tracker material present before the ECAL, on average one third of the photons pair produce before reaching the ECAL. Since these converted photons spread out in ϕ , the shower width in ϕ alone is not ideal to discriminate signal photons from background. However, the shower width in η can provide important discrimination between signal photons and background. As such, a shower shape variable in η ($\sigma_{i\eta i\eta}$) is an important quantity in the photon identification. It is summed over a 5×5 matrix around the seed of the SC. Mathematically,

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_{5 \times 5} w_i (\eta^i - \bar{\eta}^i)^2}{\sum_{5 \times 5} w_i}} \quad (4.3)$$

and

$$w_i = \max(0., w_0 + \ln(\frac{E_i}{E_{5 \times 5}})). \quad (4.4)$$

In the above equations, $w_0 = 4.7$ and $\bar{\eta}$ is the energy weighted mean η of the crystals present in the 5×5 with the seed crystal in the middle. A related quantity R_9 is used in distinguishing converted photons from unconverted ones. It is the ratio of the energy in the ECAL in a 3×3 matrix around the seed of the supercluster to the raw energy associated with the SC. Photons that have pair produced into electron-positron pairs due to interaction with the tracker material will have smaller values of R_9 , while the ones that have not will have larger values.

Electron veto is another useful variable for photon identification. It is applied to remove background electrons from the signal. As electrons do not constitute the largest background for the measurement covered in this thesis, conversion-safe electron veto (CSEV) is appropriate to use. CSEV is used to reject electrons other than the ones originating from photon conversion. A typical converted photon will have charged-particle tracks associated to a reconstructed conversion vertex, pointing to an ECAL cluster. This veto requires that there is no tracks with a hit in the inner layer of the pixel detector other than the ones related to converted photon [29].

The selection criteria applied on these photon identification variables are presented in Chapter 5.

4.3.2 Multivariate Photon Identification

The measurement presented in this thesis also uses a more sophisticated photon identification technique based on Multivariate Analysis (MVA). In the standard cut based identification approach, cuts are applied consecutively to many variables that have some ability to discriminate between signal and background. The caveat to using such an approach is that often times the relationships between the variables considered are quite complex and may be correlated. Applying consecutive cuts in the variables is an inefficient approach as the number of events may be drastically reduced. MVA can be trained for optimal discrimination between signal and background by simultaneously

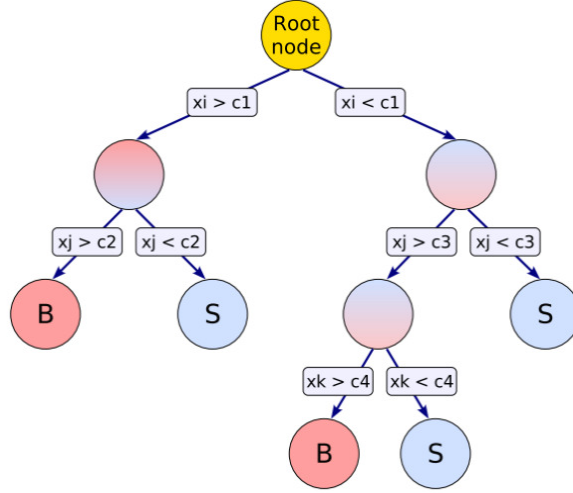


Figure 4.1: Schematic of a decision tree [10].

exploiting the discriminating properties of many variables and their correlations in order to give a single output.

A ROOT-integrated environment known as the Toolkit for Multivariate Analysis (TMVA) is used in this measurement to implement a multivariate technique. TMVA is a ‘supervised learning’ technique specifically designed for high-energy physics applications. In TMVA, a simulation-training sample with a known output is used to determine the mapping function describing the decision boundary (classification). In this analysis, a Boosted Decision Tree (BDT) classifier option is chosen. It works by making repeated binary (yes/no or left/right) decisions on single variables at a time until a stopping criterion is satisfied. The phase space of the training sample is divided into many regions, which are finally divided into either signal or background based on the final number of signal events in each phase space. In Figure 4.1 one can see that starting from the Root node the data is split into two sections based on the variable x_i . Each data branch is subjected to other binary decisions based on variable x_j . One of the branches is further split into two categories based on a binary decision on variable x_k . Both sides of the split data do not have to be separated based on the same variable. Similarly, a variable could be used at many different nodes and some may not be used at all. The decision at the final node (leaf) is either signal (S) or background (B)

depending on the nature of majority of the events that fall in those nodes, and are associated with a specific score. The path from the Root node down to a leaf represents an individual cut sequence that allows for the selection of signal or background events [10].

‘Boosted’ in the BDT term refers to enhancing the classification’s performance by making it more stable under statistical fluctuation of the training sample. It is done by building many decision trees (also known as forests) and adjusting the weights of individual events based on whether the previously trained tree classified them correctly. The final score is a weighted average of the scores the event gets from each tree in the forest. The score can be any value in between -1 and $+1$. However, the MVA is trained in order to maximize the difference between the distribution of scores for the signal and background events. Once the training is done, the scores for each tree in the forest can be applied to a statistically independent test MC sample. The results of the test sample can be compared to the MC truth information to check the performance of the MVA. In this measurement, the MVA is trained separately for barrel and endcap using a double EM enriched $\gamma + \text{jet}$ MC sample. The double EM enrichment introduces electromagnetic activity from neutral hadrons or electrons. The input variables used in this analysis along with the definitions are provided below.

- η width: Energy weighted width of the supercluster shower in η .
- ϕ width: Energy weighted width of the supercluster shower in ϕ .
- $\sigma_{i\eta i\eta}$: Described in Section 4.3.1.
- $\sigma_{i\eta i\phi}$: Shower width in $\eta - \phi$.

$$\sigma_{i\eta i\phi} = \sqrt{\frac{\sum_{5 \times 5} w_i (\eta^i - \bar{\eta}^i)(\phi_i - \bar{\phi}_i)}{\sum_{5 \times 5} w_i}}$$
. Here, w_i is similarly defined as in the case of $\sigma_{i\eta i\eta}$.
- R_9 : Described in Section 4.3.1.
- $S_4 = E_{2 \times 2} / E_{5 \times 5}$: Ratio of the highest value of energy deposit in the four possible combinations of 2×2 matrix containing the seed crystal to the energy deposited by the supercluster.
- η^{SC} : Pseudorapidity of the supercluster.
- ρ : Described in Section 4.3.1.
- $\sigma_{iRiR} = \sqrt{(\sigma_x)^2 + (\sigma_y)^2}$ is calculated only for endcap. It is the width of the supercluster in the preshower where $x - y$ coordinates are used instead of $\eta - \phi$.

Table 4.3: Linear corrections applied to MC for some variables before entering the MVA. The corrected value, $x_{corr}=a \times x + b$.

Variable	$ \eta^\gamma < 1.5$		$1.5 < \eta^\gamma < 2.5$	
x	a	b	a	b
η width	1.04302	-0.000618	0.903254	0.001346
ϕ width	1.00002	-0.000371	0.99992	-0.00000048
$\sigma_{i\eta i\eta}$	0.891832	0.0009133	0.99470	0.00003
S_4	1.01894	-0.01304	1.04969	-0.03642
R_9	1.0045	0.0010	1.0086	-0.0007

The MVA is also validated on events where electrons are reconstructed as photons only using their supercluster energy deposition. The data/MC comparison for the MVA output gives linear corrections to some variables. These corrections are applied to the MC variables before being used in the MVA. The linear corrections are provided in Table 4.3. The signal events tend to peak close to +1, while the fake events peak at both +1 and -1. The weights obtained from MVA training are later applied to signal MC, background MC, or data to get an appropriate MVA response.

4.4 Particle Flow Algorithm

The particle flow (PF) algorithm is used to identify and reconstruct all the stable particles at CMS by combining information from different sub-components of the detector. For some physics objects, this allows for a better determination of the particle species, energy, and momentum compared to the standalone algorithms. Object type, i.e. photon, electron, muon, charged hadron and neutral hadron, is determined based on the information combined to construct the object. Muons are reconstructed initially so that they are not mistaken for charged hadrons. A PF muon candidate will be reconstructed by combining tracking information from the inner tracker and the track in the muon system. Charged hadrons are identified by associating the tracks in the inner tracker with energy depositions in the ECAL and HCAL. To link the tracks to the calorimeters, they are extrapolated to the front face of the calorimeters. If the tracks fall within the energy clusters in the calorimeters, they are linked together as long as they have not already been linked previously. A PF electron will have a track in the inner tracker linked to the ECAL energy deposition. After

all the tracks are properly linked, photons and neutral hadrons are reconstructed from remaining energy depositions in the ECAL and HCAL, respectively. From the list of simple reconstructed particles, higher level physics objects such as jets, missing transverse energy (E_T^{miss}), hadronic taus, and b-jets are reconstructed [31].

4.4.1 Calorimeter Energy

The calorimeter energy is an important component in the reconstruction of many PF objects. The clustering is done separately for ECAL barrel, ECAL endcaps, HCAL barrel, HCAL endcaps, and the two layers of preshower. In the particle flow clustering for ECAL, seed crystals are selected if their energy is the highest energy deposition among their neighboring crystals and the energy is greater than 230 MeV in the barrel and 600 MeV (or $E_T > 150$ MeV) in the endcaps. The algorithm starts with the highest energy deposit then systematically searches around the detector if the crystal passes a certain energy threshold. Once it comes across a crystal below the threshold it changes direction. This process is continued until it sums up all the adjacent crystals with energy 2σ above the electronic noise observed at the beginning of the run. The process is repeated for other seeds or local maxima passing the certain threshold. In this algorithm, two or more clusters are allowed to share energy of one crystal [28]. If there is more than one seed in a cluster, then the energy of the cluster is split between the constituent seeds.

In the HCAL, the highest energy tower compared to the neighboring towers is treated as the seed, starting from which the towers are grouped together to form clusters. The cluster is formed in the HCAL in a similar manner to the ECAL.

4.4.2 Linking

In order to reconstruct the PF objects, the energy depositions and tracks from various sub-detectors need to be linked to one another. Linking refers to the process of combining information from each sub-detector while ensuring that there is no double counting of information. To that end, the linked information is removed before reconstructing the remaining particles. The link distance is defined based on the ΔR between the elements being linked. The charged particle tracks in

the silicon trackers are extrapolated to the ECAL face, and if a track lies within the boundary of the cluster, it is matched to the cluster. In the endcaps, the tracks are also propagated to the preshower layers. The linking procedure also allows collection of Bremsstrahlung photons coming from electrons as they pass through the tracker by matching ECAL clusters that are tangential to the reconstructed tracks. The ECAL and HCAL clusters are also linked in a similar way. Since the ECAL has better resolution, the cluster position is defined in the ECAL and extrapolated to the HCAL. For muons, the linking is done between silicon tracker and track segments in the muon system. The information is matched based on a fit. A detailed explanation is provided in Section 4.7.

4.5 Jet Reconstruction

After the individual charged hadrons, neutral hadrons, and photons are reconstructed, they are combined into particle flow jets by feeding the list of PF objects into a jet algorithm. Almost 65% of the jet energy is carried by charged hadrons and 25% by photons, which are measured very precisely by the tracker and ECAL. The remaining 10% contribution comes from neutral hadrons which can be measured with HCAL only. As described in Chapter 3, the HCAL resolution is very poor compared to the resolution of ECAL and the tracker. Hence, the jet energy is measured more accurately with the particle flow than with traditional jets from calorimeter towers. Furthermore, the PF jet reconstruction is more similar to the Monte Carlo jet clustering in that the inputs for the algorithm are individual particles rather than detector measurements.

4.5.1 Jet Algorithm

Jet algorithms provide mapping between QCD final states (jets) and the objects that appear in the QCD Lagrangian (quarks and gluons) or those reconstructed in the detector. The algorithms define parameters to quantify how likely it is that the two partons have come from a QCD splitting. The algorithms, then, sequentially combine the partons based on those parameters in order to construct jets. At CMS, jets are reconstructed by using the anti- k_t algorithm, which is a sequential recombination algorithm — one of two broad classes of jet algorithms. Sequential recombination algorithms take pairs of particles that are close to each other and merge them. The process is

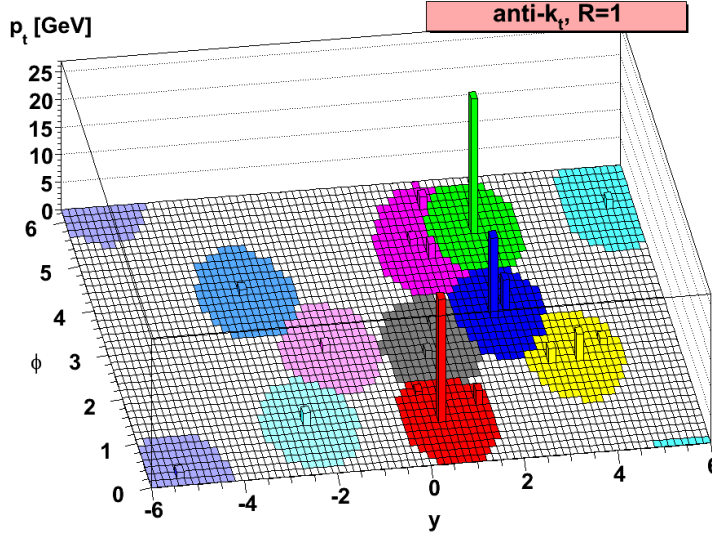


Figure 4.2: Diagram showing jet clustering using the anti- k_t algorithm. The different colors correspond to areas corresponding to different jets [11].

repeated until some criterion is met. As an example, consider a jet composed of N particles with momenta $k_1, k_2, \dots, k_N$. When k_i becomes very small or k_i and k_j become collinear, the same jets can be obtained by leaving out k_i in the first case or replacing k_i and k_j with $k_i + k_j$ in the latter case. Therefore, in an anti- k_t algorithm, the soft radiation from hadronization and underlying event model will not affect jet reconstruction.

The algorithm begins with a list of objects for which we can define distance parameters d_{iB} and d_{ij} such that

$$d_{iB} = k_{ti}^{2p} \quad (4.5)$$

and

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p})[(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2]/R^2, \quad (4.6)$$

where d_{iB} is the closest approach of the object to the beamline, d_{ij} is the distance between two objects, and $p = -1$. The distance parameter (R) is chosen as 0.5 for the measurement presented in this thesis. It ensures that any pair of final jets is separated at least by 0.5 in $\eta - \phi$ space. For every i , the algorithm selects the smallest of the d_{iB} and d_{ij} as d_{min} . If d_{min} is equal to d_{ij} , then

objects i and j are combined into a single particle, and will be added to the list of objects. If d_{min} is equal to d_{iB} , then jet i will not be merged and removed from the list of objects, and added to the list of jets. The procedure is repeated until no more objects remain.

The value of p controls which particles are combined first. The negative value in the anti- k_t algorithm ensures that the higher momentum particles are added before adding the softer particles, which gives a smooth surface to the jets as shown in Figure 4.2. Also, the anti- k_t algorithm is infrared safe, which means that the soft and collinear divergences do not affect the result of the jet algorithm. If a hard particle only has soft particles nearby, the final jet will have a jet size with radius R . However, if the hard jet has another hard jet between the radial distance of R and $2R$, then the soft particles surrounding them will be split between the two jets weighted by the jets' relative energy.

4.5.2 Jet Energy Corrections

The energy of a jet needs to be corrected to account for several effects such as contributions from pile-up, and the non-uniform and non-linear response of the detectors. The goal is to get the jet energy as close to its true value as possible. Therefore, the corrections map the measured jet energy deposition to the particle-level jet energy. They are performed in multiple steps, which are applied sequentially. In each steps, the jet's four-momentum is scaled by a correction factor. Level 1 ($L1$) correction is applied to correct for the effects due to pile-up. This correction is determined from QCD Monte Carlo events simulated with and without pile-up as a function of ρ (described in Section 4.3.1), A (area of the jet), p_T , and η . $L2$ corrects for the variation of response in η using MC truth simulations and data-driven techniques. The amount of tracker material and material from other CMS sub-components that jets have to traverse is dependent on the pseudorapidity. At lower pseudorapidity, jet energies are measured more reliably; whereas at higher values of the pseudorapidity, the increase in material leads to poor jet energy measurement. Therefore, the η -dependent $L2$ correction is obtained by comparing jets of known energies at some higher values of η with jets produced at low pseudo-rapidity. The third correction ($L3$) is applied to correct for the non-uniform response of the detector as a function of p_T . After the $L2$ correction is applied, the jet

p_T needs to be adjusted to the true jet p_T . This correction is also obtained from MC simulations [31].

4.5.3 Jet Identification

The PF jets are identified by using variables that help in discriminating jets from noise. Most of the variables used for jet identification are geared towards separating real jets from the unphysical ones. Since jets are composed of multiple particles, the number of constituents in PF jets can help distinguish background. About 65% of the jet energy is carried by charged hadrons, 25% by photons, and 10% by neutral hadrons. Therefore, jet identification includes variables that quantify the fraction of the energy measured in ECAL/HCAL by neutral hadrons. It is highly unlikely for a real jet to have almost all of its ECAL/HCAL energy come from neutral hadrons. The tracker fiducial region ends at $|\eta| = 2.4$. Quantities such as charged hadron multiplicity and fraction of energy deposited in ECAL/HCAL by charged hadrons can also be applied for $|\eta| < 2.4$, where the PF jets are expected to have at least one charged hadron object, and their contribution to the ECAL/HCAL energy is expected to be non-negligible. The selection criteria applied on the jet identification variables are presented in Chapter 5.

4.6 Electron Reconstruction

Since electrons are charged electromagnetic objects, they leave hits in the inner tracker and deposit energy in the ECAL through shower generation. Therefore, electron reconstruction at CMS relies on there being a track in the inner silicon tracker and an associated energy deposition in the ECAL, reconstructed as a supercluster. The algorithms used for clustering the ECAL energy deposits have already been discussed in Section 4.3. Once the SCs are reconstructed, either ECAL-driven or tracker-driven seeding methods are implemented to find tracker seeds, which are generated from two or three hits in the tracker to start building electron trajectories.

The tracker-driven seeding uses general charged particle tracks, extrapolates the trajectories to the ECAL surface, and links to the SCs. Tracks are reconstructed using the standard Kalman filter track reconstruction algorithm for charged particles. Since electrons can lose a lot of energy to

Bremsstrahlung, a large number of hits may be missed. Therefore, a dedicated track reconstruction procedure is implemented for electrons.

In the ECAL-driven seeding, the algorithm starts from an energy cluster in the ECAL that is spread out in ϕ due to bremsstrahlung. The algorithm extrapolates the trajectory towards the collision vertex and selects trajectories in the pixel detector corresponding to isolated high-energy electrons. The backward propagation of the trajectory is performed for both positive and negative charge hypotheses. A requirement on the supercluster of $E_T > 4$ GeV and $H/E < 0.15$ is placed to limit number of misidentified seeds.

Once the electron seeds are generated, they are used to build tracks using combinatorial Kalman Filter method. When compatible hits are found in each layer of the tracker, up to five candidate trajectories are built. Since electron loses a lot of its energy to bremsstrahlung, its change in curvature is large in the tracker. Using the standard Kalman filter with a Gaussian energy loss assumption can result in low hit collection efficiency and a poor estimation of the track parameters because of the changing curvature. Therefore, after all the hits have been collected, a Gaussian Sum Filter (GSF) is used to fit and estimate the track parameters. The GSF accounts for the electron’s non-Gaussian energy loss mechanism [28].

4.7 Muon Reconstruction

Muons are reconstructed using information from the inner tracker and hits in one or more of the muon systems (drift tubes, cathode strip chambers, and resistive plate chambers). Additionally, muons can leave some of their energy in the calorimeters. Based on the difference in reconstruction method, muons are categorized into “global muons” and “tracker muons”. Global muons are seeded from standalone tracks in the muon systems, whereas tracker muons are seeded from the tracker tracks. Standalone tracks are those that are reconstructed solely from the hits and segments in the muon system. Similarly, tracker tracks are reconstructed solely from hits in the inner tracks. Both kinds of tracks begin with seeds generated from track segments and consisting of estimates for position, direction vector, and momentum of the muon. Track fit is performed iteratively by

running over all available tracks using a Kalman filter technique.

Once the standalone tracks and tracker tracks are reconstructed, for each standalone track the global muon fitter searches for matching tracker tracks. All the hits in the linked tracks are fit using a Kalman filter technique to reconstruct global muons. The tracker muon filter works in reverse in that it extrapolates the tracker track into the muon detectors to link them to standalone tracks. The fitting of linked tracks is performed using Kalman filter technique as in the case of global muons. This approach assumes that all the tracks in the inner tracker within a specific range of p_T originate from muons. For low p_T muons that do not have enough energy to leave enough hits in the muon system, this approach is particularly beneficial. Therefore, tracker muons are more efficient than global muons for low p_T muons.

CHAPTER 5

ANALYSIS

“When a truth is necessary, the reason for it can be found by analysis, that is, by resolving it into simpler ideas and truths until the primary ones are reached.” - Gottfried Leibniz

The previous chapters on the motivation, theory, experiment, and event reconstruction lay the foundation for this measurement. Once the physical objects are reconstructed, the measurement of the triple differential cross section for photon+jet events as a function of photon transverse momentum (p_T^γ), photon pseudorapidity ($|\eta^\gamma|$) and jet pseudorapidity ($|\eta^{jet}|$) can be experimentally determined from the following expression:

$$\frac{d^3\sigma}{dp_T^\gamma d|\eta^\gamma| d|\eta^{jet}|} = \frac{1}{\Delta p_T^\gamma \Delta |\eta^\gamma| \Delta |\eta^{jet}|} U \frac{Np}{\epsilon \mathcal{L}'} \quad (5.1)$$

Here, N is the number of candidate events, p is the signal purity, U is the unfolding (i.e. the mapping between the true and measured quantities), ϵ is the efficiency of detection, $\mathcal{L}'$ is the effective integrated luminosity, and Δi is the bin size for object i .

The measurement is accomplished in the following steps:

- **Selection of the candidate events from data:** This is done by triggering on single photon events and applying selection cuts to the photon and jet candidates.
- **Extraction of signal from the candidate events/Estimating the purity:** Signal fraction/purity is obtained by fitting a distribution for the candidate events with signal and background templates. Signal templates are obtained from simulation, and background templates are obtained from data sidebands. The data sideband is optimized to reduce bias introduced through the correlation between the fit variable and the sideband variable.

- **Efficiency calculations:** Inefficiencies that are introduced to the measurement from multiple areas such as triggers, photon clustering, selection cuts, etc. are corrected.
- **Triple differential cross section measurement:** The measurement is made in 224 bin combinations of p_T^γ , $|\eta^\gamma|$, and $|\eta^{jet}|$ using Equation 5.1. The bins corresponding to $|\eta^\gamma|$ and $|\eta^{jet}|$ are: 0–0.8, 0.8–1.5, 1.5–2.1 and 2.1–2.5. The p_T^γ bins for $|\eta^\gamma| < 1.5$ are: 40–45, 45–65, 65–72, 72–90, 90–97, 97–105, 105–117, 117–165, 165–172, 172–180, 180–188, 188–196, 196–204, 204–215, 215–230, 230–245, 245–260, 260–285, 285–335, and 335–1000 GeV, and for $1.5 < |\eta^\gamma| < 2.5$ are: 40–65, 65–90, 90–105, 105–165, 165–180, 180–200, 200–225, and 225–1000 GeV.
- **Unfolding measurement:** The reconstructed p_T^γ is mapped to true values of p_T^γ to account for detector effects that smear the distribution. This process of unsmearing the distribution is accomplished via a process known as unfolding.
- **Uncertainty measurement:** The uncertainties associated with each of the above steps (purity measurement, efficiency measurement, unfolding, and additionally the luminosity measurement) are individually estimated. Error propagation is performed on equation 5.1 to estimate the total uncertainty.

5.1 Dataset

The data from CMS is separated into different categories based on the triggers used. This measurement uses 19.7 fb^{-1} of single photon triggered data at a center-of-mass energy ($\sqrt{s}$) of 8 TeV. The data were collected during 2012 and were reconstructed in January of 2013. They are categorized into ‘Run2012A’, ‘Run2012B’, ‘Run2012C’, and ‘Run2012D’. Any event passing one of the single photon triggers was recorded. The datasets used for this measurement are summarized in Table 5.1, where the AOD refers to “Analysis Object Data”, which is one of the reconstructed data formats at CMS. AOD optimizes the event information based on the size and complexity of the event information. All of the information required for this measurement can be extracted from the AOD data tier. Table 5.1 also lists double muon and single electron datasets used in the uncertainty estimations and the efficiency measurements. The double muon dataset consists of events with at least two muons in the final state passing one of the double muon triggers, while the single electron dataset consists of events with at least one electron in the final state that passes one of the single electron triggers.

Table 5.1: Datasets used in the photon+jet triple differential cross section measurement.

Dataset
Single Photon
/Photon/Run2012A-22Jan2013-v1/AOD
/SinglePhoton/Run2012B-22Jan2013-v1/AOD
/SinglePhoton/Run2012C-22Jan2013-v1/AOD
/SinglePhotonParked/Run2012D-22Jan2013-v1/AOD
Double Muon
/DoubleMu/Run2012A-13Jul2012-v1/AOD
/DoubleMu/Run2012A-recover-06Aug2012-v1/AOD
/DoubleMu/Run2012B-13Jul2012-v3/AOD
/DoubleMu/Run2012C-24Aug2012-v1/AOD
/DoubleMu/Run2012C-PromptReco-v2/AOD
/DoubleMu/Run2012D-PromptReco-v1/AOD
Double Electron
/SingleElectron/Run2012A-22Jan2013-v1/AOD
/SingleElectron/Run2012B-22Jan2013-v1/AOD
/SingleElectron/Run2012C-22Jan2013-v1/AOD
/SingleElectron/Run2012D-22Jan2013-v1/AOD

All of the events a datasets are not necessarily of good quality. If the sub-detector conditions were not optimal during a specific run, those runs are excluded by using text files known as JSON files. A ‘golden JSON file’: /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions12/8TeV-/Reprocessing/Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt is used for this measurement to ensure that only those runs and luminosity sections that are characterized as “Good” are included.

5.2 Triggers

5.2.1 Triggers Used on Single Photon Dataset

This analysis uses high-level trigger (HLT) paths with thresholds applied on online transverse momentum (p_T). Some triggers also have an additional threshold on a shower shape quantity i.e. Calorimeter-only ID (CaloId). This threshold is associated with $\sigma_{i\eta i\eta} < 0.024$ in the barrel and less than 0.040 in the endcaps. The shower shape cut applied online at trigger level is much looser than the shower shape cut applied at selection (refer to Section 5.4.1); hence, the difference between the

Table 5.2: HLT trigger paths used in the photon+jet triple differential cross section measurement along with the L1 triggers, the effective integrated luminosity, the offline p_T^γ range and the prescale values corresponding to the paths.

HLT trigger path	L1 trigger	$\int L(t)dt$ [pb ⁻¹]	p_T^γ range [GeV]	Prescale
HLT_Photon30_CaloIdVL_v*	L1_SingleEG20 or L1_SingleEG22	2.87	40–65	7000
HLT_Photon50_CaloIdVL_v*	L1_SingleEG20 or L1_SingleEG22	22.35	65–90	900
HLT_Photon75_CaloIdVL_v*	L1_SingleEG20 or L1_SingleEG22	134.11	90–105	150
HLT_Photon90_CaloIdVL_v*	L1_SingleEG20 or L1_SingleEG22	280.73	105–165	80
HLT_Photon135_v*	L1_SingleEG30	13771	165–180	1
HLT_Photon150_v*	L1_SingleEG30	19712	> 180	1

triggers in shower shape is not accounted for in the analysis. The effective luminosity for some of the HLT paths is different than the recorded luminosity because of prescaling. A prescale value of x reduces the number of events recorded by a factor of x . Triggers with lower p_T thresholds come with higher prescale. The caveat to using only the lower p_T threshold triggers is that the single photon production cross section is expected to fall steeply as p_T^γ rises; therefore, at large p_T^γ values statistics will be very limited. Therefore, in order to maximize the number of events without compromising on the phase space, datasets passing different triggers are used in different ranges of photon p_T . The ranges of p_T^γ in which the particular triggers are used are also presented in Table 5.2. The choice is made by requiring maximally efficient triggers in particular p_T^γ ranges.

As shown in Table 5.2, the first four HLT paths share a common L1 trigger and so do the last two. With p_T thresholds of 20 or 22 GeV for first four and 30 GeV for the last two paths, the L1 seed thresholds that are considered in the analysis are much larger than the HLT path threshold. Therefore, for the purpose of checking where the HLT paths are maximally efficient, it is sufficient

to measure the relative efficiency of the higher p_T threshold trigger with respect to the lower p_T threshold trigger. The relative efficiency is measured using a bootstrapping or ratio technique. In this technique, the efficiency for a higher online p_T threshold trigger (HLT_PhotonX_*) can be measured by seeing how many of the events that pass the lower p_T threshold trigger (HLT_PhotonY_*) also pass the HLT_PhotonX_* trigger. The efficiency for HLT_PhotonX_* is, therefore, the ratio of the events that pass both the triggers to the events that only pass HLT_PhotonY_*. Mathematically,

$$\epsilon_{HLT_PhotonX_*} = \frac{N_{(HLT_PhotonX_* \text{ and } HLT_PhotonY_*)}}{N_{HLT_PhotonY_*}}. \quad (5.2)$$

For the first four and the last two single photon triggers listed in Table 5.2, the only difference between the two HLT paths are their online p_T thresholds. Hence, the events that pass the test trigger (the trigger whose efficiency is being measured) can be obtained by checking which events passing the reference trigger (the trigger which is being used to measure the efficiency of the other trigger) have online p_T of the HLT object greater than the threshold of the test trigger. Mathematically,

$$\epsilon_{HLT_PhotonX_*} = \frac{N_{(HLT_PhotonY_* \text{ and } Online \ p_T > X)}}{N_{HLT_PhotonY_*}}. \quad (5.3)$$

Though their paths differ by an additional criteria on the shower shape, this method can also be extended to evaluate efficiency for HLT_Photon135_v* relative to HLT_Photon90_CaloIdVL_v* because of a tighter offline cut on shower shape as explained earlier. In the calculation of efficiency using this method, the HLT_PhotonY_* trigger is always chosen such that its p_T^γ threshold is right below that of the HLT_PhotonX trigger. This is to maximize the data in the numerator since the much higher prescaling of a lower p_T^γ threshold results in a large reduction of the statistics.

As shown in Figures 5.1 and 5.2, the triggers studied are found to be maximally efficient after approximately 15% of their p_T thresholds. To evaluate these turn-on curves, the photon selection cuts presented in Section 5.4.1 are applied. Though the bootstrapping method can be implemented to measure the trigger efficiency relative to the reference trigger, the absolute efficiency (including the efficiency of the L1 trigger) for the reference trigger still needs to be evaluated using a separate method. This value will contribute to the total efficiency measurement for this analysis will be discussed in detail in Section 5.6.

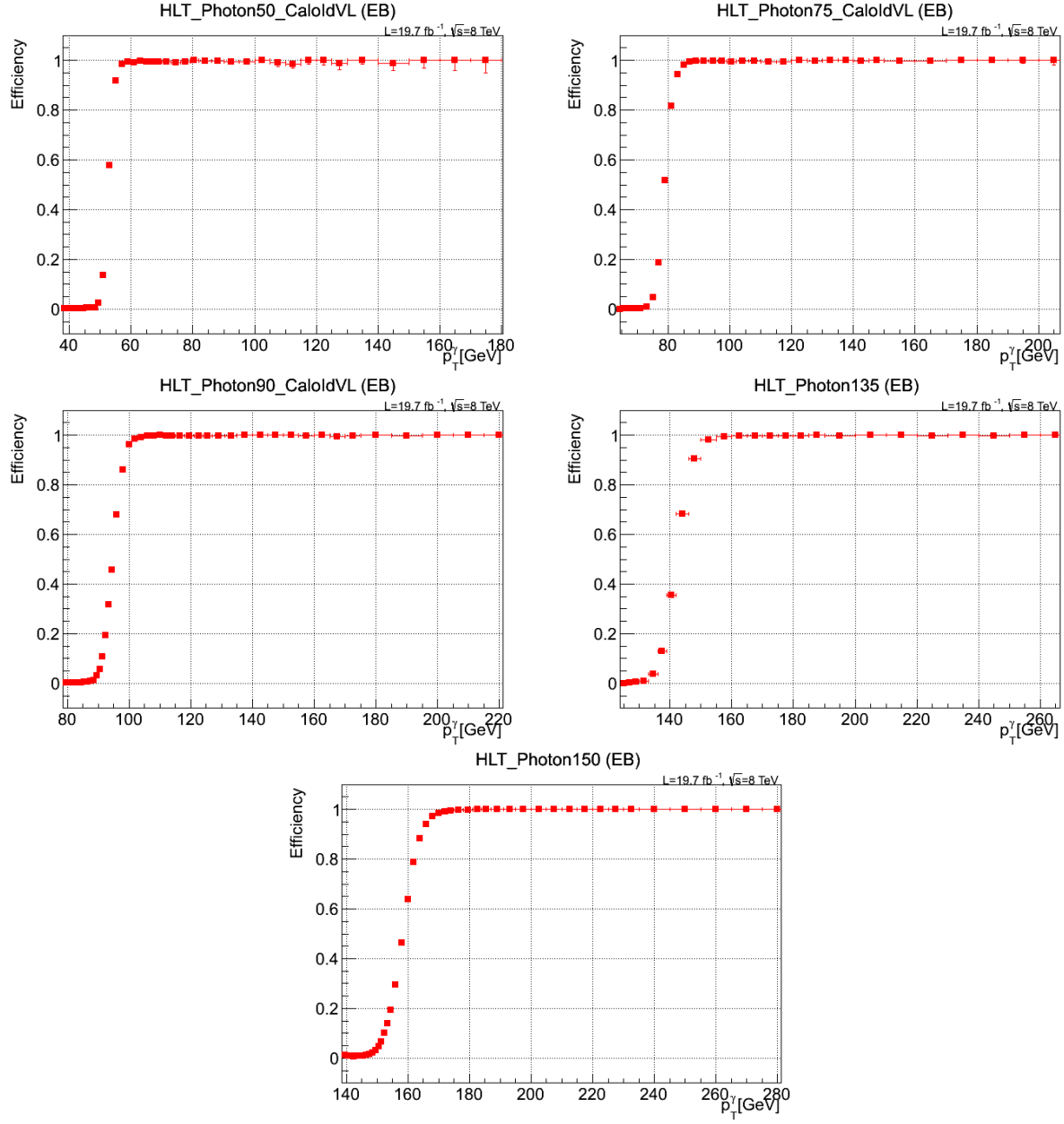


Figure 5.1: HLT efficiency curves for paths HLT_Photon50_CaloldVL_v*, HLT_Photon75_CaloldVL_v*, HLT_Photon90_CaloldVL_v*, HLT_Photon135_v*, and HLT_Photon150_v* (top to bottom and left to right) in barrel as a function of the offline p_T^γ .

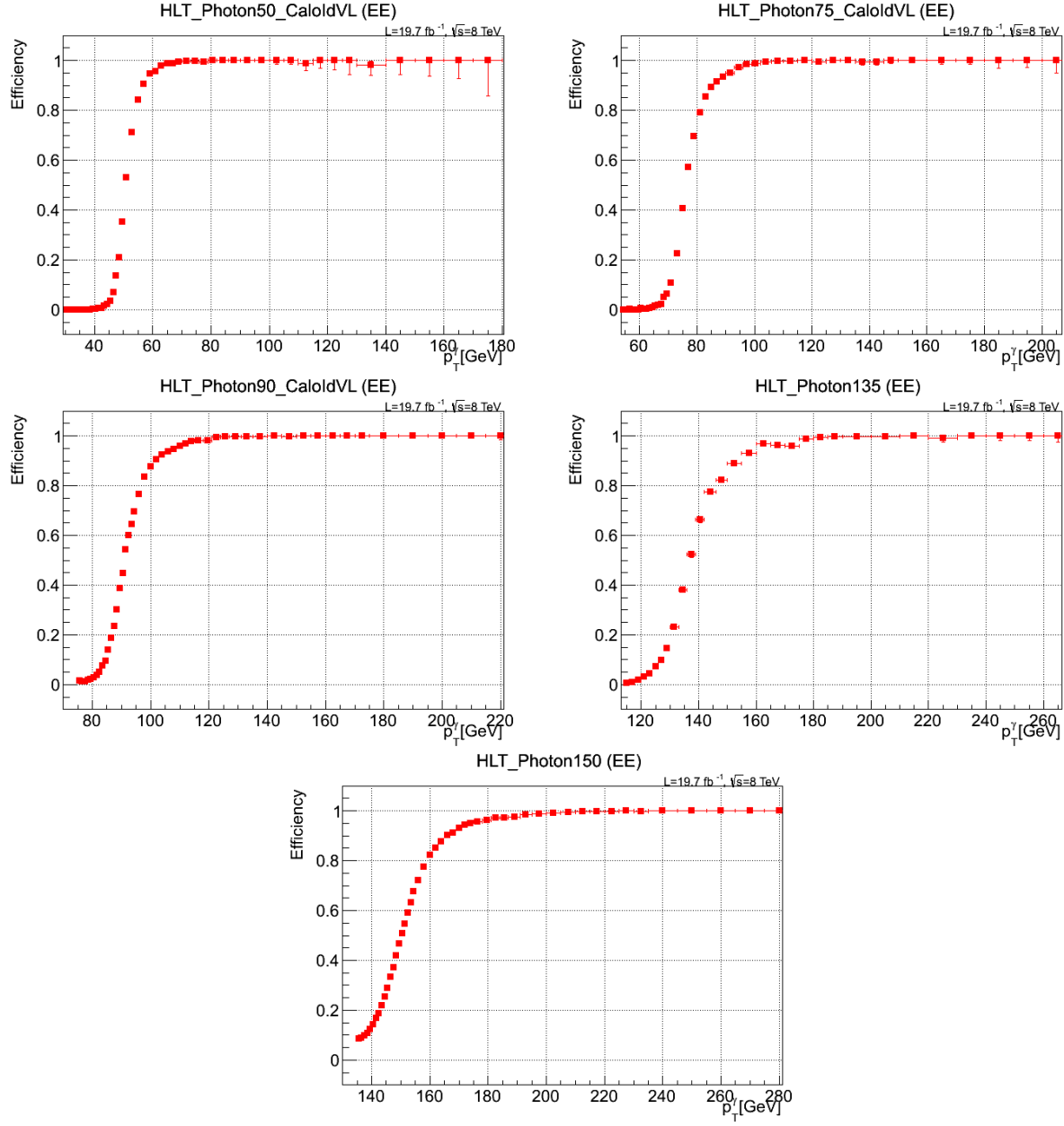


Figure 5.2: HLT efficiency curves for paths HLT_Photon50.CaloldVL_v*, HLT_Photon75.CaloldVL_v*, HLT_Photon90.CaloldVL_v*, HLT_Photon135_v*, and HLT_Photon150_v* (top to bottom and left to right) in endcaps as a function of the offline p_T^γ .

5.2.2 Triggers Used on Single Electron Dataset

The single electron events that are used for efficiency measurements are required to pass a HLT path: `HLT_Ele27_WP80_*`. Here, Ele27 refers to the p_T threshold of 27 GeV and WP80 refers to an approximate efficiency of 80% for online electron identification. The working point of 80% is associated with cuts on $\sigma_{i\eta i\eta}$, H/E , isolation measured with the ECAL in $\Delta R=0.3$ divided by electron E_T , isolation measured with the HCAL in a $\Delta R=0.3$ divided by electron E_T , tracker isolation in $\Delta R=0.3$ divided by electron E_T , η width, ϕ width, $(\frac{1}{E_e} - \frac{1}{p_e})$, and the number of hits on the electron track. The aforementioned detector isolation quantities define the amount of ECAL/HCAL/tracker energy present in a cone around the electron candidate. The L1 seed associated with the single electron HLT path is `L1_SingleEG20` or `L1_SingleEG22`, the same L1 seeds as for the majority of the single photon triggers used in this analysis. Since this single electron trigger is not prescaled, it can be used for efficiency measurements.

5.3 Monte Carlo Samples

The purpose of Monte Carlo (MC) event generators is to predict the outcome of proton-proton collisions using random number generators in which combinations of many analytical results and theory models are provided as inputs. Some of the most important and common event properties that are of relevance in a MC generation are a description of the hard interaction, parton distributions, initial/final state radiation, and hadronization. The $\gamma + \text{jet}$ measurement done with simulations using various MC generators at many stages of the analysis. $\gamma + \text{jet}$ MC events are used to construct signal templates. QCD MC events are used during optimization of the data sideband from which background templates are constructed. Similarly, a MC sample with the radiative decay of Z bosons to two leptons is used to estimate the systematic uncertainty associated with the signal template construction. A separate Drell-Yan ($Z/\gamma^* \rightarrow \ell\bar{\ell} + \text{jets}$) sample is also used in the study of systematics.

The QCD and photon+jet samples used in this measurement are generated with PYTHIA6 [32] for a center-of-mass energy of 8 TeV. Both samples use CTEQ6L [33] PDFs and a Z2* [34] tune for the underlying events. The Z boson radiative decay to leptons is simulated using Madgraph[35].

Table 5.3: MC samples used in the photon+jet measurement for signal template construction, background sideband optimization, and the study of systematics.

Process	Monte Carlo	σ [pb]
$\gamma + \text{jet}$	G_Pt-0to15_TuneZ2star_8TeV_pythia6	90601456
	G_Pt-15to30_TuneZ2star_8TeV_pythia6	200061.7
	G_Pt-30to50_TuneZ2star_8TeV_pythia6	19931.62
	G_Pt-50to80_TuneZ2star_8TeV_pythia6	3322.309
	G_Pt-80to120_TuneZ2star_8TeV_pythia6	558.287
	G_Pt-120to170_TuneZ2star_8TeV_pythia6	108.007
	G_Pt-170to300_TuneZ2star_8TeV_pythia6	30.122
	G_Pt-300to470_TuneZ2star_8TeV_pythia6	2.1
	G_Pt-470to800_TuneZ2star_8TeV_pythia6	0.212
	G_Pt-800to1400_TuneZ2star_8TeV_pythia6	0.007
QCD	QCD_Pt_20_30_EMEnriched_TuneZ2star_8TeV_pythia6	2914860
	QCD_Pt_30_80_EMEnriched_TuneZ2star_8TeV_pythia6	4615893
	QCD_Pt_80_170_EMEnriched_TuneZ2star_8TeV_pythia6	183294.9
	QCD_Pt_170_250_EMEnriched_TuneZ2star_8TeV_pythia6	4586.52
	QCD_Pt_250_350_EMEnriched_TuneZ2star_8TeV_pythia6	556.75
	QCD_Pt_350_EMEnriched_TuneZ2star_8TeV_pythia6	89.1
Systematics	ZGToLLG_8TeV-madgraph	130
	DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball	3504

The Drell-Yan sample is generated using Madgraph interfaced to PYTHIA . It is also simulated with the Z2* tune and CTEQ6L PDF set. All of the aforementioned MC samples are centrally produced by the CMS collaboration as a part of the Summer 2012 campaign. The list of samples along with their associated cross sections is provided in Table 5.3.

Pile-up effects from pile-up scenario S10 are added to all the samples used in this analysis. The pile-up interaction distribution roughly covers, without necessarily matching, the condition expected for 2012 runs. There will be discrepancies between the data and MC originating mainly from the differences in underlying event in data and MC as well as the primary vertex reconstruction procedure. In order to account for these differences, the generator level MC pile-up distribution is reweighted by the reconstructed pile-up distribution from data.

5.4 Event Selection

After the triggered data are selected, a set of offline cuts is applied on the objects reconstructed as described in Chapter 4. The main purpose of the selection is to reduce the contribution from background to the candidate sample. The same cuts are applied to the MC in order to match the trigger acceptance as well as the geometric acceptance of the detector. In this case, acceptance refers to the region of phase space that can be probed. The two main physics objects for this analysis are photons and jets. Each event is required to have one photon and at least one jet. Both photon and jet candidates are required to pass selection criteria whose details are provided below.

5.4.1 Photon Selection

In this analysis, events are required to have at least one high-energy photon passing selection cuts. In case there is more than one photon passing the criteria, the highest energy photon is used. Descriptions of the preselection variables are listed below along with the cut values:

- The transverse component of the photon candidate's reconstructed energy (p_T^γ) is required to be greater than 30 GeV. This cut is applied to all the samples. The lowest p_T^γ bin starts at a value of 40 GeV, which is well above this preselection cut.
- The pseudorapidity of the photon candidate ($|\eta^\gamma|$) is calculated from the primary vertex. A cut of $0 < |\eta^\gamma| < 1.4442$ or $1.566 < |\eta^\gamma| < 2.5$ is required for the candidate sample. This cut ensures that the photon candidate is reconstructed within the ECAL fiducial region, where barrel section stretch from $0 < |\eta^\gamma| < 1.4442$ and the endcap covers $1.566 < |\eta^\gamma| < 2.5$.
- Conversation-safe electron veto is required.
- The PU corrected charged hadron isolation within a cone of radius $\Delta R = 0.3$ is required to be less than 5 GeV.
- The $\sigma_{i\eta i\eta}$ quantity is required to be < 0.014 for barrel photons and < 0.034 for endcap photons.
- For photons in the barrel region, H/E is required to be less than 0.082 if $R_9 > 0.94$, and it is required to be less than 0.075 if $R_9 < 0.94$. In the case of endcap photons, H/E is required to be less than 0.075 for all values of R_9 .
- Though no cut is placed in the R_9 variable during preselection, the cut on H/E depends on the R_9 values.

Table 5.4: Summary of the preselection cuts.

Variable	Cut (barrel)	Cut (endcaps)
p_T^γ	$> 30 \text{ GeV}$	
$ \eta^\gamma $	< 1.4442	$> 1.566 \text{ and } < 2.5$
CSEV	$=1$	
CH Isolation ($\Delta R = 0.3$)	$< 5 \text{ GeV}$	
$\sigma_{i\eta i\eta}$	< 0.014	< 0.034
H/E	$< 0.082 \text{ if } R_9 > 0.94$ $< 0.075 \text{ if } R_9 < 0.94$	< 0.075

The selection cuts are summarized in Table 5.4.

5.4.2 Generator Level Photon Selection

Simulated events can be reconstructed just as data is, and the preselection cuts that are applied to the data can be applied to the MC as well. However, one of the major advantages of MC over data is that the true information is available for MC. In certain cases, it is advantageous to use the truth level (generator level) MC information to match with the reconstructed MC sample. In such situations, the reconstructed signal events from MC are required to match the following generator level criteria for isolated prompt photons.

- $\text{pdgID} = 22$: pdgID is a Particle Data Group Identifier that is different for different particles. Since the value of the ID is 22 for photons, applying this condition on the generator level particle will select only photons.
- $\text{statusMC} = 3 \text{ OR } \text{statusMC} = 1 \text{ and } |\text{mother pdgID}| < 10$: $\text{StatusMC} = 3$ corresponds to the partons that are used in the matrix element calculation, including immediate decays of resonances. $\text{StatusMC} = 1$ corresponds to particles that are not fragmented or decayed and are the final state particles given by the generator. $\text{mother } |\text{pdgID}| < 10$ corresponds to all quarks and gluons. Therefore, the OR selection criteria results in the selection of particles from the hard interaction or final state particles whose mother is a quark or a gluon.
- Particle level Isolation ($\Delta R = 0.2$) less than 4 GeV. Here, the particle level isolation is calculated by summing the transverse momenta of all generator level particles excluding the photon itself in the cone. This criterion, though implemented to match the signal acceptance to the background acceptance, removes most of the fragmentation photons.

The matching between a reconstructed photon candidate and the generator level signal photon is done using a cone of $\Delta R = 0.1$. Objects selected by matching reconstructed photon candidates

with generator level particles that fail at least one of the above generator level signal criteria are called fakes.

5.4.3 Jet Selection

In this study, candidate events are also required to have at least one particle flow (PF) jet that passes selection criteria applied to reject fake jets, badly reconstructed jets and noise while keeping the efficiency for selecting real jets high. Details of the PF jet reconstruction are provided in Chapter 4. The anti- k_t jet algorithm with distance parameter $\Delta R = 0.5$ is used in the reconstruction of the jets used for this measurement. Since in the particle flow jet reconstruction, the algorithms are fed with PF objects, photons are also reconstructed as jets. These photons are removed by requiring that the jet does not overlap with the leading photon within a radius of $\Delta R = 0.5$. After removal of the fakes, the jets are required to pass the set of conditions summarized in Table 5.5. A more detailed description of each of the variables and the cuts is provided below:

- p_T^{jet} is the transverse momentum of the jet. It is required to be greater than 25 GeV.
- Neutral Hadron Fraction is required to be less than 0.90.
- The neutral electromagnetic fraction is required to be less than 0.90.
- The jet must have at least two constituents.
- The fraction of jet energy coming from muons is required to be less than 0.8.

The following cuts related to the charged hadrons are applied only for jets within the fiducial region of the tracker with $|\eta^{jet}| < 2.4$.

- The fraction of jet energy in HCAL coming from charged objects is required to be greater than zero.
- The number of PF charged hadron objects in the jet is required to be greater than zero as well. This cut is dependent in the previous cut on the charged hadron fraction.
- The fraction of jet energy in the ECAL coming from charged objects is required to be less than 0.90.

Table 5.5: Summary of the jet selection cuts.

Variable	Cut (barrel)
p_T^{jet}	$> 25 \text{ GeV}$
Neutral Hadron Fraction	< 0.90
Neutral EM Fraction	< 0.90
Number of Constituents	> 1
Muon Fraction	< 0.8
If $ \eta^{jet} < 2.4$:	
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged ElectroMagnetic Fraction	< 0.90

Furthermore, jets coming from pile-up are rejected using the jet ID algorithm [36], which uses discriminating variables based on the jet topology, object multiplicity, and trajectories of the tracks associated with the jets. Once the above cuts are applied, the jet energy is corrected as described in Chapter 4 using the prescription provided by CMS [31]. The triple differential cross section is carried out in four η^{jet} bins: $|\eta^{jet}| < 0.8$, $0.8 < |\eta^{jet}| < 1.5$, $1.5 < |\eta^{jet}| < 2.1$ and $2.1 < |\eta^{jet}| < 2.5$.

The number of candidate events passing the event selection cuts along with additional cuts on photon isolation and MVA response as described in Section 5.5 are provided in Tables 5.6, 5.7, 5.8, and 5.9.

5.5 Purity Estimate

Purity is the fraction of signal photons present in the candidate sample. The pool of data after the selection criteria have been applied merely forms the candidate sample. The cross section for the background, comprised mostly of QCD events, is many orders of magnitude larger than the cross section for signal process. Therefore, even after applying the photon and event selection cuts, a significant amount of background remains. To obtain the purity from the candidate sample, a template fitting technique is applied. In this technique, a variable is chosen such that the shape

Table 5.6: Number of candidate events along with statistical uncertainty for different p_T^γ and $|\eta^{jet}|$ bins at $|\eta^\gamma| < 0.8$.

p_T^γ [GeV]	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–45	1552 ± 39.40	1128 ± 33.59	630 ± 25.10	268 ± 16.37
45–65	2375 ± 48.73	1552 ± 39.40	905 ± 30.08	388 ± 19.70
65–72	2181 ± 46.70	1506 ± 38.81	809 ± 28.44	299 ± 17.29
72–90	2809 ± 53.00	1826 ± 42.73	972 ± 31.18	355 ± 18.84
90–97	3216 ± 56.71	2114 ± 45.98	1010 ± 31.78	406 ± 20.15
97–105	2573 ± 50.72	1685 ± 41.05	869 ± 29.48	322 ± 17.94
105–117	5394 ± 73.44	3398 ± 58.29	1666 ± 40.82	626 ± 25.02
117–165	7987 ± 89.37	4817 ± 69.40	2316 ± 48.12	786 ± 28.04
165–172	21560 ± 146.83	12766 ± 112.99	5773 ± 75.98	1896 ± 43.54
172–180	20499 ± 143.17	11907 ± 109.12	5302 ± 72.81	1681 ± 41.00
180–188	23582 ± 153.56	14025 ± 118.43	6059 ± 77.84	1814 ± 42.59
188–196	19445 ± 139.45	11110 ± 105.40	4827 ± 69.48	1484 ± 38.52
196–204	15797 ± 125.69	9271 ± 96.29	3844 ± 62.00	1143 ± 33.81
204–215	17229 ± 131.26	9903 ± 99.51	4172 ± 64.59	1171 ± 34.22
215–230	17792 ± 133.39	10013 ± 100.06	4094 ± 63.98	1113 ± 33.36
230–245	12773 ± 113.02	7173 ± 84.69	2906 ± 53.91	758 ± 27.53
245–260	9336 ± 96.62	5148 ± 71.75	2132 ± 46.17	516 ± 22.72
260–285	10642 ± 103.16	5876 ± 76.66	2159 ± 46.47	524 ± 22.89
285–335	11079 ± 105.26	6062 ± 77.86	2216 ± 47.07	444 ± 21.07
335–1000	10525 ± 102.59	5111 ± 71.49	1531 ± 39.13	215 ± 14.66

Table 5.7: Number of candidate events along with statistical uncertainty for different p_T^γ and $|\eta^{jet}|$ bins at $0.8 < |\eta^\gamma| < 1.44$.

p_T^γ [GeV]	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–45	1300 $\pm$ 36.06	1033 $\pm$ 32.14	587 $\pm$ 24.23	246 $\pm$ 15.68
45–65	2004 $\pm$ 44.77	1387 $\pm$ 37.24	882 $\pm$ 29.70	363 $\pm$ 19.05
65–72	1793 $\pm$ 42.34	1287 $\pm$ 35.87	765 $\pm$ 27.66	326 $\pm$ 18.06
72–90	2308 $\pm$ 48.04	1638 $\pm$ 40.47	954 $\pm$ 30.89	394 $\pm$ 19.85
90–97	2601 $\pm$ 51.00	1862 $\pm$ 43.15	1028 $\pm$ 32.06	416 $\pm$ 20.40
97–105	2140 $\pm$ 46.26	1471 $\pm$ 38.35	870 $\pm$ 29.50	340 $\pm$ 18.44
105–117	4258 $\pm$ 65.25	3114 $\pm$ 55.80	1703 $\pm$ 41.27	652 $\pm$ 25.53
117–165	6308 $\pm$ 79.42	4365 $\pm$ 66.07	2456 $\pm$ 49.56	824 $\pm$ 28.71
165–172	16345 $\pm$ 127.85	11588 $\pm$ 107.65	6165 $\pm$ 78.52	2051 $\pm$ 45.29
172–180	15400 $\pm$ 124.10	10554 $\pm$ 102.73	5509 $\pm$ 74.22	1845 $\pm$ 42.95
180–188	17745 $\pm$ 133.21	12460 $\pm$ 111.62	6355 $\pm$ 79.72	2017 $\pm$ 44.91
188–196	14175 $\pm$ 119.06	10100 $\pm$ 100.50	5056 $\pm$ 71.11	1658 $\pm$ 40.72
196–204	11811 $\pm$ 108.68	8183 $\pm$ 90.46	4178 $\pm$ 64.64	1348 $\pm$ 36.72
204–215	12669 $\pm$ 112.56	8912 $\pm$ 94.40	4443 $\pm$ 66.66	1318 $\pm$ 36.30
215–230	12893 $\pm$ 113.55	9044 $\pm$ 95.10	4279 $\pm$ 65.41	1303 $\pm$ 36.10
230–245	9293 $\pm$ 96.40	6328 $\pm$ 79.55	3005 $\pm$ 54.82	842 $\pm$ 29.02
245–260	6978 $\pm$ 83.53	4604 $\pm$ 67.85	2128 $\pm$ 46.13	562 $\pm$ 23.71
260–285	7785 $\pm$ 88.23	5251 $\pm$ 72.46	2228 $\pm$ 47.20	592 $\pm$ 24.33
285–335	8180 $\pm$ 90.44	5290 $\pm$ 72.73	2227 $\pm$ 47.19	472 $\pm$ 21.73
335–1000	7385 $\pm$ 85.94	4368 $\pm$ 66.09	1434 $\pm$ 37.87	192 $\pm$ 13.86

Table 5.8: Number of candidate events along with statistical uncertainty for different p_T^γ and $|\eta^{jet}|$ bins at $1.56 < |\eta^\gamma| < 2.1$.

p_T^γ [GeV]	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–65	2492 $\pm$ 49.92	1935 $\pm$ 43.99	1240 $\pm$ 35.21	541 $\pm$ 23.26
65–90	2902 $\pm$ 53.87	2293 $\pm$ 47.89	1603 $\pm$ 40.04	739 $\pm$ 27.18
90–105	3287 $\pm$ 57.33	2680 $\pm$ 51.77	1775 $\pm$ 42.13	873 $\pm$ 29.55
105–165	6886 $\pm$ 82.98	5738 $\pm$ 75.75	3898 $\pm$ 62.43	1576 $\pm$ 39.70
165–180	20619 $\pm$ 143.59	17103 $\pm$ 130.78	10664 $\pm$ 103.27	4034 $\pm$ 63.51
180–200	23782 $\pm$ 154.21	19657 $\pm$ 140.20	12365 $\pm$ 111.20	4383 $\pm$ 66.20
200–225	16547 $\pm$ 128.64	13468 $\pm$ 116.05	8144 $\pm$ 90.24	2584 $\pm$ 50.83
225–1000	22243 $\pm$ 149.14	17054 $\pm$ 130.59	9067 $\pm$ 95.22	2315 $\pm$ 48.11

Table 5.9: Number of candidate events along with statistical uncertainty for different p_T^γ and $|\eta^{jet}|$ bins at $2.1 < |\eta^\gamma| < 2.5$.

p_T^γ [GeV]	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–65	1726 ± 41.55	1424 ± 37.74	954 ± 30.89	494 ± 22.23
65–90	2023 ± 44.98	1763 ± 41.99	1281 ± 35.79	567 ± 23.81
90–105	2152 ± 46.39	1870 ± 43.24	1363 ± 36.92	690 ± 26.27
105–165	4562 ± 67.54	3879 ± 62.28	2809 ± 53.00	1317 ± 36.29
165–180	11607 ± 107.74	9922 ± 99.61	6741 ± 82.10	2538 ± 50.38
180–200	12717 ± 112.77	10893 ± 104.37	7160 ± 84.62	2585 ± 50.84
200–225	8334 ± 91.29	6929 ± 83.24	4469 ± 66.85	1441 ± 37.96
225–1000	8763 ± 93.61	6904 ± 83.09	3624 ± 60.20	947 ± 30.77

distributions of signal and background events are different. The distinction in the background and signal shape allows one to fit the candidate sample with the sum of signal and background templates. The signal fraction that is associated with the fit gives the signal purity.

A composite model $F(x)$ can be constructed from the unity-normalized signal template $S(x)$ and unity-normalized background template $B(x)$ as

$$F(x) = f_{sig} \cdot S(x) + (1 - f_{sig}) \cdot B(x). \quad (5.4)$$

Here, f_{sig} represents the purity. Since both $S(x)$ and $B(x)$ are unity normalized, $F(x)$ is also unity normalized by construction. A fit is performed to maximize the likelihood, which is equivalent to minimizing the negative of the log likelihood defined as

$$-\log L(f_{sig}; x_1, x_2, \dots, x_N) = -\sum_N \log F(x_i | f_{sig}). \quad (5.5)$$

In the above equation, $L(f_{sig}; x_1, x_2, \dots, x_N)$ is the likelihood function as a function of parameter f_{sig} , the x_i represent the individual observed values, N represents the total number of data points. The maximum likelihood estimates the parameter f_{sig} that maximizes the probability for each independent measurement x_i given the composite model $F(x)$.

Before beginning the purity estimate, a fit distribution is obtained. The weights obtained from the MVA training of the EM enriched $\gamma + \text{jet}$ sample described in Section 4.3.2 is applied to the data

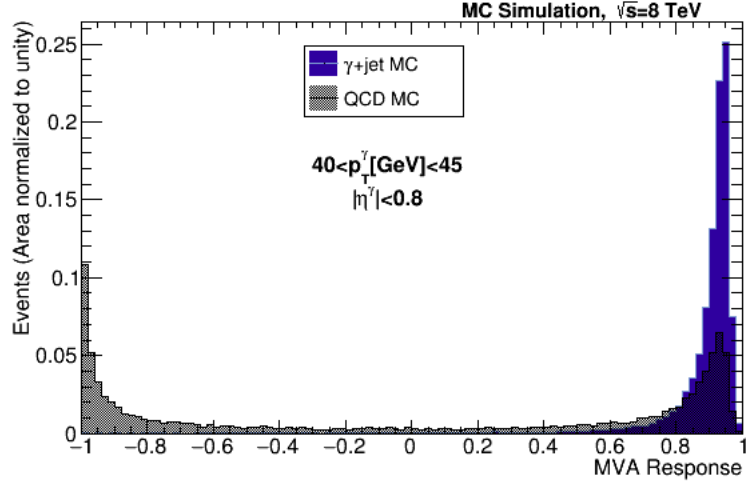


Figure 5.3: An example of the MVA response distribution for truth matched signal and background samples from MC simulation. Both signal and background areas are normalized to unity.

and MC to get MVA response distributions for all the samples used in the fitting. Since the MVA response distributions are distinct for signal and fake events obtained from the truth matched MC as shown in Figure 5.3, the MVA response distributions are ideal for performing a template fit. The signal templates are obtained directly from the MC signal histogram. The background templates are obtained from single photon data sidebands using data-driven techniques. The optimization of the data sidebands for constructing background templates is discussed in the next section.

5.5.1 Background Template Construction

The process of obtaining background templates from data sidebands makes an assumption that the fit variable is independent of the sideband variable. In this measurement, the photon isolation variable defined in Chapter 4 is used as a sideband variable. The choice is driven by the fact that no cuts are applied to the photon isolation during preselection and it is one of the few variables that is not used in the training of the MVA. Though it is expected to be mostly uncorrelated with the MVA response, slight correlations are observed. Measures are taken to 1) reduce the correlation between MVA response and the photon isolation variable and 2) optimize a sideband region where

the bias due to correlation is a minimum.

1. **Reducing the correlation between MVA response and photon isolation:** As described in the Section 4.3.1, the photon isolation estimate involves summing up the energy associated with the PF photon candidate in a cone around the photon candidate. The photon candidate, though being a similar name, is a reconstructed supercluster (SC) with the standalone photon reconstruction. During the photon isolation estimation, some of the SC energy may leak into the isolation cone, which can cause a correlation between the photon isolation and MVA response variable. The standard photon isolation algorithm used in this analysis performs a cleaning of this leakage using a method known as supercluster footprint removal. It is done using an algorithm that extrapolates the PF photons to the surface of the ECAL crystals in the SC, and removing them from the isolation sum if they are within a certain distance of the crystal boundaries measured in the units of crystal width. The distance parameter (also called enlargement parameter) can be adjusted such that better footprint removal is performed.

A study of whether or not MVA response and photon isolation are correlated can be performed by studying if the photon isolation distribution for background events is similar at different slices of MVA response values. In the absence of any bias, these distributions are expected to be similar to each other, while the reverse is an indication of correlation between the variables used in this analysis. QCD multijet events that pass all the event selection criteria and are matched to the generator level condition for fake photons are selected for the bias study. The study is performed for each bin in p_T^γ and $|\eta^\gamma|$. For each bin, unity-normalized photon isolation distributions are compared for two different cut values of MVA: MVA response > 0.8 and MVA response < -0.8 .

Comparison shows that when using a default value of the footprint removal enlargement factor, there is a considerable amount of correlation between the MVA response variable and the photon isolation. As the area of footprint removal is increased by tuning the enlargement parameter, the differences between the two distributions become smaller. When no further improvement is made, the process is stopped. Figures 5.4 and 5.5 show examples of the comparison between photon isolation distributions at the two different MVA response cut values using the default footprint removal enlargement setting, and similar distributions when the footprint area is tuned to minimize the bias. More plots for different combinations of p_T^γ , $|\eta^\gamma|$, and $|\eta^{jet}|$ bins are provided in Appendix A. Based on these studies, the enlargement parameter is chosen as 7.0 for $|\eta^\gamma| < 1.5$, 5.0 for $1.5 < |\eta^\gamma| < 2.1$, and 3.0 for $2.1 < |\eta^\gamma| < 2.5$.

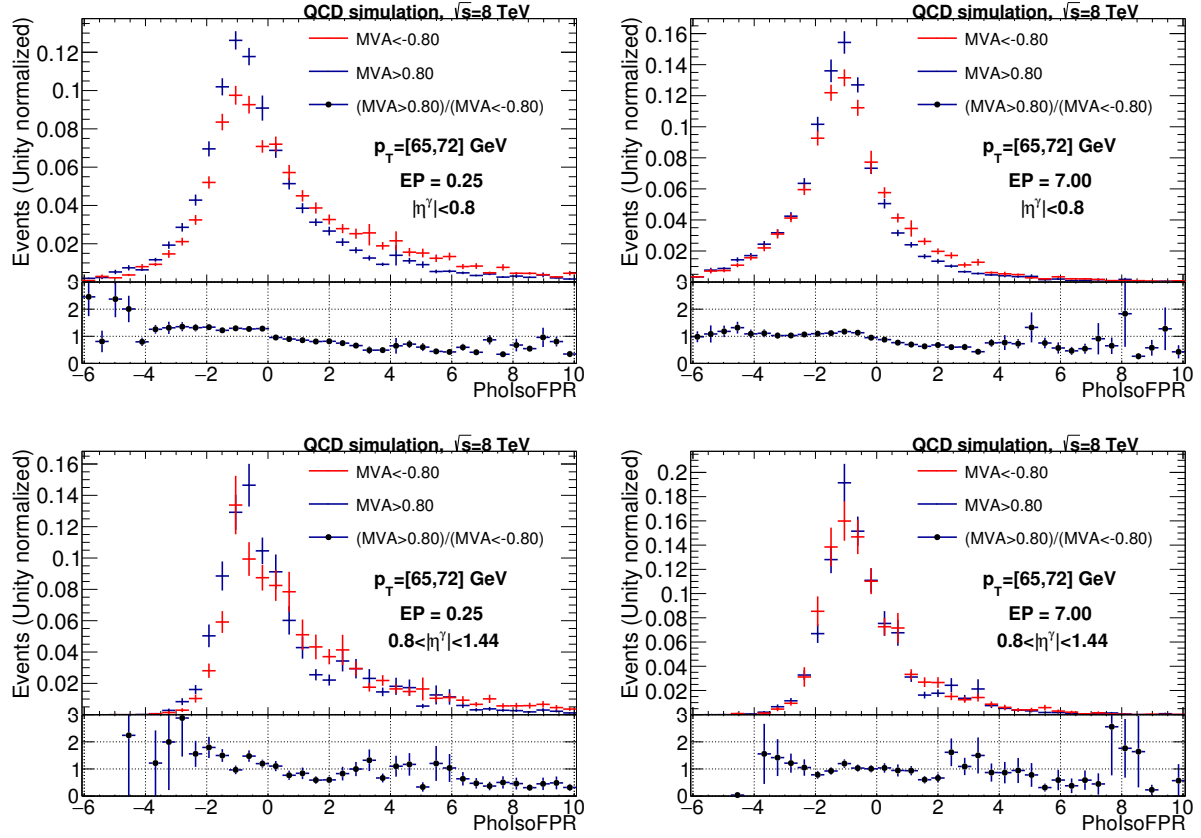


Figure 5.4: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD fakes in the barrel region using MC simulation.

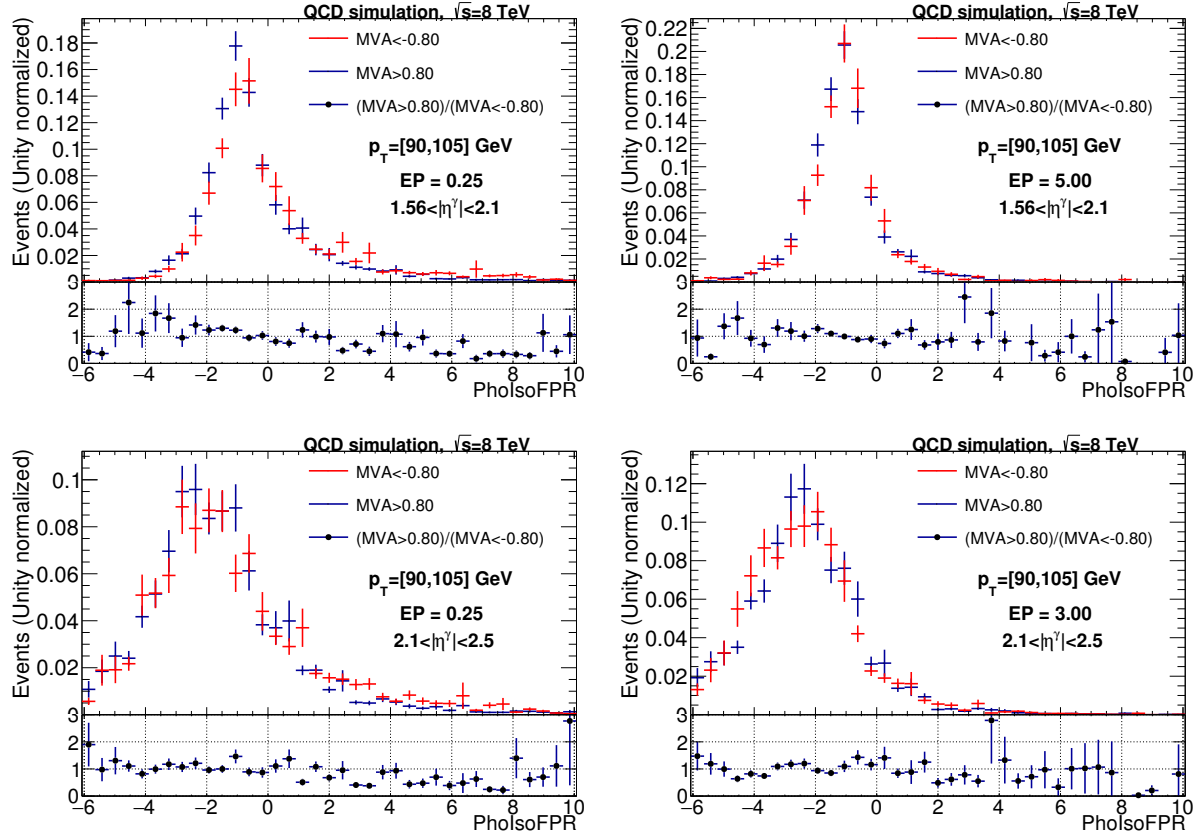


Figure 5.5: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD fakes in the endcaps region using MC simulation.

2. **Sideband Optimization:** Though the correlation between MVA response and the photon isolation variable is reduced by changing the amount of footprint removal, a residual correlation is expected to be present. For each bin of p_T^γ and $|\eta^\gamma|$, sideband regions are optimized individually so as to select those regions of photon isolation that give the minimum residual correlation for that particular bin. The goal is to find a cut window in the photon isolation variable such that the data MVA response distribution in that region adequately describes the fake MVA response distribution present in the data in the nominal region. Here, the nominal region is defined as a region that is orthogonal to the sideband region. The nominal region is selected differently for the different $|\eta^\gamma|$ bins based on the distribution of the photon isolation variable and where the signal photons are dominant. The nominal region are defined as photon isolation less than 0 GeV, -0.5 GeV and -1.0 GeV for $|\eta^\gamma| < 1.5$, $1.5 < |\eta^\gamma| < 2.1$, and $2.1 < |\eta^\gamma| < 2.5$, respectively.

The sideband optimization is performed by creating toy experiments for individual bins in the different combinations of p_T^γ and $|\eta^\gamma|$. To that end, for each bin, pseudo data is generated using the sum of photon MC events and fake MC events. Photon objects in the $\gamma + \text{jet}$ sample that pass the event selection criteria are matched to the MC particles that pass the generator level signal criteria provided in Section 5.4.2 to get true photon MC events. Similarly, the photon objects in the QCD MC sample passing the same selection criteria but matching to generator level particles that fail at least one of the signal criteria are used to get the fake MC events. In order to generate the pseudo events, a certain signal fraction is arbitrarily chosen keeping in mind certain expectations based on known physics. In data, the signal fraction is expected to increase as a function of p_T^γ since the QCD background contribution is reduced at higher transverse momenta. Therefore, the pseudo data is generated with an input fraction of 0.4 for $p_T < 90$ GeV, 0.6 for $p_T < 180$ and an input fraction of 0.8 for $p_T > 180$ GeV for the barrel. Also, the purity is expected to be lower in higher rapidity regions, therefore, it is chosen differently for endcaps (EE). The values chosen are 0.3 for $p_T < 90$ GeV, 0.45 for $p_T < 180$ and 0.6 for $p_T > 180$ GeV in EE. With the signal fraction f_{in} and templates from true signal events $S(x)$ and true fake events $B_1(x)$, pseudo data is generated from the composite template $G(x)$;

$$G(x) = f_{in} \cdot S(x) + (1 - f_{in}) \cdot B_1(x). \quad (5.6)$$

The number of generated events is kept the same as the number of candidate events in a corresponding bin in data to keep the uncertainty and bias close to what is expected from data.

In order to fit the pseudo data, several tentative templates ($B_t(x)$) are constructed. The templates are constructed from an admixture of QCD MC and $\gamma + \text{jet}$ MC. This sample,

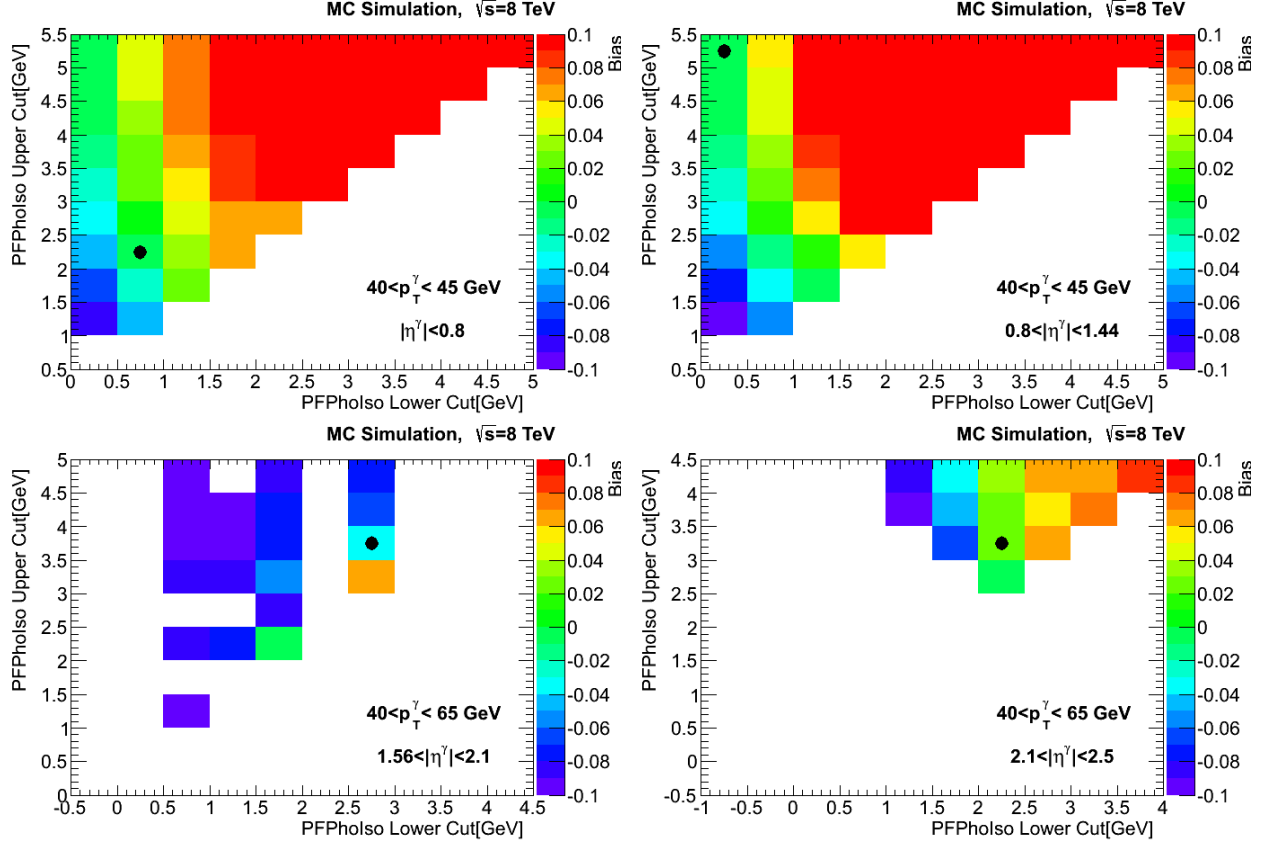


Figure 5.6: Sideband variation plots. The z axis shows bias values obtained for MC toy experiments done with tentative background templates with the sideband window corresponding to the photon isolation values represented by x and y axes. Black dots represent the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin are used to construct the background template.

from here on referred to as Mixture-1, is made by adding the cross section reweighted MC events that pass the photon selection criteria. The events are treated like data in that no generator level information is used. To construct the tentative templates, different sets of photon isolation cuts are applied to Mixture-1. Forty-five tentative templates per bin are considered for this process, which differ from one another only in the isolation cut. The cuts are all orthogonal to the nominal region defined above.

Each of the 45 tentative templates along with the signal template are fit to the pseudodataset. Since simply performing a likelihood fit does not give information about the number of events and only fits the shape, an extended maximum likelihood fit is performed. In other words, the measured number of events is a random number generated from a Poisson distribution

about the number of generated events. For each template, the fit is repeated 500 times with a different number of events and a fixed signal fraction. As a result, for each of the tentative templates, a distribution of output signal fraction (f_{out}) is obtained. This accounts for the variation seen in f_{out} due to the statistical uncertainty in the number of events.

Similarly, bias and pull distributions are obtained where bias (b) is defined as

$$b = (f_{out} - f_{in})/f_{in}, \quad (5.7)$$

and pull (p) as

$$p = (f_{out} - f_{in})/\sigma_{f_{out}}, \quad (5.8)$$

where, $\sigma_{f_{out}}$ is the fitting uncertainty associated with each of the f_{out} values. Based on the MVA output distribution of signal events, the signal events are expected to accumulate at a MVA output value close to +1. Though the signal purity is expected to be higher at higher MVA response values, fitting in the entire range of the MVA response allows for a better signal/background discrimination because the shape difference between the fake and true events at negative MVA response values is found to be more striking, and acts as an handle to control the shape of the composite model being fit to the dataset. Hence, it is important to first perform the fitting over the entire range of MVA response and apply a cut on MVA response value afterwards. A working point is chosen at a MVA response value of 0.8. Therefore, the output signal fraction, bias, and pull calculations are confined within the MVA output range of 0.8 to 1.0.

Bias values close to zero imply that the output signal fraction is close to the input signal fraction. Hence, the photon isolation cut applied to the corresponding template $B_t(x)$ gives the best agreement between the MVA response distribution in Mixture-1 and the similar distribution of fakes from QCD MC in the nominal region. This window can thus be applied to the data to get a background template to estimate the background events present in the candidate sample. Results from some toy experiments with bias as a function of the upper and lower values of photon isolation cuts are shown in Figure 5.6. More plots are provided in Appendix B. Each bin in these figures corresponds to a toy experiment performed with a particular choice of tentative template obtained from indicated photon isolation cuts. The black dot corresponds to the bin that resulted in a bias value closest to zero. The photon isolation cuts that result in the lowest bias for each p_T^γ and $|\eta^\gamma|$ bins are summarized in Table 5.10. With these cut values, a background region is found where the contamination of the signal photons is minimal, and the background template shape agrees with

Table 5.10: Photon isolation cuts applied to data for background templates in each bin of p_T^γ and $|\eta^\gamma|$.

p_T^γ (GeV)	Photon isolation (GeV)		p_T^γ (GeV)	Photon isolation (GeV)	
	$ \eta^\gamma < 0.8$	$0.8 < \eta^\gamma < 1.5$		$1.5 < \eta^\gamma < 2.1$	$2.1 < \eta^\gamma < 2.5$
40–45	0.5–2.0	0.0–5.0	40–65	2.5–3.5	2.5–3.5
45–65	0.5–2.5	0.5–2.0	65–90	0.5–4.5	1.5–3.5
65–72	1.0–3.0	0.0–5.0	90–105	3.0–4.5	1.5–3.5
72–90	1.0–4.0	0.5–3.5	105–165	3.0–4.5	2.0–3.0
90–97	1.5–5.0	0.5–5.0	165–180	1.5–4.5	3.0–4.5
97–105	1.0–5.0	0.5–5.0	180–200	3.0–4.0	2.0–4.5
105–117	1.5–4.5	1.0–5.0	200–225	2.0–4.0	2.0–3.5
117–165	2.0–3.5	1.0–2.5	225–1000	1.5–4.5	2.0–4.5
165–172	1.5–3.5	2.0–3.5			
172–180	2.0–5.0	1.0–4.5			
180–188	2.0–5.0	1.0–3.0			
188–196	1.5–2.5	1.5–3.5			
196–204	1.5–4.5	1.5–2.5			
204–215	2.0–4.0	3.0–4.5			
215–230	4.0–5.0	1.5–2.5			
230–245	4.0–5.0	1.0–3.0			
245–260	2.5–5.0	1.0–4.5			
260–285	3.0–4.0	1.0–3.0			
285–335	2.0–4.0	1.0–1.5			
335–1000	1.5–2.5	0.0–5.0			

the template shape from MC truth.

5.5.2 Fitting Procedure

Once the sideband is optimized, the MVA response distribution of the data in the sideband region is taken as a non-parametric background template. This background along with the $\gamma +$ jet signal template is used to fit the candidate sample using a binned maximum likelihood fit that minimizes the negative log likelihood equation provided in Equation 5.5. A few examples of such fits are provided in Figures 5.7 and 5.8. Plots for all combinations of p_T^γ , $|\eta^\gamma|$ and $|\eta^{jets}|$ are provided in Appendix C. The raw purity values are listed in Tables 5.11, 5.12, 5.13, and 5.14 along with the fitting uncertainties. These values are further corrected for any possible residual bias arising from the correlation between the MVA and the photon isolation variable. The associated systematic uncertainties will be provided in Section 5.5.3 along with a discussion of their origin.

Table 5.11: Uncorrected values of purity and the associated fitting uncertainty for different p_T^γ and $|\eta^{jet}|$ bins at $|\eta^\gamma| < 0.8$.

p_T^γ [GeV]	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–45	$0.553 \pm 3.07\%$	$0.515 \pm 4.09\%$	$0.397 \pm 8.03\%$	$0.347 \pm 14.44\%$
45–65	$0.657 \pm 1.78\%$	$0.539 \pm 3.19\%$	$0.498 \pm 4.82\%$	$0.445 \pm 8.97\%$
65–72	$0.722 \pm 1.51\%$	$0.694 \pm 2.01\%$	$0.589 \pm 3.85\%$	$0.530 \pm 7.65\%$
72–90	$0.770 \pm 1.14\%$	$0.725 \pm 1.66\%$	$0.675 \pm 2.76\%$	$0.540 \pm 6.93\%$
90–97	$0.829 \pm 0.78\%$	$0.797 \pm 1.10\%$	$0.695 \pm 2.34\%$	$0.701 \pm 3.64\%$
97–105	$0.800 \pm 1.09\%$	$0.756 \pm 1.59\%$	$0.697 \pm 2.73\%$	$0.613 \pm 5.98\%$
105–117	$0.835 \pm 0.63\%$	$0.803 \pm 0.92\%$	$0.739 \pm 1.68\%$	$0.707 \pm 3.10\%$
117–165	$0.862 \pm 0.46\%$	$0.827 \pm 0.70\%$	$0.781 \pm 1.23\%$	$0.713 \pm 2.66\%$
165–172	$0.844 \pm 0.35\%$	$0.809 \pm 0.52\%$	$0.753 \pm 0.96\%$	$0.732 \pm 1.81\%$
172–180	$0.887 \pm 0.25\%$	$0.859 \pm 0.39\%$	$0.831 \pm 0.66\%$	$0.774 \pm 1.48\%$
180–188	$0.883 \pm 0.25\%$	$0.858 \pm 0.37\%$	$0.820 \pm 0.67\%$	$0.770 \pm 1.50\%$
188–196	$0.839 \pm 0.41\%$	$0.797 \pm 0.64\%$	$0.724 \pm 1.27\%$	$0.693 \pm 2.54\%$
196–204	$0.873 \pm 0.35\%$	$0.851 \pm 0.50\%$	$0.798 \pm 0.97\%$	$0.777 \pm 1.91\%$
204–215	$0.877 \pm 0.33\%$	$0.853 \pm 0.49\%$	$0.816 \pm 0.89\%$	$0.768 \pm 2.02\%$
215–230	$0.930 \pm 0.18\%$	$0.913 \pm 0.28\%$	$0.881 \pm 0.53\%$	$0.867 \pm 1.09\%$
230–245	$0.917 \pm 0.26\%$	$0.903 \pm 0.39\%$	$0.876 \pm 0.71\%$	$0.869 \pm 1.43\%$
245–260	$0.923 \pm 0.30\%$	$0.905 \pm 0.47\%$	$0.884 \pm 0.82\%$	$0.852 \pm 1.91\%$
260–285	$0.904 \pm 0.36\%$	$0.887 \pm 0.54\%$	$0.857 \pm 1.04\%$	$0.806 \pm 2.55\%$
285–335	$0.893 \pm 0.43\%$	$0.866 \pm 0.67\%$	$0.843 \pm 1.23\%$	$0.840 \pm 2.83\%$
335–1000	$0.852 \pm 0.74\%$	$0.810 \pm 1.27\%$	$0.756 \pm 2.73\%$	$0.825 \pm 5.92\%$

In order to correct for the bias for a particular choice of a sideband, a relationship between input signal fraction and output signal fraction is established using yet another set of MC toy experiments. The procedure for the toy experiments is similar to that used in the sideband optimization. The MC samples used are also the same as what was used before, namely, γ +jet MC and QCD MC with truth matching for pseudo data generation and γ +jet MC and Mixture-1 in the optimized sideband region for constructing templates. In the sideband optimization, a constant value of input signal fraction (f_{in}) is fit multiple times with different hypothesis for the measured number of events to get a distribution of output signal fraction (f_{out}). However, in this toy experiment, f_{in} is varied for a constant value of measured number of events resulting in a range of output signal fractions f_{out} , establishing a relationship between f_{in} and f_{out} . The points along the curve describing the f_{in} vs. f_{out} relationship give the correction for purity values obtained by fitting data. Examples showing

Table 5.12: Uncorrected values of purity and the associated fitting uncertainty for different p_T^γ and $|\eta^{jet}|$ bins at $0.8 < |\eta^\gamma| < 1.44$.

p_T^γ [GeV]	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–45	$0.469 \pm 4.51\%$	$0.528 \pm 4.22\%$	$0.457 \pm 7.07\%$	$0.399 \pm 13.47\%$
45–65	$0.537 \pm 3.00\%$	$0.501 \pm 4.06\%$	$0.540 \pm 4.47\%$	$0.512 \pm 7.63\%$
65–72	$0.551 \pm 3.18\%$	$0.558 \pm 3.70\%$	$0.555 \pm 4.89\%$	$0.568 \pm 7.10\%$
72–90	$0.638 \pm 2.13\%$	$0.606 \pm 2.79\%$	$0.646 \pm 3.26\%$	$0.555 \pm 6.80\%$
90–97	$0.696 \pm 1.63\%$	$0.724 \pm 1.73\%$	$0.675 \pm 2.73\%$	$0.658 \pm 4.56\%$
97–105	$0.714 \pm 1.71\%$	$0.726 \pm 1.96\%$	$0.696 \pm 2.82\%$	$0.640 \pm 5.55\%$
105–117	$0.740 \pm 1.08\%$	$0.768 \pm 1.14\%$	$0.737 \pm 1.72\%$	$0.712 \pm 3.06\%$
117–165	$0.753 \pm 0.91\%$	$0.761 \pm 1.06\%$	$0.746 \pm 1.50\%$	$0.729 \pm 2.74\%$
165–172	$0.820 \pm 0.40\%$	$0.843 \pm 0.43\%$	$0.833 \pm 0.62\%$	$0.831 \pm 1.08\%$
172–180	$0.792 \pm 0.51\%$	$0.806 \pm 0.58\%$	$0.815 \pm 0.77\%$	$0.787 \pm 1.49\%$
180–188	$0.771 \pm 0.54\%$	$0.800 \pm 0.58\%$	$0.799 \pm 0.80\%$	$0.784 \pm 1.53\%$
188–196	$0.807 \pm 0.48\%$	$0.839 \pm 0.50\%$	$0.828 \pm 0.74\%$	$0.837 \pm 1.26\%$
196–204	$0.781 \pm 0.63\%$	$0.797 \pm 0.71\%$	$0.812 \pm 0.94\%$	$0.786 \pm 1.83\%$
204–215	$0.861 \pm 0.36\%$	$0.877 \pm 0.39\%$	$0.879 \pm 0.55\%$	$0.871 \pm 1.05\%$
215–230	$0.822 \pm 0.50\%$	$0.836 \pm 0.56\%$	$0.844 \pm 0.78\%$	$0.842 \pm 1.43\%$
230–245	$0.777 \pm 0.80\%$	$0.793 \pm 0.91\%$	$0.797 \pm 1.30\%$	$0.778 \pm 2.67\%$
245–260	$0.833 \pm 0.69\%$	$0.840 \pm 0.83\%$	$0.837 \pm 1.24\%$	$0.820 \pm 2.60\%$
260–285	$0.821 \pm 0.73\%$	$0.833 \pm 0.85\%$	$0.830 \pm 1.30\%$	$0.862 \pm 2.24\%$
285–335	$0.776 \pm 1.05\%$	$0.784 \pm 1.28\%$	$0.823 \pm 1.72\%$	$0.795 \pm 4.24\%$
335–1000	$0.771 \pm 1.38\%$	$0.798 \pm 1.63\%$	$0.738 \pm 3.49\%$	$0.785 \pm 7.79\%$

Table 5.13: Uncorrected values of purity and the associated fitting uncertainty for different p_T^γ and $|\eta^{jet}|$ bins at $1.56 < |\eta^\gamma| < 2.1$.

p_T^γ [GeV]	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–65	$0.338 \pm 4.22\%$	$0.379 \pm 4.22\%$	$0.417 \pm 4.61\%$	$0.377 \pm 7.85\%$
65–90	$0.351 \pm 4.35\%$	$0.404 \pm 4.04\%$	$0.474 \pm 3.81\%$	$0.499 \pm 5.23\%$
90–105	$0.629 \pm 1.45\%$	$0.699 \pm 1.28\%$	$0.735 \pm 1.37\%$	$0.763 \pm 1.79\%$
105–165	$0.551 \pm 1.34\%$	$0.624 \pm 1.19\%$	$0.673 \pm 1.24\%$	$0.666 \pm 2.00\%$
165–180	$0.548 \pm 0.93\%$	$0.635 \pm 0.80\%$	$0.688 \pm 0.86\%$	$0.719 \pm 1.28\%$
180–200	$0.593 \pm 0.78\%$	$0.674 \pm 0.67\%$	$0.746 \pm 0.67\%$	$0.769 \pm 1.05\%$
200–225	$0.629 \pm 0.84\%$	$0.705 \pm 0.74\%$	$0.760 \pm 0.80\%$	$0.787 \pm 1.30\%$
225–1000	$0.622 \pm 0.85\%$	$0.683 \pm 0.81\%$	$0.753 \pm 0.91\%$	$0.742 \pm 1.89\%$

Table 5.14: Uncorrected values of purity and the associated fitting uncertainty for different p_T^γ and $|\eta^{jet}|$ bins at $2.1 < |\eta^\gamma| < 2.5$.

p_T^γ [GeV]	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–65	$0.272 \pm 6.59\%$	$0.334 \pm 5.60\%$	$0.378 \pm 5.81\%$	$0.468 \pm 5.88\%$
65–90	$0.354 \pm 4.06\%$	$0.418 \pm 3.49\%$	$0.507 \pm 3.07\%$	$0.514 \pm 4.42\%$
90–105	$0.402 \pm 3.62\%$	$0.482 \pm 2.98\%$	$0.582 \pm 2.48\%$	$0.658 \pm 2.73\%$
105–165	$0.542 \pm 1.45\%$	$0.606 \pm 1.28\%$	$0.694 \pm 1.11\%$	$0.724 \pm 1.48\%$
165–180	$0.466 \pm 1.31\%$	$0.558 \pm 1.06\%$	$0.655 \pm 0.96\%$	$0.698 \pm 1.36\%$
180–200	$0.504 \pm 1.09\%$	$0.592 \pm 0.90\%$	$0.682 \pm 0.84\%$	$0.742 \pm 1.13\%$
200–225	$0.503 \pm 1.44\%$	$0.596 \pm 1.18\%$	$0.724 \pm 0.97\%$	$0.780 \pm 1.40\%$
225–1000	$0.507 \pm 1.53\%$	$0.615 \pm 1.24\%$	$0.718 \pm 1.23\%$	$0.743 \pm 2.21\%$

the output signal fraction vs. the input signal fraction are provided in Figure 5.9. A second order polynomial is used for fitting. Parameters associated with these fits are provided in Tables 5.15 and 5.16, and all the plots are provided in Appendix D. The final purity values after applying the correction are listed in Tables 5.17 and 5.18.

5.5.3 Systematic Uncertainties in the Purity Measurement

Apart from statistical uncertainties, there are other non-random sources of uncertainty on any measurement. These uncertainties, known as systematic uncertainties, are often difficult to predict and depend on the methodology of the measurement. In this measurement, several sources of systematic uncertainties are investigated: uncertainties due to the background template shape, uncertainties due to the signal template shape, and uncertainties associated with the bias correction. Each of these contributions to the overall systematic uncertainty is described below.

1. **Uncertainties due to the background template shape:** The choice of the background template shape contributes to the total uncertainty on the purity. Since the background template shape is obtained from a data sideband, the technique that is used for optimization of the sideband region introduces uncertainty in the background template. The sideband optimization makes use of Mixture-1 MC in order to predict the background template. Therefore, it is susceptible to acquiring uncertainties due to data-MC disagreement. Furthermore, since the final background template is generated with data, statistics are limited. Random fluctuations caused by the statistical uncertainty of the data in the sideband region may cause mis-modeling of the background template shape.

Table 5.15: The output signal fraction is corrected as $f_{out}^c = p0 + p1 \times f_{out} + p2 \times f_{out}^2$.

p_T^γ [GeV]	$ \eta^\gamma < 0.8$			$0.8 < \eta^\gamma < 1.5$		
	p0	p1	p2	p0	p1	p2
40–45	0.00	0.99	0.00	0.02	0.95	0.03
45–65	−0.01	1.01	0.00	0.03	0.95	0.03
65–72	−0.02	1.05	−0.03	0.04	0.91	0.05
72–90	0.02	0.95	0.03	−0.01	0.98	0.02
90–97	0.00	1.03	−0.03	0.02	0.98	0.00
97–105	0.02	0.96	0.03	0.04	0.92	0.04
105–117	0.07	0.83	0.09	0.02	0.97	0.02
117–165	0.02	0.96	0.02	0.05	0.90	0.04
165–172	0.06	0.85	0.09	0.10	0.77	0.13
172–180	0.02	0.95	0.02	0.07	0.83	0.09
180–188	0.08	0.87	0.04	0.06	0.88	0.06
188–196	0.07	0.85	0.08	0.09	0.82	0.09
196–204	0.06	0.89	0.04	0.05	0.85	0.09
204–215	0.00	0.99	0.01	−0.48	2.00	−0.51
215–230	0.04	0.92	0.04	0.02	0.95	0.02
230–245	−0.01	1.03	−0.02	0.06	0.84	0.10
245–260	0.01	1.02	−0.02	0.02	0.96	0.02
260–285	0.03	0.94	0.03	0.15	0.72	0.13
285–335	0.05	0.90	0.05	0.09	0.82	0.09
335–1000	0.07	0.87	0.06	0.07	0.88	0.05

Table 5.16: The output signal fraction is corrected as $f_{out}^c = p0 + p1 \times f_{out} + p2 \times f_{out}^2$.

p_T^γ [GeV]	$1.5 < \eta^\gamma < 2.1$			$2.1 < \eta^\gamma < 2.5$		
	p0	p1	p2	p0	p1	p2
40–65	0.01	1.10	−0.09	−0.04	1.04	0.00
65–90	−0.03	1.09	−0.06	0.06	0.84	0.09
90–105	−0.05	1.22	−0.17	0.07	0.84	0.09
105–165	0.05	0.88	0.07	−0.05	1.05	0.00
165–180	0.01	1.00	−0.01	−0.01	0.95	0.06
180–200	0.06	0.89	0.05	−0.01	1.03	−0.02
200–225	0.01	0.95	0.03	−0.02	1.04	−0.02
225–1000	0.03	0.93	0.03	0.02	0.93	0.04

Table 5.17: Bias corrected values of purity for different p_T^γ bins in the barrel region. JetBin1, JetBin2, JetBin3 and JetBin4 correspond to $0.8 < |\eta^{jet}| < 1.5$, $1.5 < |\eta^{jet}| < 2.1$, and $2.1 < |\eta^{jet}| < 2.5$ respectively.

p_T^γ [GeV]	$ \eta^\gamma < 0.8$				$0.8 < \eta^\gamma < 1.5$			
	JetBin1	JetBin2	JetBin3	JetBin4	JetBin1	JetBin2	JetBin3	JetBin4
40–45	0.553	0.516	0.399	0.349	0.473	0.531	0.462	0.406
45–65	0.654	0.535	0.493	0.439	0.542	0.507	0.545	0.517
65–72	0.724	0.696	0.591	0.530	0.555	0.561	0.559	0.571
72–90	0.769	0.725	0.675	0.542	0.630	0.598	0.638	0.547
90–97	0.832	0.801	0.701	0.707	0.703	0.730	0.683	0.666
97–105	0.799	0.755	0.696	0.613	0.717	0.729	0.700	0.645
105–117	0.834	0.802	0.739	0.709	0.740	0.768	0.737	0.712
117–165	0.862	0.827	0.781	0.714	0.757	0.765	0.750	0.733
165–172	0.842	0.807	0.752	0.732	0.817	0.839	0.829	0.827
172–180	0.888	0.860	0.832	0.775	0.791	0.805	0.814	0.787
180–188	0.887	0.864	0.828	0.781	0.774	0.803	0.802	0.787
188–196	0.839	0.797	0.726	0.696	0.810	0.841	0.829	0.838
196–204	0.875	0.854	0.803	0.783	0.775	0.791	0.806	0.780
204–215	0.875	0.851	0.814	0.766	0.858	0.876	0.878	0.869
215–230	0.929	0.912	0.880	0.866	0.822	0.837	0.845	0.842
230–245	0.917	0.903	0.876	0.869	0.774	0.789	0.793	0.774
245–260	0.925	0.907	0.887	0.856	0.833	0.840	0.837	0.820
260–285	0.902	0.885	0.855	0.804	0.826	0.838	0.834	0.864
285–335	0.893	0.866	0.844	0.840	0.781	0.789	0.825	0.799
335–1000	0.853	0.812	0.760	0.827	0.776	0.801	0.744	0.789

Table 5.18: Bias corrected values of purity for different p_T^γ bins in the endcaps. JetBin1, JetBin2, JetBin3 and JetBin4 correspond to $0.8 < |\eta^{jet}| < 1.5$, $1.5 < |\eta^{jet}| < 2.1$, and $2.1 < |\eta^{jet}| < 2.5$ respectively.

p_T^γ [GeV]	$1.5 < \eta^\gamma < 2.1$				$2.1 < \eta^\gamma < 2.5$			
	JetBin1	JetBin2	JetBin3	JetBin4	JetBin1	JetBin2	JetBin3	JetBin4
40–65	0.17	0.21	0.26	0.24	0.25	0.26	0.30	0.30
65–90	0.39	0.41	0.45	0.45	0.46	0.44	0.45	0.48
90–105	0.71	0.70	0.74	0.74	0.51	0.53	0.53	0.55
105–165	0.60	0.60	0.65	0.65	0.60	0.59	0.64	0.64
165–180	0.62	0.62	0.64	0.67	0.54	0.55	0.56	0.60
180–200	0.67	0.67	0.70	0.72	0.59	0.59	0.61	0.64
200–225	0.69	0.69	0.69	0.71	0.59	0.60	0.62	0.63
225–1000	0.68	0.66	0.69	0.71	0.59	0.60	0.60	0.60

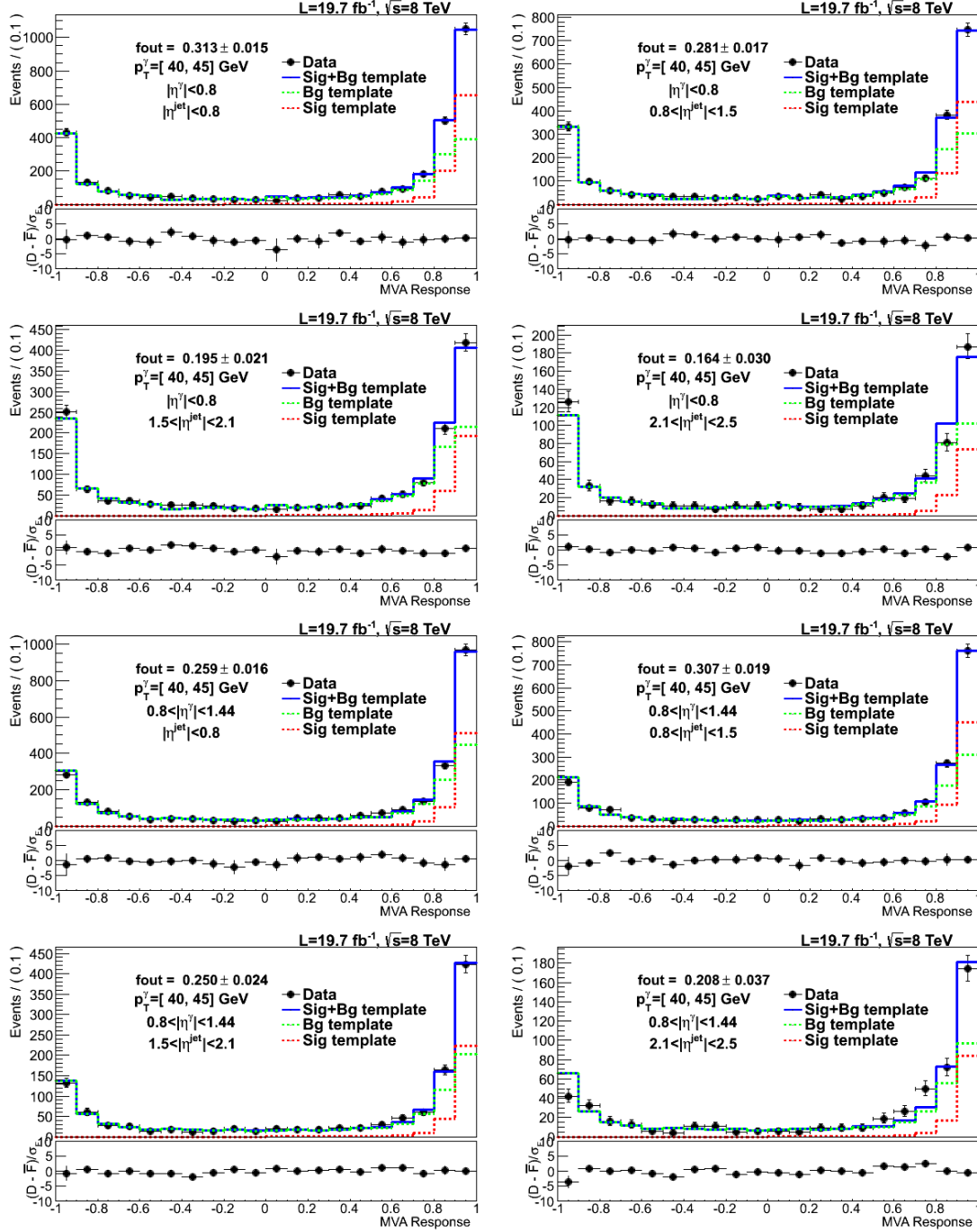


Figure 5.7: The purity fits before the bias correction is applied. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of the purity when fitting the candidate data with Gaussian variations in signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

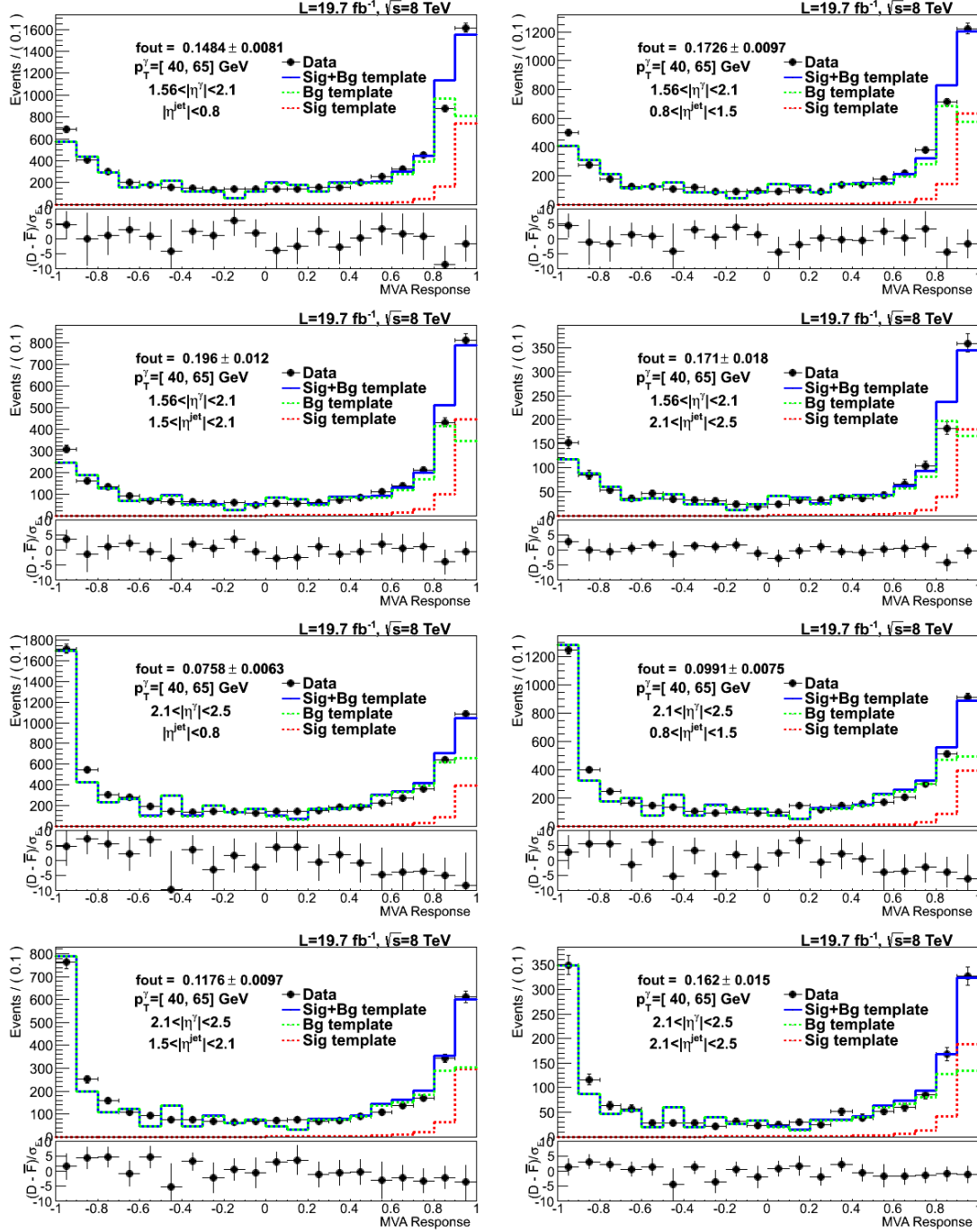


Figure 5.8: The purity fits before the bias correction is applied. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of the purity when fitting the candidate data with Gaussian variations in signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

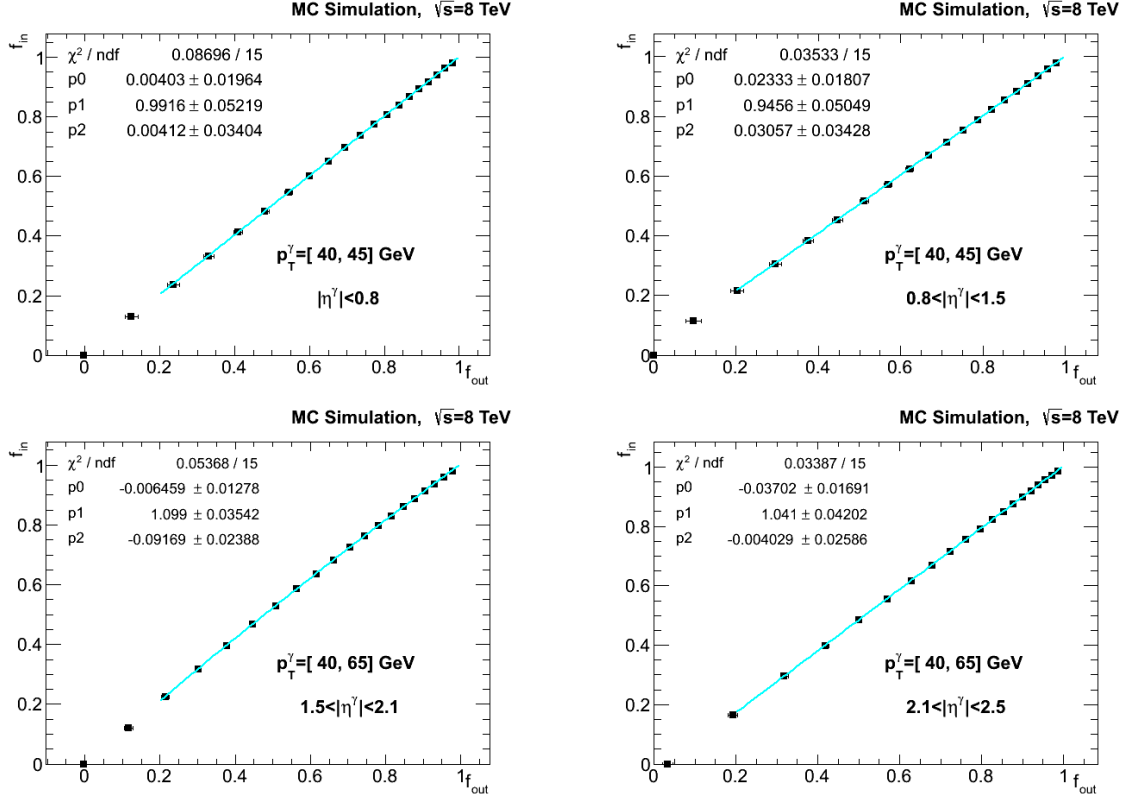


Figure 5.9: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins in $|\eta^\gamma|$.

In order to estimate the systematic uncertainty due to background template, these two contributions to template mis-modeling are considered. Several copies of the nominal background template are made by smearing the nominal template bin by bin. For each bin, the content of the smeared template is taken as a random number generated from a Gaussian distribution about the nominal template. The mean value of the Gaussian is taken as the bin content of the nominal background template. The spread of the Gaussian is taken as the sum in quadrature of the two uncertainties in the background template shape discussed above i.e. statistical uncertainty of the sideband data and data-MC in the sideband region. The smeared background templates along with the nominal signal template are used to fit the data. This process results in a distribution of purity values which envelope the purity measurement made with the nominal background template. From the distribution, the systematic uncertainty is estimated by finding purity values that lie 68% (1σ) above and below the mean purity from the nominal background template fit.

2. **Uncertainties due to the signal template shape:** As in the case of the background template uncertainty, the choice of signal template also contributes to the total systematic uncertainty in the purity measurement. The signal template is constructed using a $\gamma + \text{jet}$ MC sample containing a very large number of events. Hence, the signal template uncertainty is dominated by data-MC disagreement. This data-MC shape difference is estimated using pure photon signals from final state radiation (FSR) events in $Z \rightarrow \gamma\mu\mu$ data and MC. To select the FSR photons from $Z \rightarrow \gamma\mu\mu$ data, events are required to pass one of the single photon triggers discussed previously. Both the data and MC are then required to pass the following selection criteria for muon and event selection along with the standard photon selection criteria presented in Chapter 5.4.1:

- At least two muons with $p_T^\mu > 20(10)$ GeV.
- At least one photon with $p_T^\gamma > 15$ GeV.
- A photon passing a loose H/E cut of 0.05. This cut is tighter than the analogous cut presented in Section 3.1 for events in the endcap.
- $\Delta R(\mu_1, \mu_2) > 1.0$.

Additional cuts are tuned to select final state radiation (FSR) events from the Z boson decay to ensure the selection of a maximally pure photon sample:

- $\min(\Delta R(\gamma, \mu)) < 0.8$
- Two body invariant mass of the two muons, $40 \text{ GeV} < M_{\mu\mu} < 70 \text{ GeV}$.

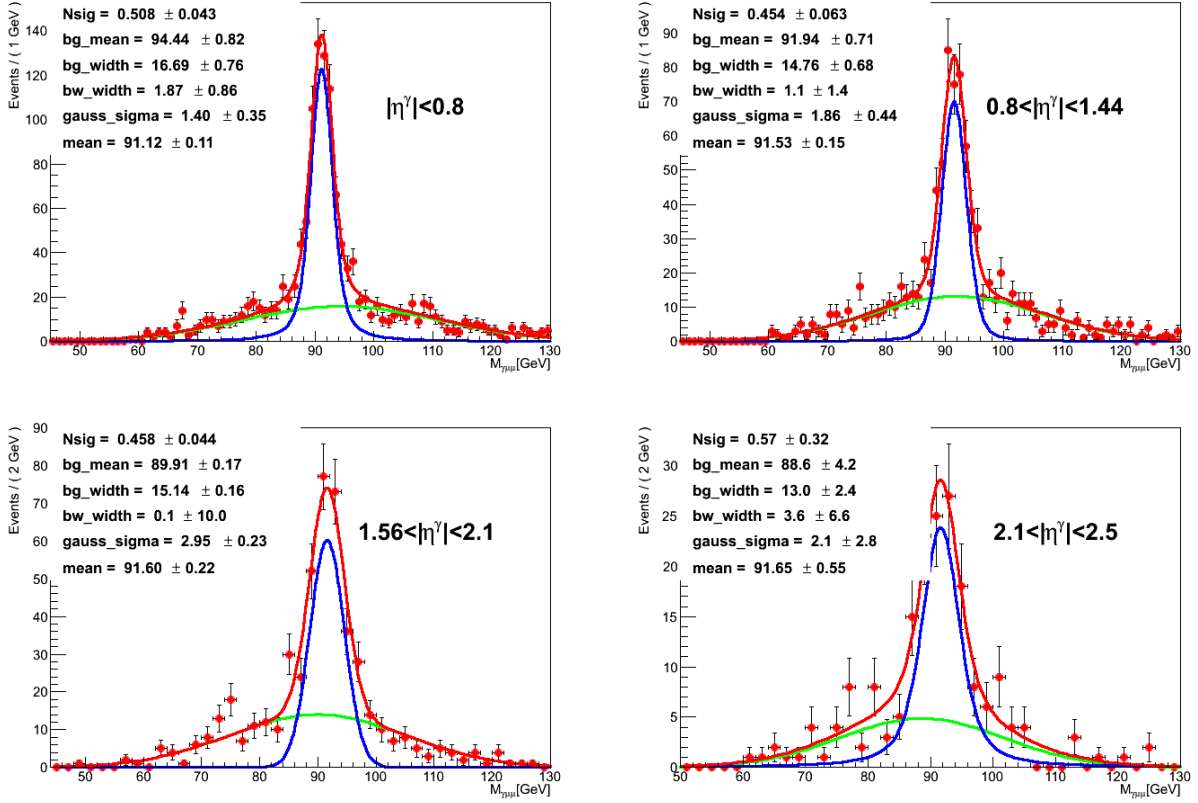


Figure 5.10: Fits of the $Z\gamma$ data (red circles) with the sum of Voigtian function and Gaussian (red line). The blue lines correspond to the Voigtian function describing the signal and green lines correspond to the Gaussian related to the background.

- Three body invariant mass of the two muons and a photon, $80 \text{ GeV} < M_{\mu\mu\gamma} < 100 \text{ GeV}$.

After passing the $Z \rightarrow \gamma\mu\mu$ event selection, there is still a considerable amount of background present in the data. In order to subtract these events from the data, a sideband subtraction technique is applied. The three body invariant mass ($M_{\gamma\mu\mu}$) distribution is fit with a sum of a Voigtian function to describe the signal and a Gaussian to estimate the background present under the Z mass peak as seen in Figure 5.10.

The MVA response distribution of QCD MC is then scaled by the fraction of background present in the signal region. This distribution is then subtracted from the MVA response distribution in a signal region to get a background subtracted distribution. The signal distribution from data is validated using a MC sample. Figure 5.11 shows the background and $Z\gamma$ stacked on top of each other after scaling them to the background and signal fractions

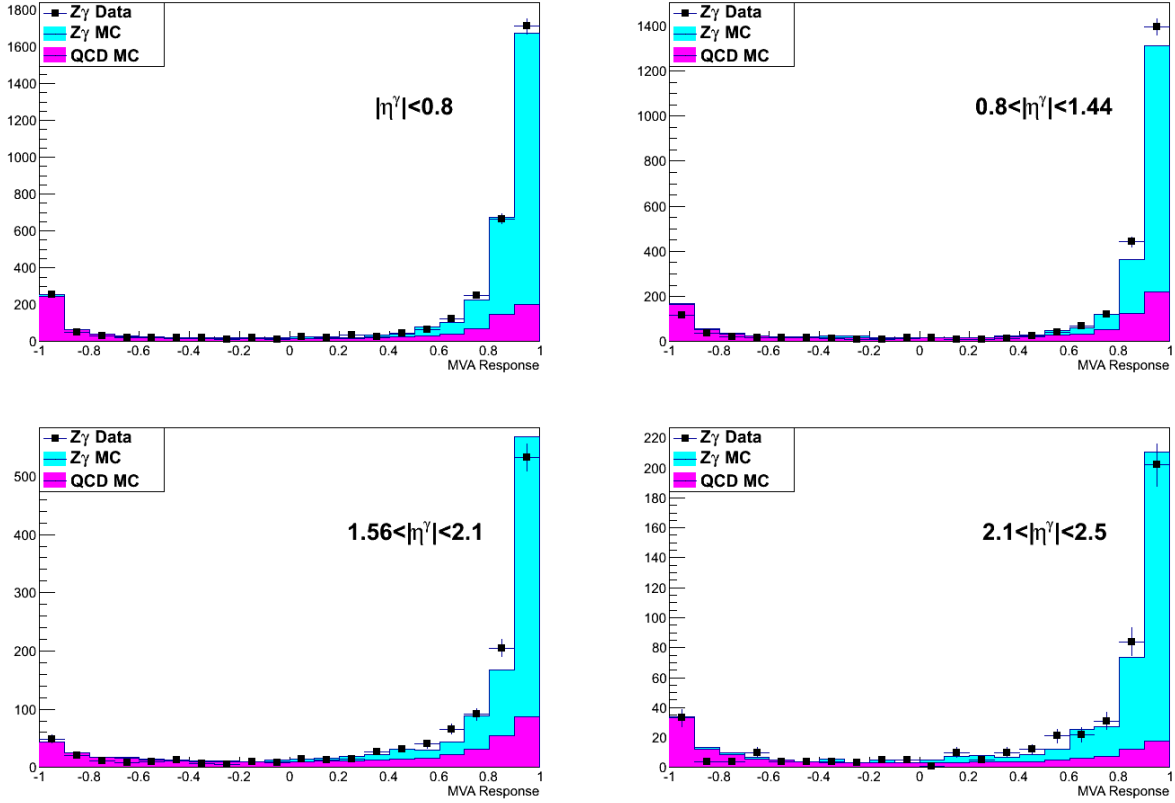


Figure 5.11: The MVA response distributions for QCD background scaled to background fraction from Figure 5.10 in the mass region of $80 < M_{\gamma\mu\mu} < 100$ GeV is shown in cyan, $Z\gamma$ MC FSR distribution scaled to the signal fraction in the same three-body mass region is shown in pink, and the $Z\gamma$ data distribution is represented by the black squares with error bars.

respectively along with the data.

The QCD background is subtracted from $Z\gamma$ data to get a background subtracted $Z\gamma$ FSR distribution from data. The shape difference between this distribution and the distribution from $Z\gamma$ MC is used to assign systematic uncertainty to the γ + jet signal template. There are separate distributions for separate bins in photon eta. However, because of the low statistics of the $Z\gamma$ FSR photon sample the data-MC difference obtained with $p_T^\gamma > 30$ GeV is used for signal templates in all the bins of p_T^γ . There is no observed dependance in the photon MVA distribution obtained from $Z\gamma$ data with the photon p_T .

Once the shape difference is obtained, a similar procedure as in the case of the background

template uncertainty is performed. Smeared signal templates are constructed in a similar way. However, only the data-MC shape difference is considered for smearing since the statistical uncertainty associated with the signal MC is very small. The data events are fit with composite models made with smeared signal templates and nominal background templates. These purity distributions are used to assign the systematic uncertainty on the purity measurement due to signal template choice.

3. **Uncertainties in the bias correction:** Uncertainty associated with the bias correction method presented in Section 5.5.2 is also accounted for in the total uncertainty calculation. In order to estimate the uncertainty, the bias correction can be expressed in terms of outputs of a toy experiment. The pull is defined in terms of the output signal fraction f_{out} and input signal fraction f_{in} as

$$p = (f_{out} - f_{in})/\sigma_{f_{out}}. \quad (5.9)$$

The corrected output signal fraction (f_{out}^c) can thus be expressed in terms of the pull as

$$f_{out}^c = f_{out} - \sigma_{f_{out}} p. \quad (5.10)$$

The uncertainty on the corrected output signal fraction is then

$$df_{out}^c = \sqrt{df_{out}^2 + p^2 d\sigma_{f_{out}}^2 + \sigma_{f_{out}}^2 dp^2}. \quad (5.11)$$

In order to find the correct uncertainty associated with the bias correction, a toy experiment similar to the one for sideband optimization is done. However, since there is an estimate for the corrected purity, this value that is unique for each bin is used as an input to generate the pseudo data. As before, the number of events is allowed to vary within a poisson distribution. As a result, distributions of output signal fraction, uncertainty associated with the signal fraction, and pull are obtained. Using the parameters from these distribution, the relevant uncertainties are estimated.

The total corrected purity values with respect to p_T^γ along with the total uncertainty is provided in Figures 5.12 and 5.13. The breakdown of the uncertainty on the purity into different sources is provided in Tables 5.19 through 5.34.

5.6 Efficiency

One of the most important factors that must be considered in a CMS analysis is the efficiency of the event selection. The total efficiency refers to the ratio of the number of signal events passing all

Table 5.19: Total uncertainties on the purity along with the breakdown of the uncertainties into different categories for $|\eta^\gamma| < 0.8$, $|\eta^{jet}| < 0.8$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total Syst.	
			Low	High	Low	High	Low	High
40–45	3.07%	2.61%	0.78%	1.25%	10.40%	10.07%	10.75%	10.48%
45–65	1.79%	1.75%	0.70%	0.98%	3.20%	6.58%	3.71%	6.88%
65–72	1.51%	1.28%	0.95%	1.26%	5.09%	5.93%	5.33%	6.20%
72–90	1.14%	1.14%	0.70%	1.14%	3.21%	3.50%	3.48%	3.86%
90–97	0.78%	0.86%	0.57%	0.89%	2.67%	4.31%	2.87%	4.49%
97–105	1.09%	0.98%	0.85%	1.17%	1.69%	2.87%	2.13%	3.25%
105–117	0.63%	0.72%	0.78%	1.22%	2.17%	2.25%	2.42%	2.66%
117–165	0.46%	0.50%	0.70%	1.18%	1.32%	1.64%	1.58%	2.08%
165–172	0.35%	0.33%	0.83%	1.38%	1.41%	1.73%	1.67%	2.24%
172–180	0.25%	0.25%	0.53%	1.14%	1.71%	1.81%	1.81%	2.15%
180–188	0.25%	0.25%	0.57%	1.23%	1.28%	2.36%	1.42%	2.67%
188–196	0.41%	0.39%	0.90%	1.78%	1.85%	4.03%	2.09%	4.42%
196–204	0.35%	0.35%	0.61%	1.25%	1.54%	1.92%	1.70%	2.32%
204–215	0.33%	0.27%	0.63%	1.30%	2.28%	2.33%	2.38%	2.68%
215–230	0.18%	0.23%	0.51%	0.94%	1.42%	1.87%	1.53%	2.10%
230–245	0.26%	0.27%	0.63%	1.07%	1.48%	2.19%	1.63%	2.45%
245–260	0.30%	0.32%	0.57%	0.99%	1.16%	1.47%	1.33%	1.80%
260–285	0.36%	0.36%	0.65%	1.14%	1.44%	2.26%	1.62%	2.56%
285–335	0.43%	0.38%	0.64%	1.44%	1.55%	1.82%	1.72%	2.35%
335–1000	0.74%	0.55%	1.39%	2.27%	2.92%	6.27%	3.28%	6.69%

Table 5.20: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $|\eta^\gamma| < 0.8$, $0.8 < |\eta^{jet}| < 1.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total Syst.	
			Low	High	Low	High	Low	High
40–45	4.08%	4.49%	0.82%	1.42%	10.87%	10.51%	11.79%	11.51%
45–65	3.22%	3.07%	0.85%	1.08%	3.70%	7.64%	4.88%	8.31%
65–72	2.00%	2.22%	0.82%	1.30%	5.69%	6.75%	6.16%	7.22%
72–90	1.66%	1.43%	0.65%	1.48%	3.14%	3.73%	3.51%	4.26%
90–97	1.10%	1.13%	0.49%	1.01%	2.69%	4.46%	2.96%	4.71%
97–105	1.60%	1.63%	0.99%	1.24%	1.73%	3.15%	2.57%	3.75%
105–117	0.92%	0.94%	0.83%	1.26%	2.30%	2.38%	2.62%	2.85%
117–165	0.70%	0.67%	0.56%	1.06%	1.49%	1.80%	1.73%	2.19%
165–172	0.53%	0.45%	0.83%	1.42%	1.46%	1.83%	1.74%	2.36%
172–180	0.39%	0.38%	0.62%	1.17%	1.86%	2.03%	2.00%	2.38%
180–188	0.37%	0.35%	0.54%	1.34%	1.38%	2.42%	1.53%	2.79%
188–196	0.64%	0.53%	0.97%	2.08%	1.93%	4.35%	2.22%	4.85%
196–204	0.50%	0.47%	0.71%	1.43%	1.66%	1.98%	1.86%	2.49%
204–215	0.49%	0.44%	0.65%	1.43%	2.46%	2.37%	2.58%	2.80%
215–230	0.28%	0.30%	0.48%	0.98%	1.54%	1.99%	1.64%	2.24%
230–245	0.39%	0.37%	0.77%	1.18%	1.61%	2.26%	1.82%	2.57%
245–260	0.47%	0.42%	0.68%	1.10%	1.25%	1.56%	1.49%	1.95%
260–285	0.54%	0.57%	0.62%	1.32%	1.63%	2.30%	1.83%	2.71%
285–335	0.67%	0.60%	0.62%	1.49%	1.58%	1.91%	1.80%	2.50%
335–1000	1.27%	0.90%	1.97%	2.73%	3.32%	6.14%	3.97%	6.79%

Table 5.21: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $|\eta^\gamma| < 0.8$, $1.5 < |\eta^{jet}| < 2.1$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–45	8.00%	7.53%	0.79%	1.63%	15.08%	14.50%	16.87%	16.42%
45–65	4.87%	4.61%	0.93%	1.11%	4.41%	8.96%	6.45%	10.14%
65–72	3.84%	3.35%	0.89%	1.23%	6.70%	7.39%	7.54%	8.20%
72–90	2.76%	2.71%	1.13%	1.54%	3.26%	4.02%	4.38%	5.09%
90–97	2.32%	2.48%	0.62%	1.05%	3.03%	4.74%	3.97%	5.46%
97–105	2.73%	2.36%	1.00%	1.37%	2.07%	3.81%	3.29%	4.68%
105–117	1.68%	1.80%	1.01%	1.49%	2.50%	2.59%	3.24%	3.49%
117–165	1.23%	1.28%	0.57%	1.18%	1.66%	1.99%	2.17%	2.64%
165–172	0.96%	0.92%	0.85%	1.57%	1.70%	2.16%	2.11%	2.82%
172–180	0.66%	0.70%	0.65%	1.30%	1.83%	1.88%	2.07%	2.39%
180–188	0.66%	0.59%	0.58%	1.42%	1.61%	2.72%	1.81%	3.12%
188–196	1.26%	1.07%	1.03%	2.19%	2.27%	5.18%	2.72%	5.73%
196–204	0.97%	1.00%	0.71%	1.46%	1.77%	2.22%	2.15%	2.84%
204–215	0.89%	0.86%	0.66%	1.57%	2.58%	2.59%	2.80%	3.15%
215–230	0.53%	0.54%	0.42%	1.08%	1.54%	2.04%	1.68%	2.37%
230–245	0.71%	0.67%	0.46%	1.33%	1.67%	2.39%	1.86%	2.81%
245–260	0.82%	0.83%	0.70%	1.35%	1.26%	1.63%	1.66%	2.27%
260–285	1.04%	0.98%	0.76%	1.45%	1.81%	2.65%	2.19%	3.18%
285–335	1.22%	1.10%	0.69%	1.76%	1.39%	1.68%	1.91%	2.67%
335–1000	2.71%	1.95%	1.56%	3.55%	2.65%	6.25%	3.64%	7.45%

Table 5.22: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $|\eta^\gamma| < 0.8$, $2.1 < |\eta^{jet}| < 2.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–45	14.37%	15.06%	0.75%	1.83%	17.58%	17.06%	23.16%	22.83%
45–65	9.08%	8.75%	1.38%	1.36%	5.21%	11.58%	10.28%	14.58%
65–72	7.64%	6.98%	0.96%	1.02%	8.76%	10.03%	11.24%	12.26%
72–90	6.90%	8.56%	1.24%	1.45%	5.29%	5.20%	10.13%	10.12%
90–97	3.61%	3.28%	0.64%	1.03%	3.58%	5.79%	4.90%	6.73%
97–105	5.98%	5.08%	1.13%	1.40%	2.59%	4.09%	5.81%	6.67%
105–117	3.09%	2.88%	0.99%	1.56%	2.71%	2.80%	4.08%	4.31%
117–165	2.66%	2.68%	0.81%	1.40%	1.81%	2.11%	3.34%	3.69%
165–172	1.81%	1.69%	0.80%	1.42%	1.64%	2.05%	2.49%	3.01%
172–180	1.47%	1.79%	0.69%	1.37%	2.07%	2.33%	2.82%	3.24%
180–188	1.48%	1.18%	0.63%	1.48%	1.49%	2.75%	2.00%	3.34%
188–196	2.52%	2.24%	1.04%	2.29%	2.50%	5.38%	3.51%	6.26%
196–204	1.90%	1.75%	0.77%	1.49%	1.72%	2.22%	2.57%	3.20%
204–215	2.02%	2.00%	0.70%	1.76%	3.13%	3.22%	3.79%	4.19%
215–230	1.10%	1.26%	0.46%	1.44%	1.66%	2.13%	2.14%	2.87%
230–245	1.43%	1.29%	0.40%	1.46%	1.47%	2.20%	2.00%	2.94%
245–260	1.90%	1.92%	0.50%	1.32%	1.24%	1.49%	2.34%	2.76%
260–285	2.55%	2.93%	0.89%	1.50%	1.62%	2.60%	3.46%	4.19%
285–335	2.83%	2.44%	1.35%	2.42%	1.28%	1.72%	3.07%	3.85%
335–1000	5.91%	4.32%	2.31%	4.10%	3.22%	5.88%	5.86%	8.37%

Table 5.23: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $0.8 < |\eta^\gamma| < 1.5$, $|\eta^{jet}| < 0.8$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–45	4.47%	4.58%	3.16%	2.07%	7.64%	8.71%	9.45%	10.06%
45–65	2.98%	2.91%	3.31%	2.33%	7.78%	8.33%	8.94%	9.12%
65–72	3.16%	3.04%	4.96%	2.75%	6.91%	7.08%	9.04%	8.18%
72–90	2.16%	2.12%	4.04%	2.50%	3.46%	4.32%	5.73%	5.42%
90–97	1.61%	1.57%	4.46%	2.44%	3.80%	4.51%	6.07%	5.36%
97–105	1.70%	1.57%	3.95%	2.19%	4.09%	4.53%	5.90%	5.27%
105–117	1.08%	1.14%	3.48%	2.24%	2.97%	3.06%	4.71%	3.96%
117–165	0.91%	0.88%	4.67%	2.50%	3.72%	3.76%	6.03%	4.60%
165–172	0.40%	0.42%	4.27%	1.99%	3.82%	5.87%	5.75%	6.21%
172–180	0.51%	0.52%	5.30%	2.42%	2.45%	2.09%	5.86%	3.24%
180–188	0.53%	0.53%	5.39%	2.61%	3.81%	3.58%	6.62%	4.47%
188–196	0.48%	0.52%	5.96%	2.26%	2.56%	2.76%	6.51%	3.60%
196–204	0.63%	0.63%	5.45%	2.58%	1.93%	2.54%	5.82%	3.68%
204–215	0.36%	0.40%	4.01%	1.81%	3.06%	5.50%	5.06%	5.80%
215–230	0.50%	0.52%	7.12%	2.33%	1.87%	2.22%	7.38%	3.26%
230–245	0.80%	0.78%	7.54%	2.85%	2.67%	3.16%	8.04%	4.33%
245–260	0.69%	0.74%	7.34%	2.51%	1.44%	1.61%	7.52%	3.07%
260–285	0.72%	0.70%	7.19%	3.03%	1.51%	2.09%	7.38%	3.75%
285–335	1.05%	0.85%	10.28%	3.65%	4.70%	6.91%	11.33%	7.86%
335–1000	1.38%	0.93%	14.16%	4.29%	4.97%	7.37%	15.03%	8.58%

Table 5.24: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $0.8 < |\eta^\gamma| < 1.5$, $0.8 < |\eta^{jet}| < 1.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–45	4.19%	4.41%	3.55%	2.02%	6.57%	7.39%	8.67%	8.84%
45–65	4.01%	4.14%	4.13%	2.53%	7.84%	8.55%	9.78%	9.83%
65–72	3.67%	3.75%	5.53%	3.17%	6.76%	7.04%	9.50%	8.58%
72–90	2.82%	2.88%	3.98%	2.68%	3.50%	4.42%	6.03%	5.91%
90–97	1.72%	1.68%	3.90%	2.03%	3.60%	3.86%	5.57%	4.68%
97–105	1.96%	2.10%	3.30%	1.84%	3.74%	4.03%	5.41%	4.91%
105–117	1.14%	1.23%	3.21%	1.98%	2.63%	2.68%	4.33%	3.55%
117–165	1.06%	1.13%	4.52%	2.26%	3.71%	3.77%	5.96%	4.54%
165–172	0.43%	0.50%	4.19%	1.78%	3.53%	5.51%	5.50%	5.82%
172–180	0.58%	0.63%	4.96%	2.14%	2.36%	2.00%	5.53%	3.00%
180–188	0.57%	0.48%	5.31%	2.49%	3.61%	3.37%	6.44%	4.22%
188–196	0.50%	0.51%	5.71%	2.03%	2.37%	2.46%	6.20%	3.23%
196–204	0.72%	0.58%	5.39%	2.60%	2.06%	2.66%	5.80%	3.76%
204–215	0.39%	0.41%	3.83%	1.65%	3.22%	5.66%	5.02%	5.91%
215–230	0.56%	0.55%	7.04%	2.15%	1.82%	2.24%	7.30%	3.15%
230–245	0.91%	0.84%	6.57%	2.61%	2.47%	3.17%	7.07%	4.20%
245–260	0.83%	0.82%	7.55%	2.47%	1.54%	1.69%	7.75%	3.10%
260–285	0.84%	0.85%	7.64%	3.06%	1.54%	2.05%	7.84%	3.78%
285–335	1.28%	0.97%	11.71%	3.97%	4.01%	6.60%	12.41%	7.76%
335–1000	1.63%	1.19%	13.66%	3.65%	5.11%	6.89%	14.63%	7.89%

Table 5.25: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $0.8 < |\eta^\gamma| < 1.5$, $1.5 < |\eta^{jet}| < 2.1$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–45	7.00%	7.32%	4.37%	2.51%	7.53%	8.39%	11.37%	11.42%
45–65	4.43%	4.21%	3.43%	2.11%	7.45%	7.79%	9.22%	9.10%
65–72	4.86%	4.98%	5.96%	3.19%	6.77%	6.99%	10.30%	9.15%
72–90	3.30%	3.42%	4.25%	2.79%	2.79%	3.18%	6.13%	5.44%
90–97	2.70%	2.67%	3.82%	2.28%	3.95%	5.22%	6.11%	6.29%
97–105	2.81%	2.88%	3.72%	2.11%	4.07%	4.42%	6.22%	5.68%
105–117	1.72%	1.79%	3.56%	2.38%	2.63%	2.71%	4.77%	4.03%
117–165	1.49%	1.38%	4.67%	2.58%	3.59%	3.64%	6.05%	4.67%
165–172	0.62%	0.66%	3.98%	1.85%	3.67%	5.53%	5.46%	5.87%
172–180	0.77%	0.68%	4.74%	1.90%	2.23%	1.90%	5.29%	2.77%
180–188	0.80%	0.66%	4.76%	2.22%	3.51%	3.31%	5.95%	4.04%
188–196	0.74%	0.79%	5.89%	2.10%	2.27%	2.42%	6.36%	3.30%
196–204	0.95%	0.91%	5.57%	2.44%	1.82%	2.33%	5.93%	3.49%
204–215	0.55%	0.60%	3.73%	1.58%	3.20%	5.50%	4.96%	5.75%
215–230	0.78%	0.82%	7.15%	2.03%	1.78%	2.08%	7.41%	3.02%
230–245	1.31%	1.33%	7.41%	2.79%	2.45%	3.10%	7.92%	4.38%
245–260	1.24%	1.01%	8.53%	2.64%	1.59%	1.72%	8.74%	3.31%
260–285	1.30%	1.10%	5.96%	2.60%	1.60%	1.97%	6.27%	3.45%
285–335	1.71%	1.27%	11.75%	3.53%	4.17%	5.79%	12.53%	6.90%
335–1000	3.46%	2.00%	13.16%	4.68%	5.03%	7.38%	14.23%	8.96%

Table 5.26: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $0.8 < |\eta^\gamma| < 1.5$, $2.1 < |\eta^{jet}| < 2.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–45	13.25%	12.04%	4.28%	4.21%	9.24%	10.47%	15.77%	16.50%
45–65	7.55%	7.53%	3.69%	2.20%	6.55%	6.96%	10.64%	10.48%
65–72	7.07%	6.30%	5.24%	2.83%	6.83%	7.30%	10.67%	10.05%
72–90	6.91%	7.36%	5.52%	3.91%	4.49%	5.21%	10.24%	9.83%
90–97	4.51%	4.65%	3.86%	2.45%	4.02%	5.73%	7.26%	7.78%
97–105	5.51%	5.99%	5.36%	3.06%	3.77%	3.96%	8.88%	7.81%
105–117	3.05%	3.02%	4.01%	2.70%	2.69%	2.77%	5.70%	4.91%
117–165	2.72%	2.52%	4.76%	2.57%	3.77%	3.73%	6.57%	5.19%
165–172	1.08%	1.11%	3.80%	1.84%	3.63%	5.49%	5.38%	5.90%
172–180	1.49%	1.38%	5.12%	2.30%	2.53%	2.16%	5.88%	3.45%
180–188	1.52%	1.47%	5.91%	2.59%	3.70%	3.43%	7.13%	4.54%
188–196	1.26%	1.31%	6.77%	2.25%	2.23%	2.33%	7.25%	3.50%
196–204	1.84%	1.59%	5.27%	2.66%	2.00%	2.68%	5.86%	4.10%
204–215	1.05%	1.19%	3.51%	1.54%	2.82%	5.26%	4.66%	5.61%
215–230	1.43%	1.50%	6.56%	2.09%	1.58%	1.82%	6.91%	3.15%
230–245	2.68%	2.41%	9.24%	3.65%	2.67%	3.42%	9.92%	5.55%
245–260	2.60%	2.56%	9.36%	3.53%	1.36%	1.52%	9.80%	4.62%
260–285	2.24%	2.57%	7.51%	2.63%	1.77%	2.07%	8.13%	4.22%
285–335	4.22%	3.29%	12.25%	4.42%	3.20%	4.88%	13.08%	7.36%
335–1000	7.75%	6.52%	6.66%	4.35%	5.81%	7.04%	10.98%	10.53%

Table 5.27: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $1.5 < |\eta^\gamma| < 2.1$, $|\eta^{jet}| < 0.8$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–65	4.03%	4.10%	1.61%	2.16%	9.69%	12.76%	10.64%	13.57%
65–90	4.39%	4.46%	3.44%	3.80%	16.62%	21.15%	17.55%	21.94%
90–105	1.40%	1.36%	2.64%	2.53%	15.90%	22.84%	16.18%	23.02%
105–165	1.33%	1.49%	2.17%	3.15%	16.76%	26.23%	16.96%	26.46%
165–180	0.93%	0.99%	4.74%	4.72%	11.86%	10.81%	12.82%	11.83%
180–200	0.76%	0.85%	5.10%	5.55%	5.60%	12.85%	7.62%	14.02%
200–225	0.85%	0.92%	5.08%	5.92%	7.71%	8.54%	9.28%	10.43%
225–1000	0.85%	0.70%	6.93%	6.08%	4.41%	4.20%	8.24%	7.42%

Table 5.28: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $1.5 < |\eta^\gamma| < 2.1$, $0.8 < |\eta^{jet}| < 1.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–65	4.03%	4.12%	2.35%	2.44%	8.50%	11.10%	9.73%	12.09%
65–90	4.05%	3.59%	3.13%	3.43%	13.80%	17.44%	14.60%	18.13%
90–105	1.24%	1.11%	2.53%	2.42%	14.30%	20.85%	14.56%	21.02%
105–165	1.19%	1.36%	2.25%	2.94%	14.68%	23.87%	14.92%	24.08%
165–180	0.79%	0.82%	5.07%	4.25%	10.31%	9.39%	11.51%	10.34%
180–200	0.66%	0.70%	5.11%	4.73%	4.97%	11.53%	7.17%	12.48%
200–225	0.75%	0.83%	5.04%	5.16%	6.83%	7.55%	8.53%	9.18%
225–1000	0.81%	0.83%	6.64%	5.62%	4.11%	4.02%	7.86%	6.96%

Table 5.29: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $1.5 < |\eta^\gamma| < 2.1$, $1.5 < |\eta^{jet}| < 2.1$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–65	4.41%	4.42%	2.02%	2.43%	7.34%	9.59%	8.80%	10.83%
65–90	3.80%	4.00%	2.95%	3.27%	10.75%	14.20%	11.84%	15.11%
90–105	1.34%	1.34%	2.10%	2.04%	13.30%	17.97%	13.54%	18.14%
105–165	1.25%	1.31%	2.16%	2.82%	13.20%	18.60%	13.44%	18.86%
165–180	0.86%	1.00%	5.08%	4.07%	8.71%	8.16%	10.13%	9.17%
180–200	0.67%	0.62%	4.92%	4.27%	3.96%	9.18%	6.35%	10.14%
200–225	0.81%	0.83%	5.12%	4.58%	6.21%	6.78%	8.09%	8.23%
225–1000	0.91%	0.99%	6.58%	5.11%	3.52%	3.40%	7.53%	6.22%

Table 5.30: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $1.5 < |\eta^\gamma| < 2.1$, $2.1 < |\eta^{jet}| < 2.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–65	7.49%	6.79%	1.58%	2.02%	8.27%	10.50%	10.82%	12.67%
65–90	5.21%	5.02%	3.18%	3.50%	9.85%	12.20%	11.50%	13.65%
90–105	1.75%	1.80%	2.46%	2.17%	12.10%	16.91%	12.48%	17.15%
105–165	2.01%	2.61%	2.22%	3.00%	12.60%	19.23%	13.05%	19.64%
165–180	1.28%	1.37%	5.09%	4.10%	7.86%	7.27%	9.46%	8.45%
180–200	1.05%	1.06%	5.34%	4.25%	3.63%	8.40%	6.55%	9.48%
200–225	1.30%	1.45%	5.08%	4.31%	5.79%	6.30%	7.84%	7.77%
225–1000	1.89%	2.18%	7.46%	5.77%	3.36%	3.26%	8.47%	6.98%

Table 5.31: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $2.1 < |\eta^\gamma| < 2.5$, $|\eta^{jet}| < 0.8$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–65	7.29%	6.86%	3.33%	5.93%	20.64%	41.46%	22.00%	42.44%
65–90	3.89%	3.85%	0.54%	2.47%	61.01%	65.51%	61.14%	65.67%
90–105	3.47%	3.43%	1.39%	2.73%	46.90%	53.48%	47.05%	53.66%
105–165	1.51%	1.53%	0.56%	2.11%	36.14%	35.93%	36.18%	36.02%
165–180	1.37%	1.42%	1.80%	2.13%	41.54%	42.36%	41.60%	42.44%
180–200	1.09%	0.99%	0.68%	2.66%	33.00%	33.48%	33.03%	33.60%
200–225	1.45%	1.86%	1.58%	3.21%	30.96%	34.10%	31.06%	34.30%
225–1000	1.53%	1.90%	1.95%	2.98%	25.61%	26.64%	25.75%	26.87%

Table 5.32: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $2.1 < |\eta^\gamma| < 2.5$, $0.8 < |\eta^{jet}| < 1.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–65	6.02%	6.22%	2.35%	6.51%	15.99%	31.14%	17.31%	32.42%
65–90	3.40%	3.86%	0.61%	2.34%	51.90%	58.48%	52.04%	58.65%
90–105	2.91%	2.99%	1.19%	2.11%	38.30%	43.35%	38.44%	43.51%
105–165	1.32%	1.41%	0.74%	1.81%	32.95%	32.66%	32.99%	32.74%
165–180	1.10%	1.14%	1.69%	2.08%	33.66%	34.49%	33.72%	34.57%
180–200	0.90%	0.84%	0.80%	2.27%	27.99%	28.31%	28.01%	28.42%
200–225	1.18%	1.19%	1.36%	2.70%	25.74%	28.26%	25.81%	28.41%
225–1000	1.24%	1.43%	2.02%	2.59%	20.53%	21.28%	20.68%	21.49%

Table 5.33: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $2.1 < |\eta^\gamma| < 2.5$, $1.5 < |\eta^{jet}| < 2.1$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–65	6.17%	7.92%	2.26%	5.33%	13.46%	25.84%	15.78%	27.55%
65–90	3.05%	2.68%	0.68%	1.90%	43.60%	48.59%	43.69%	48.70%
90–105	2.46%	2.61%	1.05%	1.95%	31.42%	36.36%	31.54%	36.51%
105–165	1.13%	1.29%	0.47%	1.77%	26.16%	24.82%	26.20%	24.92%
165–180	0.98%	1.07%	1.67%	2.01%	27.70%	28.29%	27.77%	28.38%
180–200	0.84%	0.97%	0.96%	1.95%	22.90%	23.26%	22.94%	23.36%
200–225	0.97%	1.16%	1.32%	2.32%	20.12%	22.18%	20.19%	22.33%
225–1000	1.23%	1.48%	1.54%	2.42%	17.95%	18.47%	18.08%	18.68%

Table 5.34: Total uncertainty on purity along with breakdown of the uncertainties into different categories for $2.1 < |\eta^\gamma| < 2.5$, $2.1 < |\eta^{jet}| < 2.5$. Here, ‘Stat.’ refers to statistical uncertainty, ‘Bias Corr.’ refers to uncertainty due to bias correction, ‘Sig Temp.’ refers to uncertainty due to signal template, and ‘Bg Temp.’ refers to uncertainty due to background template selection, respectively. The terms ‘Low’ and ‘High’ refer to the lower and higher uncertainty values for the cases where asymmetric uncertainty calculations are made.

p_T^γ (GeV)	Stat.	Bias Corr.	Sig Temp.		Bg Temp.		Total	
			Low	High	Low	High	Low	High
40–65	6.12%	6.24%	1.14%	3.94%	10.51%	20.10%	12.28%	21.42%
65–90	4.39%	4.84%	0.58%	1.95%	42.28%	45.13%	42.56%	45.43%
90–105	2.73%	2.61%	0.96%	1.70%	26.90%	31.31%	27.04%	31.46%
105–165	1.51%	1.68%	0.66%	1.57%	23.77%	24.03%	23.84%	24.14%
165–180	1.39%	1.56%	1.58%	1.86%	23.64%	24.14%	23.75%	24.27%
180–200	1.13%	1.22%	0.59%	2.04%	19.76%	20.20%	19.80%	20.34%
200–225	1.40%	1.56%	1.30%	1.80%	17.97%	19.92%	18.08%	20.07%
225–1000	2.22%	2.53%	1.74%	2.39%	17.13%	17.65%	17.40%	17.99%

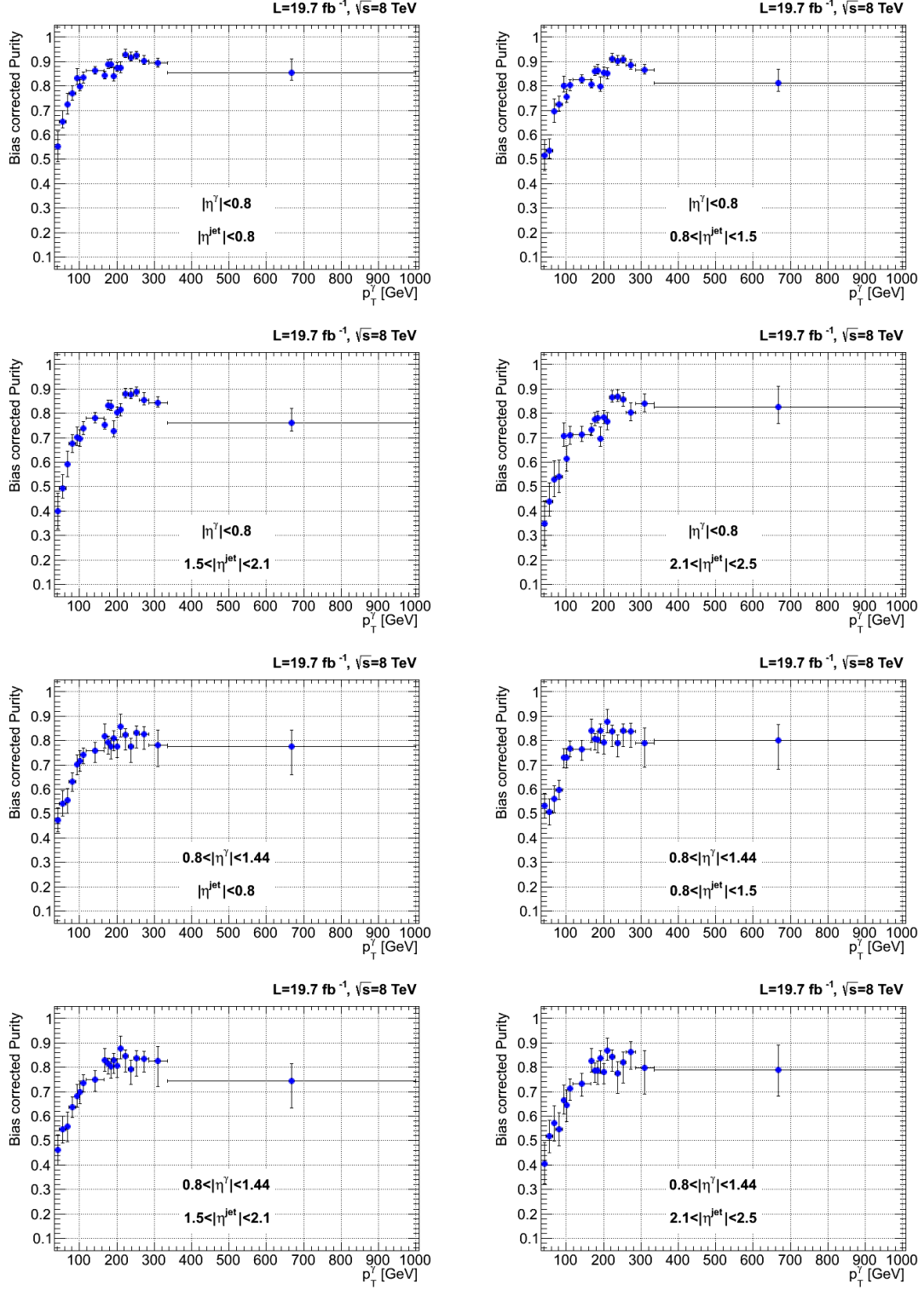


Figure 5.12: Bias corrected purity values as a function of p_T^γ for various $|\eta^\gamma|$, and $|\eta^{jet}|$ combinations, where the photon is in the barrel.

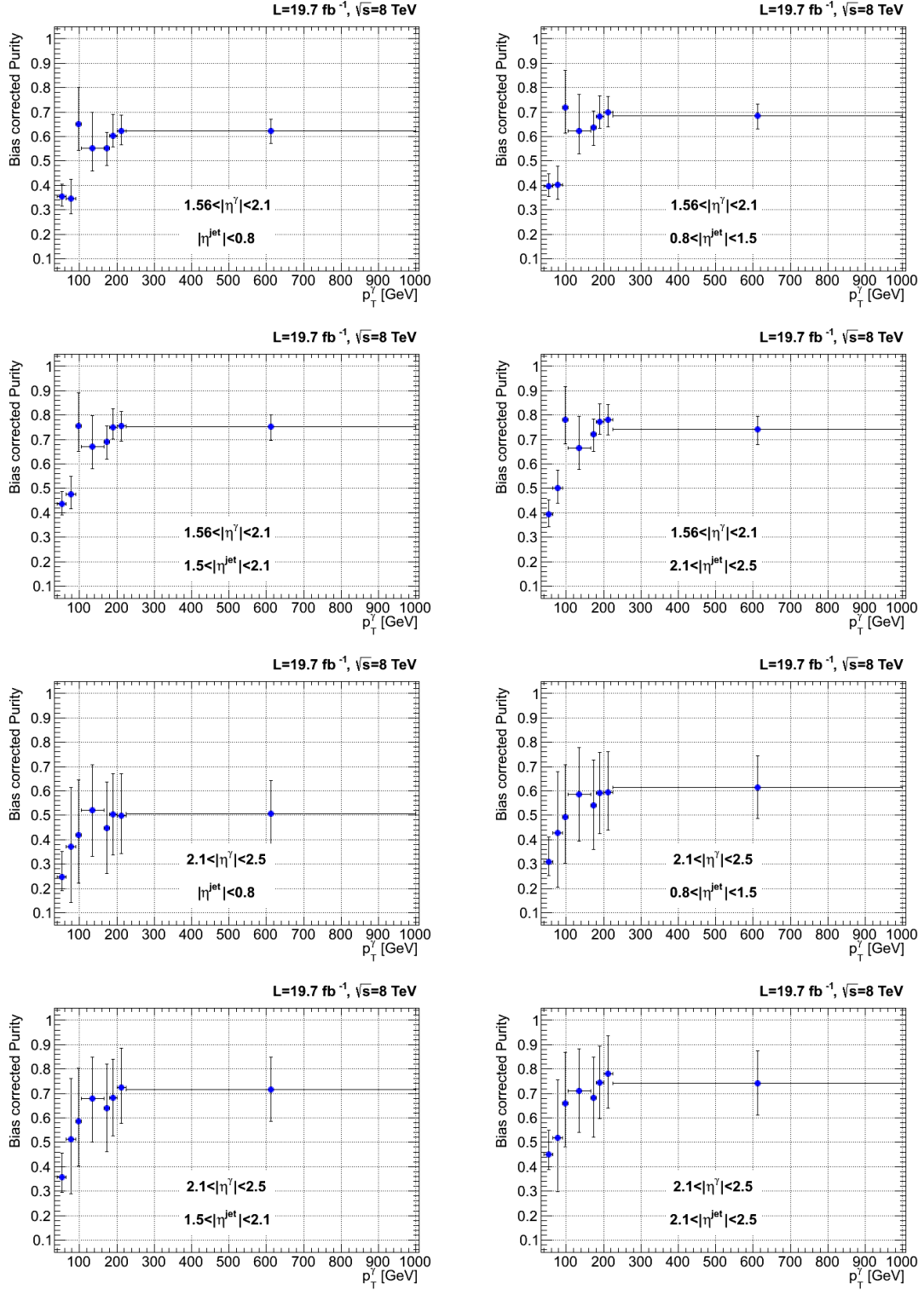


Figure 5.13: Bias corrected purity values as a function of p_T^γ for various $|\eta^\gamma|$, and $|\eta^{jet}|$ combinations, where the photon is in the endcaps.

the selection cuts to the signal events that were produced in the detector. Efficiencies for this measurement include photon trigger efficiencies ($\epsilon_{trigger}$), cluster efficiencies (ϵ_{CLUS}), photon selection efficiencies (ϵ_{SEL}), conversion safe electron veto efficiency (ϵ_{CSEV}), and photon MVA efficiencies (ϵ_{MVA}). The total efficiency for the measurement is therefore,

$$\epsilon_{total} = \epsilon_{trigger}\epsilon_{CLUS}\epsilon_{SEL}\epsilon_{CSEV}\cdot\epsilon_{MVA} \quad (5.12)$$

In the above equation, $\epsilon_{trigger}$ refers to the efficiency for a reconstructed photon to pass the trigger system, ϵ_{CLUS} refers to the efficiency for the photon produced inside the detector acceptance to be clustered using reconstruction algorithms, ϵ_{SEL} is the efficiency for the reconstructed photon to pass the photon selection criteria for this analysis, ϵ_{CSEV} is the efficiency for the reconstructed photon that passes the selection criteria to also pass the conversion safe electron veto, and ϵ_{MVA} is the efficiency for the reconstructed photon that passes all the above criteria to also pass the MVA cut. The efficiencies are not correlated with each other; therefore the total efficiency can be taken simply as the product of the individual efficiencies. Each of the individual efficiencies is estimated in a different way. However, one of the techniques that is used is a Tag and Probe method, which is briefly explained below.

5.6.1 Tag and Probe

The Tag and Probe technique exploits the di-object resonances such as $Z \rightarrow ee$ to measure some efficiency associated with one of the decay particles. The ‘tag’ is defined as one of the decay products that passes a tight identification, whereas the ‘probe’ is the particle whose efficiency is being studied. The probe is required to pass only a loose identification so as not to bias the efficiency measurement. The ‘tight’ criteria and the ‘loose’ criteria are dependent on a particular analysis. Based on whether the probe passes the criteria of the efficiency being studied, the probe is either categorized as passing or failing. The tag and probe objects are paired and their invariant mass is calculated for both passing and failing categories. From the pairs, the number of signal events is estimated by fitting the invariant mass peak. The number of signal events that are in the passing probe + tag pair form the numerator for the efficiency measurement. The number of signal events that are in the either of the two pairs (i.e. passing probe + tag or failing probe + tag) form the

denominator. Since the electrons at CMS are reconstructed by pairing ECAL energy depositions with the tracks in the tracker, electron objects can give information about the photon triggering, reconstruction, and identification if the information from tracking is simply ignored. Therefore, the Tag and Probe technique can be used with electrons to measure many of the efficiencies.

5.6.2 Trigger Efficiency

Section 5.2 discusses the measurement of relative trigger efficiencies for the various HLT paths used in the photon+jet cross section measurement. Since the HLT paths are combined at different values of p_T^γ , their relative efficiencies are relevant only in the specific section of p_T^γ range where they are used. Figure 5.16 shows the relative trigger efficiency values as a function of the p_T bins used in this analysis. The values are also provided in Table 5.35. The efficiency measurements are carried out separately for barrel and endcaps. For the p_T bins where HLT_Photon30_CaloIdVL_v* is used, the relative efficiency associated with HLT_Photon50_CaloId- VL_v* is used since bootstrapping technique is applied only for HLT paths with p_T threshold greater than 50 GeV.

Once the relative efficiencies are measured, HLT_Photon135_v* for which there is no prescale is used to estimate the absolute trigger efficiency of the HLT path. This accounts for not just the total HLT path efficiency but also the L1 trigger efficiency associated with the HLT objects. The absolute efficiency is estimated using the Tag and Probe technique using a single electron trigger. The ‘tag’ electron is required to pass a single electron HLT path detailed in Section 5.2. The ‘tag’ electron reconstructed as a photon without the tracking information is required to pass all the photon selection criteria for this measurement except for the conversion safe electron veto which is used to reduce electron backgrounds. The probe is required to be a photon object that passes a loose cut on $\sigma_{i\eta i\eta}$ and H/E . The $\sigma_{i\eta i\eta}$ cut is the same as what is used in the preselection. However, since the H/E cut in the preselection has an R_9 dependence, a simple cut of $H/E < 0.075$ is applied to all the probes. The passing probes are further required to pass the single photon HLT path whose efficiency is being studied. With the tag and probe technique, the absolute HLT path efficiency is measured separately for barrel and endcaps. The total trigger efficiency measured with this method is compared with the relative efficiency measured using the ratio method over a few bins in p_T^γ as shown in Figure 5.14.

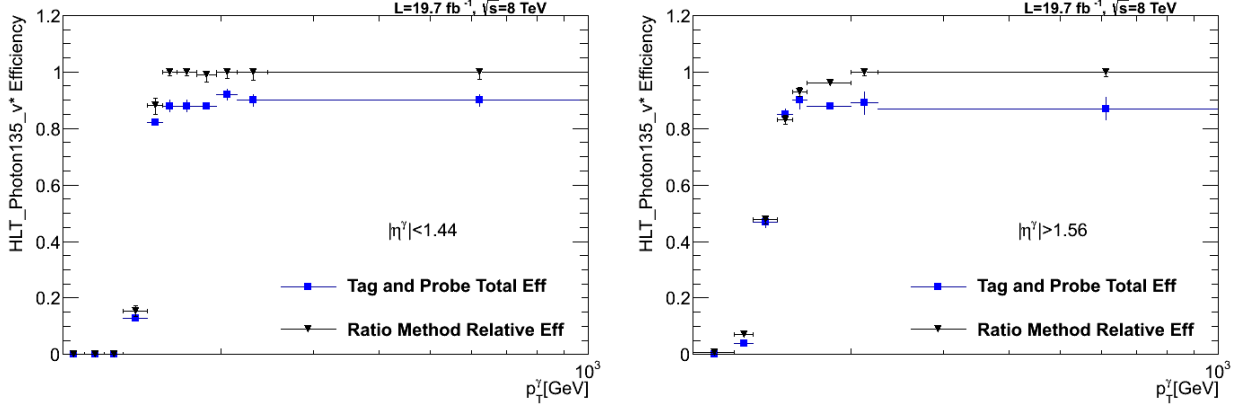


Figure 5.14: Trigger efficiency as a function of reconstructed photon p_T for the HLT_Photon135_v* path in the barrel (left) and endcap (right) regions. The relative efficiencies are found using the ratio/bootstrapping method and the absolute efficiencies are estimated using the Tag and Probe method.

Using the ratio of the absolute efficiency to the relative efficiency for the highest p_T^γ bin, where the triggers are expected to be fully turned on, scale factors are calculated. The scale factor is found to be $0.90^{+0.020}_{-0.032}$ in the barrel and $0.87^{+0.040}_{-0.043}$ in the endcap region. This scale factor (SF) is used to scale the relative efficiencies for the remaining p_T^γ bins, and compare to those calculated directly using the Tag and Probe method as shown in Figure 5.15. For the barrel region, there is very good agreement between the efficiencies measured using these two methods. In the endcaps, although there are slight discrepancies between the two methods at low p_T^γ bins, the agreement is within the statistical uncertainties in the region where the particular HLT path is used in the analysis. The SFs are used to scale not just the HLT_Photon135_v* path, but also all the other paths used in this analysis. The assumption is that the difference in efficiency between the two methods comes from the same source, i.e. the L1 trigger, for all the HLT paths. The final values of the SF-corrected trigger efficiencies as a function of p_T^γ bins used for the $\gamma + \text{jet}$ cross section measurement are presented in Figure 5.16 and provided in Table 5.35.

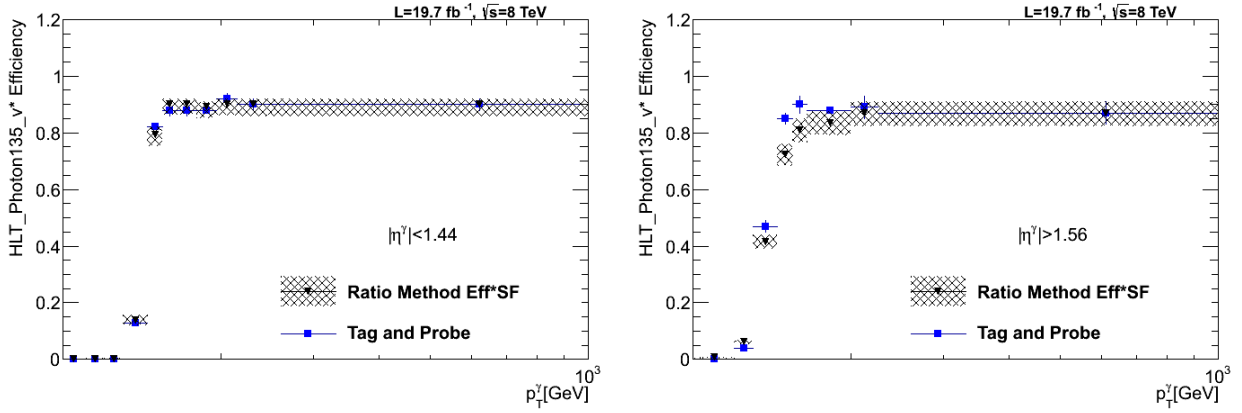


Figure 5.15: Trigger efficiency as a function of reconstructed photon p_T for the HLT_Photon135_v* path in the barrel (left) and endcap (right) region. The relative efficiencies are multiplied by a scale factor estimated by taking the ratio between the two methods (Tag and Probe to estimate absolute efficiencies and ratio/bootstrapping to calculate relative efficiency). The shaded area shows the uncertainty on the relative efficiency scaled by the SF, whereas the blue error bar shows the statistical uncertainty associated with the Tag and Probe method.

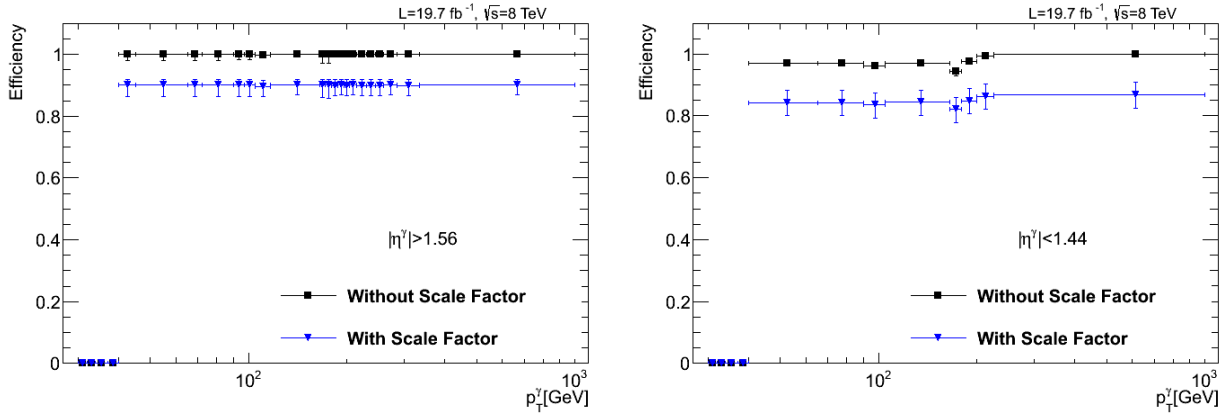


Figure 5.16: Trigger efficiency as a function of reconstructed photon p_T for barrel (left) and endcap (right) regions. The values correspond to the relative efficiency measurements for HLT_Photon50_CaloIdVL_v*, HLT_Photon75_CaloIdV_v*, HLT_Photon90_CaloIdVL_v*, HLT_Photon135_v*, and HLT_Photon150_v* with and without the application of the Scale factor at the p_T^γ ranges where these paths are used.

Table 5.35: Trigger efficiency values for different p_T^γ bins in the barrel ($|\eta^\gamma| < 1.44$) ($1.56 < |\eta^\gamma| < 2.5$) region.

p_T^γ [GeV]	Relative eff.	Absolute eff.	p_T^γ [GeV]	Relative eff.	Absolute eff.
$ \eta^\gamma < 1.44$			$1.56 < \eta^\gamma < 2.5$		
40–45	$1.00^{+0.00\%}_{-1.98\%}$	$0.90^{+3.65\%}_{-3.18\%}$	40–65	$0.97^{+0.68\%}_{-0.85\%}$	$0.84^{+4.19\%}_{-4.17\%}$
45–65	$1.00^{+0.00\%}_{-1.90\%}$	$0.90^{+3.61\%}_{-3.18\%}$	65–90	$0.97^{+0.68\%}_{-0.85\%}$	$0.84^{+4.19\%}_{-4.17\%}$
65–72	$1.00^{+0.00\%}_{-1.98\%}$	$0.90^{+3.65\%}_{-3.18\%}$	90–105	$0.96^{+0.70\%}_{-0.83\%}$	$0.84^{+4.15\%}_{-4.13\%}$
72–90	$1.00^{+0.00\%}_{-1.90\%}$	$0.90^{+3.61\%}_{-3.18\%}$	105–165	$0.97^{+0.27\%}_{-0.29\%}$	$0.84^{+4.14\%}_{-4.14\%}$
90–97	$1.00^{+0.00\%}_{-1.79\%}$	$0.90^{+3.57\%}_{-3.18\%}$	165–180	$0.94^{+1.25\%}_{-1.54\%}$	$0.82^{+4.24\%}_{-4.16\%}$
97–105	$1.00^{+0.00\%}_{-1.82\%}$	$0.90^{+3.58\%}_{-3.18\%}$	180–200	$0.98^{+0.12\%}_{-0.12\%}$	$0.85^{+4.15\%}_{-4.15\%}$
105–117	$1.00^{+0.23\%}_{-0.46\%}$	$0.90^{+3.20\%}_{-3.18\%}$	200–225	$0.99^{+0.08\%}_{-0.09\%}$	$0.86^{+4.23\%}_{-4.23\%}$
117–165	$1.00^{+0.00\%}_{-0.26\%}$	$0.90^{+3.19\%}_{-3.18\%}$	225–1000	$1.00^{+0.04\%}_{-0.05\%}$	$0.87^{+4.25\%}_{-4.25\%}$
165–172	$1.00^{+0.00\%}_{-2.75\%}$	$0.90^{+4.03\%}_{-3.18\%}$			
172–180	$1.00^{+0.00\%}_{-2.93\%}$	$0.90^{+4.13\%}_{-3.18\%}$			
180–188	$1.00^{+0.03\%}_{-0.07\%}$	$0.90^{+3.18\%}_{-3.18\%}$			
188–196	$1.00^{+0.00\%}_{-0.06\%}$	$0.90^{+3.18\%}_{-3.18\%}$			
196–204	$1.00^{+0.05\%}_{-0.11\%}$	$0.90^{+3.18\%}_{-3.18\%}$			
204–215	$1.00^{+0.00\%}_{-0.07\%}$	$0.90^{+3.18\%}_{-3.18\%}$			
215–230	$1.00^{+0.03\%}_{-0.09\%}$	$0.90^{+3.18\%}_{-3.18\%}$			
230–245	$1.00^{+0.05\%}_{-0.13\%}$	$0.90^{+3.18\%}_{-3.18\%}$			
245–260	$1.00^{+0.06\%}_{-0.18\%}$	$0.90^{+3.18\%}_{-3.18\%}$			
260–285	$1.00^{+0.00\%}_{-0.13\%}$	$0.90^{+3.19\%}_{-3.18\%}$			
285–335	$1.00^{+0.06\%}_{-0.17\%}$	$0.90^{+3.18\%}_{-3.18\%}$			
335–1000	$1.00^{+0.00\%}_{-0.16\%}$	$0.90^{+3.19\%}_{-3.18\%}$			

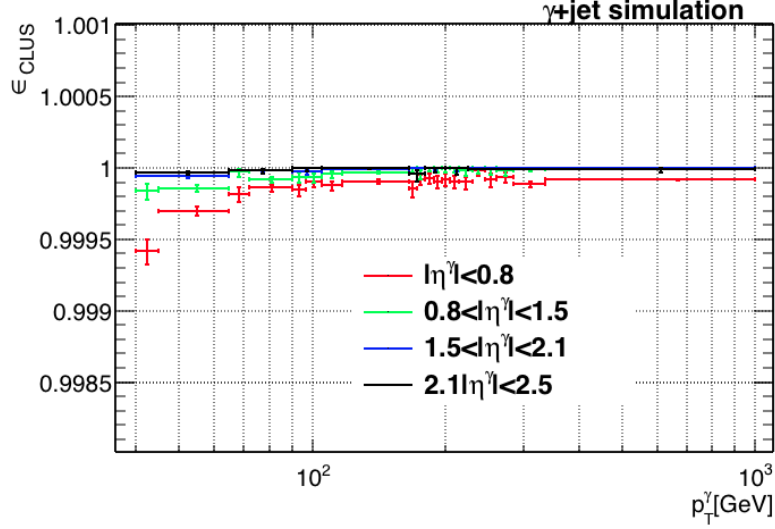


Figure 5.17: Clustering efficiency as a function of generated photon p_T in four different generated $|\eta^\gamma|$ regions.

5.6.3 Cluster Efficiency

The efficiency of the photon clustering is measured using γ +jet MC simulation. The efficiency is evaluated by taking the generated photons that match reconstructed photons in the numerator and dividing by all the generated photons. The generated photons are required to pass all the selection criteria for generator level photon objects as described in Chapter 4. The efficiency are approximately 100% in all bins as can be seen from Figure 5.17.

5.6.4 Selection Efficiency

The selection efficiency measures the efficiency for a reconstructed photon to pass the photon isolation cut and the selection cuts described in Section 5.4.1 with the exception of the conversion safe electron veto. To obtain the efficiency, γ +jet MC simulation is used. The reconstructed photons are required to match a generator level signal photon. From these truth matched reconstructed photons, the number of those that pass the selection cuts under study is taken as the numerator, whereas the total number of reconstructed photons regardless of whether or not they pass the selection cuts is taken as the denominator. The efficiency measured using this method is indicated as

‘Gen γ +jet’ in Figure 5.18. Three other methods are used to measure the same efficiency, and the difference between the methods is taken as the systematic uncertainty in the efficiency measurement.

In the second method (labeled as ‘Gen $Z \rightarrow ee$ ’ in Figure 5.18), generator level information from $Z \rightarrow ee$ MC is used. A reconstructed supercluster is required to match MC electron coming from Z decay. The reconstructed events that pass the photon selection criteria under study are taken in the numerator, while all the events regardless of whether they pass or fail the criteria are taken in the denominator to estimate the efficiency. Comparing the results from first method to this method gives an uncertainty in the efficiency due to inherent differences between the photon and electron objects.

The third and fourth methods use the Tag and Probe technique in MC and data, respectively. In this method, the Z resonance is reconstructed from the invariant mass of tag and a probe electron. Tag electrons are required to pass tight selection criteria, which is the full event selection defined in Section 5.4.1. The probes that pass the selection cuts under study are categorized as passing probes. The passing probe + tag and failing probe + tag are fit separately to obtain the amount of signal present under the Z peak. The efficiency is the ratio of the signal yield of the passing probe + tag combination to the signal yield of any probe + tag combination. The third and fourth methods are labeled as ‘MC T&P’ and ‘Data T&P’, respectively, in Figure 5.18.

The comparison of the efficiencies obtained using the second and third methods gives the uncertainty in efficiency due to the choice of technique. Finally, the comparison of efficiencies obtained using the third and fourth methods gives the uncertainty due to data-MC differences. All three sources of uncertainty are added in quadrature to get the total systematic uncertainty in the efficiency measurement from the first method.

5.6.5 Conversion Safe Electron Veto Efficiency

The conversion safe electron veto efficiency is obtained from the official estimate by the CMS collaboration. It is reported as $(99.1 \pm 0.1)\%$ in the barrel and $(97.8 \pm 0.2)\%$ in the endcaps for 8

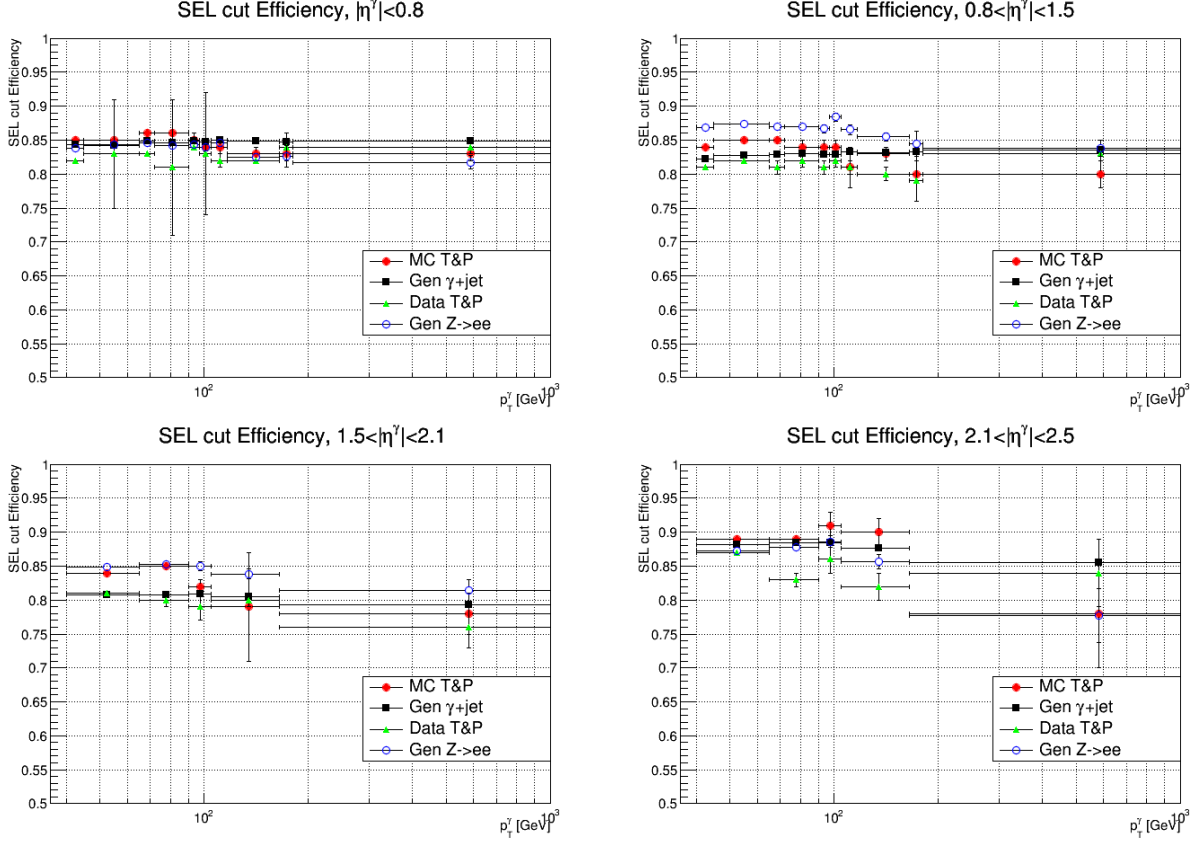


Figure 5.18: Selection efficiencies as a function of generated p_T^γ in four different generated $|\eta^\gamma|$ regions using four different methods.

Table 5.36: Selection efficiencies for $|\eta^\gamma| < 0.8$.

p_T^γ (GeV)	Efficiency	Uncertainty				
		Stat.	γ/e	Tag and Probe	MC/data	Total
40–45	0.84	0.47%	0.61%	1.03%	3.33%	3.57%
45–65	0.84	0.32%	0.17%	1.26%	2.60%	2.91%
65–72	0.85	0.22%	0.30%	1.58%	3.82%	4.15%
72–90	0.85	0.21%	0.47%	1.53%	5.91%	6.12%
90–97	0.85	0.19%	0.60%	0.86%	1.80%	2.09%
97–105	0.85	0.19%	0.85%	0.02%	1.48%	1.72%
105–117	0.85	0.16%	0.44%	0.41%	2.94%	3.00%
117–165	0.85	0.09%	2.74%	0.48%	1.20%	3.03%
165–180	0.85	0.15%	2.55%	0.50%	0.62%	2.68%
180–1000	0.85	0.07%	3.66%	1.50%	0.88%	4.05%

Table 5.37: Selection Efficiency for $0.8 < |\eta^\gamma| < 1.5$.

p_T^γ (GeV)	Efficiency	Uncertainty				
		Stat.	γ/e	Tag and Probe	MC/data	Total
40–45	0.82	0.56%	5.61%	4.01%	3.22%	7.63%
45–65	0.83	0.20%	5.59%	3.13%	3.16%	7.14%
65–72	0.83	0.30%	5.03%	2.66%	4.35%	7.17%
72–90	0.83	0.17%	4.81%	3.95%	2.29%	6.63%
90–97	0.83	0.23%	4.60%	3.25%	3.62%	6.70%
97–105	0.83	0.22%	6.75%	5.31%	2.42%	8.92%
105–117	0.83	0.19%	3.99%	6.69%	0.00%	7.79%
117–165	0.83	0.10%	2.89%	3.00%	3.61%	5.51%
165–180	0.83	0.17%	1.43%	5.35%	1.20%	5.67%
180–1000	0.83	0.08%	0.33%	4.49%	3.59%	5.76%

Table 5.38: Selection efficiency for $1.5 < |\eta^\gamma| < 2.1$.

p_T^γ (GeV)	Efficiency	Uncertainty				
		Stat.	γ/e	Tag and Probe	MC/data	Total
40–65	0.81	0.48%	5.20%	1.30%	3.74%	6.55%
65–90	0.81	0.22%	5.55%	0.70%	5.82%	8.07%
90–105	0.81	0.22%	5.10%	3.71%	3.71%	7.32%
105–165	0.81	0.14%	4.14%	6.04%	1.24%	7.43%
165–1000	0.79	0.13%	2.59%	4.26%	2.52%	5.59%

Table 5.39: Selection efficiency for $2.1 < |\eta^\gamma| < 2.5$.

p_T^γ (GeV)	Efficiency	Uncertainty				
		Stat.	γ/e	Tag and Probe	MC/data	Total
40–65	0.88	0.17%	1.10%	2.51%	2.78%	3.91%
65–90	0.88	0.21%	0.68%	1.51%	6.52%	6.73%
90–105	0.89	0.19%	0.12%	2.71%	5.65%	6.27%
105–165	0.88	0.14%	2.38%	4.99%	9.12%	10.67%
165–1000	0.86	0.16%	9.15%	0.33%	7.01%	11.53%

TeV. The efficiencies are obtained from photons in $Z \rightarrow \gamma\mu\mu$ events [29].

5.6.6 MVA Efficiency

The MVA efficiency is the efficiency for a reconstructed photon passing all the selection cuts including CSEV to also pass the MVA response cut. The methods implemented to measure this efficiency are similar to those used to estimate the selection efficiency discussed in Section 5.6.4. The generator level signal photon is matched to a reconstructed photon that passes all the selection criteria including CSEV. The efficiency is obtained from the ratio of the number of photons that also pass the MVA cut to the total number of photons.

In the second method, a generator level electron coming from Z decay is matched to a photon that passes all the selection criteria as well as CSEV. The efficiency is the ratio of the number of reconstructed photons that also pass the MVA cut to all of the photons. The difference between first method and the second method gives the uncertainty in the MVA efficiency due to inherent differences between the electron and photon objects. As shown in Figure 5.19, the difference in efficiencies between these two methods is substantial. The difference comes from the choice of input variables used in the training of the MVA. For example, the R_9 variable that is one of the inputs for the MVA behaves differently for electrons and photons.

In the third and fourth methods, the Tag and Probe technique is used in $Z \rightarrow ee$ data and MC. The tag electron is required to pass all of the selection criteria, while the probe electron is categorized as a passing probe if it passes all the cuts including MVA, and is categorized as failing probe if it fails just the the MVA cut. The invariant mass for tag and probe combinations are fit for both passing and failing scenario. The signal yield for passing probe + tag pairs is divided by the signal yield for all probe + tag pair to get the efficiency. The difference between second and third methods as well as third and fourth method are taken as systematic uncertainties in the efficiency measurement. All three contributions to the systematic uncertainty are added in quadrature to get the total systematic uncertainty in the MVA efficiency, while the estimates from the first method are used as the central values for the MVA efficiency.

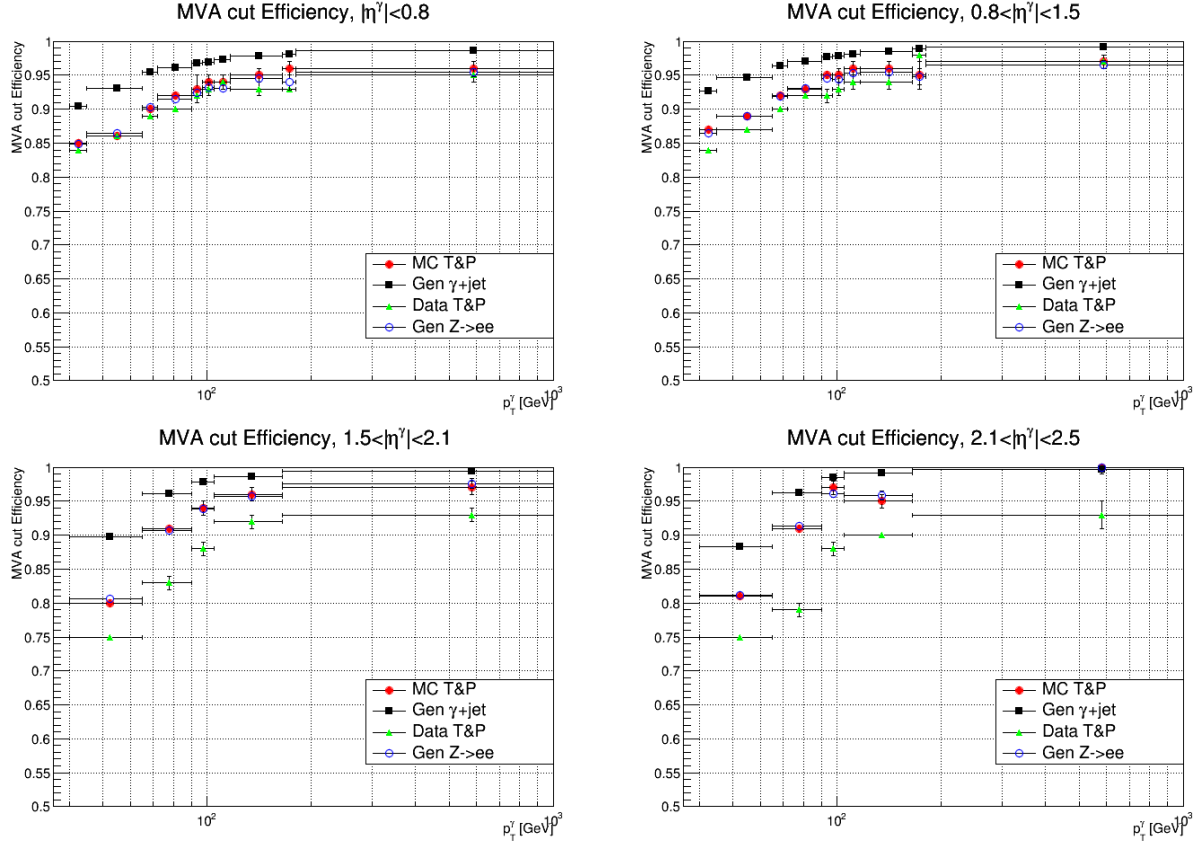


Figure 5.19: MVA cut efficiencies as a function of generated p_T^γ in four different generated $|\eta^\gamma|$ regions using four different methods.

Table 5.40: MVA cut efficiency for $|\eta^\gamma| < 0.8$.

p_T^γ (GeV)	Efficiency	Uncertainty				
		Stat.	γ/e	TP	MC/data	Total
40–45	0.90	0.20%	6.18%	0.20%	1.64%	6.40%
45–65	0.93	0.09%	7.15%	0.06%	0.81%	7.20%
65–72	0.95	0.11%	5.41%	0.16%	1.57%	5.64%
72–90	0.96	0.07%	4.77%	0.86%	2.29%	5.36%
90–97	0.97	0.10%	4.46%	0.60%	1.45%	4.72%
97–105	0.97	0.09%	3.88%	1.14%	1.44%	4.30%
105–117	0.97	0.09%	4.38%	0.63%	0.21%	4.43%
117–165	0.98	0.03%	3.44%	0.42%	2.04%	4.02%
165–180	0.98	0.06%	4.27%	2.09%	2.95%	5.60%
180–1000	0.99	0.02%	3.20%	0.76%	1.22%	3.51%

Table 5.41: MVA cut efficiency for $0.8 < |\eta^\gamma| < 1.5$.

p_T^γ (GeV)	Efficiency	Uncertainty				
		Stat.	γ/e	TP	MC/data	Total
40–45	0.93	0.17%	6.66%	0.24%	3.20%	7.39%
45–65	0.95	0.08%	6.05%	0.08%	2.31%	6.47%
65–72	0.96	0.11%	4.62%	0.53%	2.39%	5.23%
72–90	0.97	0.06%	4.09%	0.26%	1.34%	4.31%
90–97	0.98	0.08%	3.19%	0.81%	3.65%	4.92%
97–105	0.98	0.08%	3.56%	1.01%	2.45%	4.44%
105–117	0.98	0.07%	2.83%	0.66%	2.55%	3.86%
117–165	0.99	0.03%	3.11%	0.64%	1.73%	3.61%
165–180	0.99	0.05%	4.13%	0.22%	2.83%	5.01%
180–1000	0.99	0.02%	2.62%	0.53%	0.20%	2.68%

Table 5.42: MVA cut efficiency for $1.5 < |\eta^\gamma| < 2.1$.

p_T^γ (GeV)	Efficiency	Uncertainty				
		Stat.	γ/e	TP	MC/data	Total
40–65	0.90	0.15%	10.23%	0.40%	5.83%	11.78%
65–90	0.96	0.08%	5.64%	0.16%	7.70%	9.55%
90–105	0.98	0.07%	4.07%	0.09%	5.62%	6.94%
105–165	0.99	0.04%	2.96%	0.15%	3.95%	4.94%
165–1000	0.99	0.02%	1.81%	0.31%	4.33%	4.70%

Table 5.43: MVA cut efficiency for $2.1 < |\eta^\gamma| < 2.5$.

p_T^γ (GeV)	Efficiency	Uncertainty				
		Stat.	γ/e	TP	MC/data	Total
40–65	0.88	0.19%	8.08%	0.49%	6.51%	10.39%
65–90	0.96	0.11%	5.08%	0.51%	11.89%	12.94%
90–105	0.98	0.07%	2.36%	0.38%	8.63%	8.96%
105–165	0.99	0.04%	3.32%	0.59%	5.24%	6.24%
165–1000	1.00	0.02%	0.29%	0.01%	7.02%	7.03%

5.7 Unfolding

For any high-energy physics measurement, the distribution $f(x)$ of any kinematic quantity x is distorted by detector effects such as non-linear response, limited acceptance, and finite resolution of the detector. The effects can either be loss of events or migration of the data between different bins in x . Smearing of the variables is corrected for so that the measured distribution can be compared to the theory or other experiments that will have different detector resolution and acceptance. This is achieved through a process known as unfolding. The first stage of unfolding involves construction of a response matrix that relates the true quantities to the reconstructed quantities. Suppose the reconstructed distribution is divided in different bins in x : x_i^{rec} . The response/migration matrix element R_{ij} is defined as the probability that event from truth bin j is found in the reconstructed bin i . The R_{ij} is calculated from a MC training sample.

The expected number of events in bin i (μ_i) is obtained from the true number of events by summing over contributions from all the bins:

$$\mu_i = \sum_j R_{ij} x_j^{truth}, \quad (5.13)$$

where, x_j^{truth} is the true number of events in the bin j . In data, the known is the reconstructed distribution and the unknown is the truth. To get the estimate for the true value from the above equation, the response matrix has to be inverted and applied to the measured data. The RooUnfold[37] framework implemented in ROOT provides several algorithms for solving the above equation to get the estimates for the true values. The method that is used for this analysis is the iterative ‘Bayesian’ method as proposed by D’Agostini [38]. A matrix inversion technique and a bin-by-bin method of estimating the inverse of the response matrix are also used for comparison for completeness. However, these methodologies are simplistic and are not robust against bias.

The bin-by-bin method is a simplistic method where correction factors C_i are obtained from MC for each bins, i :

$$C_i = \frac{x_i^{truth}}{x_i^{rec}}. \quad (5.14)$$

The correction factors are then applied to the data in the same bin to get the unfolded value,

$$x_i^{un,BBB} = C_i x_i^{data}, \quad (5.15)$$

where $x_i^{un,BBB}$ is the bin-by-bin estimates of the true value, and x_i^{data} is the data measurement. This technique makes the assumption that any off-diagonal elements from the response matrix are equivalent to the loss in the acceptance, thereby neglecting contributions from bin migration.

If the response/migration matrix is an invertible square matrix, Equation 5.13 can be solved easily by applying the inverse matrix to the reconstructed data distribution in order to get the unfolded distribution. The matrix inversion techniques use this method. Therefore, the unfolded distribution in bin i ($x_i^{un,in}$) can be obtained as

$$x_i^{un,in} = \sum_j R_{ij}^{-1} x_j^{data}. \quad (5.16)$$

In this method, the random fluctuations are amplified. Smearing of the truth distribution washes out the sharp features, and since ‘unsmearing’ has to create those features, any small noise is amplified. In order to make the unfolded distribution match the true distribution, smoothing techniques could be implemented. However, in the matrix inversion method used for this measurement, no smoothing criterion is applied.

The Bayesian approach is a way of learning about the relationships of physical quantities from experimental data using probability theory. The goal is to find all the possible distributions that may have caused the observed distribution. The Bayesian approach implemented by ROOT consists of a regularization parameter, which is the number of iterations. The algorithm inverts the response matrix in many iterations, each time using the results from the previous iteration. In doing so, the MC model dependence is reduced in each step. However, too many iterations will cause the algorithm to be sensitive to random bin-by-bin fluctuations, thereby creating artificial effects. Therefore, regularization is achieved by stopping the process before reaching the ‘true’ inverse [37].

In the γ + jet measurement, the differential cross section measured as a function of p_T^γ , $|\eta^\gamma|$ and $|\eta^{jet}|$ is corrected for the smearing only in p_T^γ bins. Unfolding is not done for $|\eta^\gamma|$ and $|\eta^{jet}|$ because the bins in the η variables are much wider compared to the detector resolution. The p_T^γ bins used in this analysis have variable width. Though the migration between bins is not expected to be significant, it maybe more pronounced for bins with small width. A migration matrix is obtained

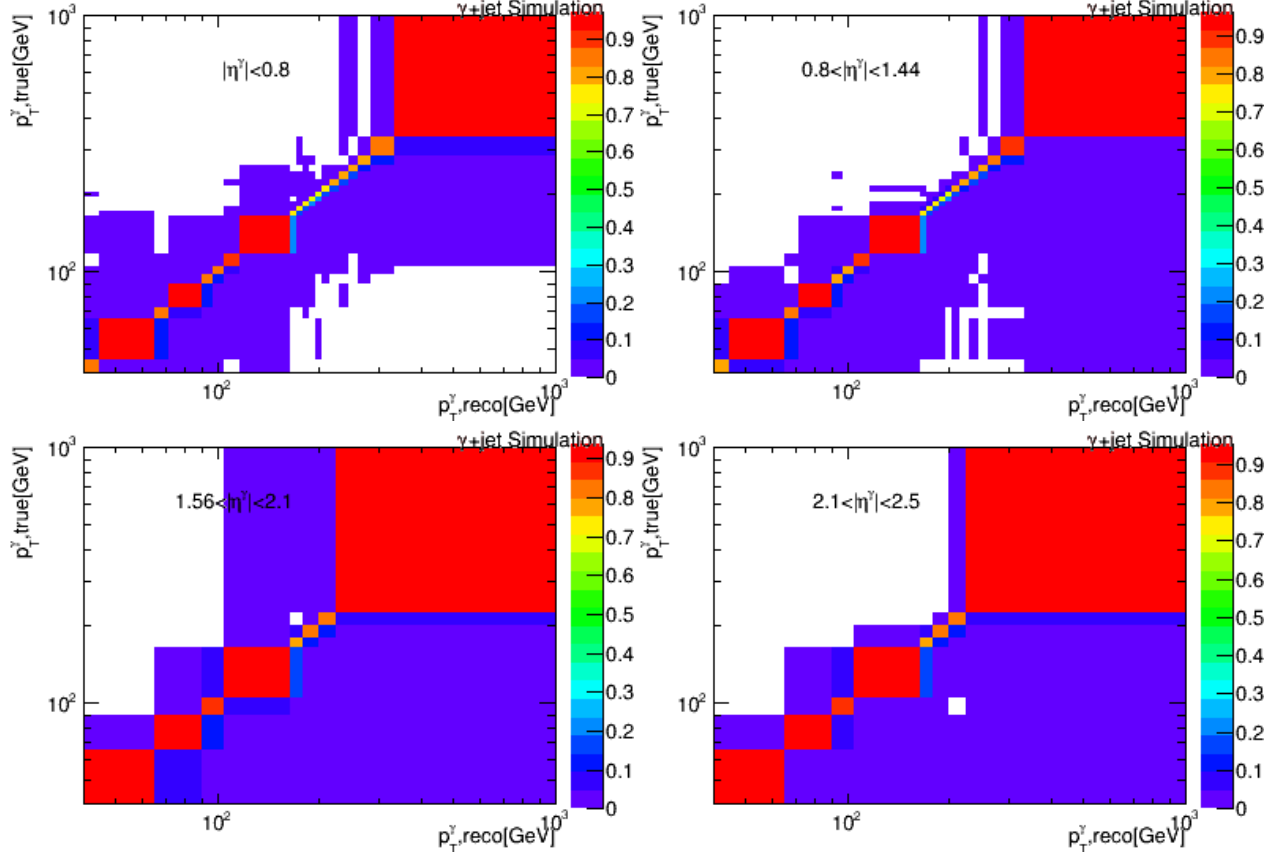


Figure 5.20: Response matrices obtained for different choices of $|\eta^\gamma|$. $\gamma + \text{jet}$ MC events passing all the event selection criteria are used for obtaining the matrices. The migration from over-flow and under-flow bins (not shown) is also considered.

from a $\gamma + \text{jet}$ MC. For the matrix construction, two additional bins (over-flow and under-flow bins) at the lower and the upper p_T^γ boundaries are included in order to correctly approximate the fraction of events in the first and last bins of the true distribution. However, these additional bins are not shown in the plots.

The MC samples used for the response matrix construction are required to pass the same selection criteria as the candidate samples in the rest of the analysis. The four response matrices are provided in Figure 5.20. The response matrices are used on a statistically independent set of $\gamma + \text{jet}$ MC events to perform a closure test. The closure test is performed for the bin-by-bin, matrix inversion, and iterative Bayesian methods. The unfolded results for MC are seen to be consistent

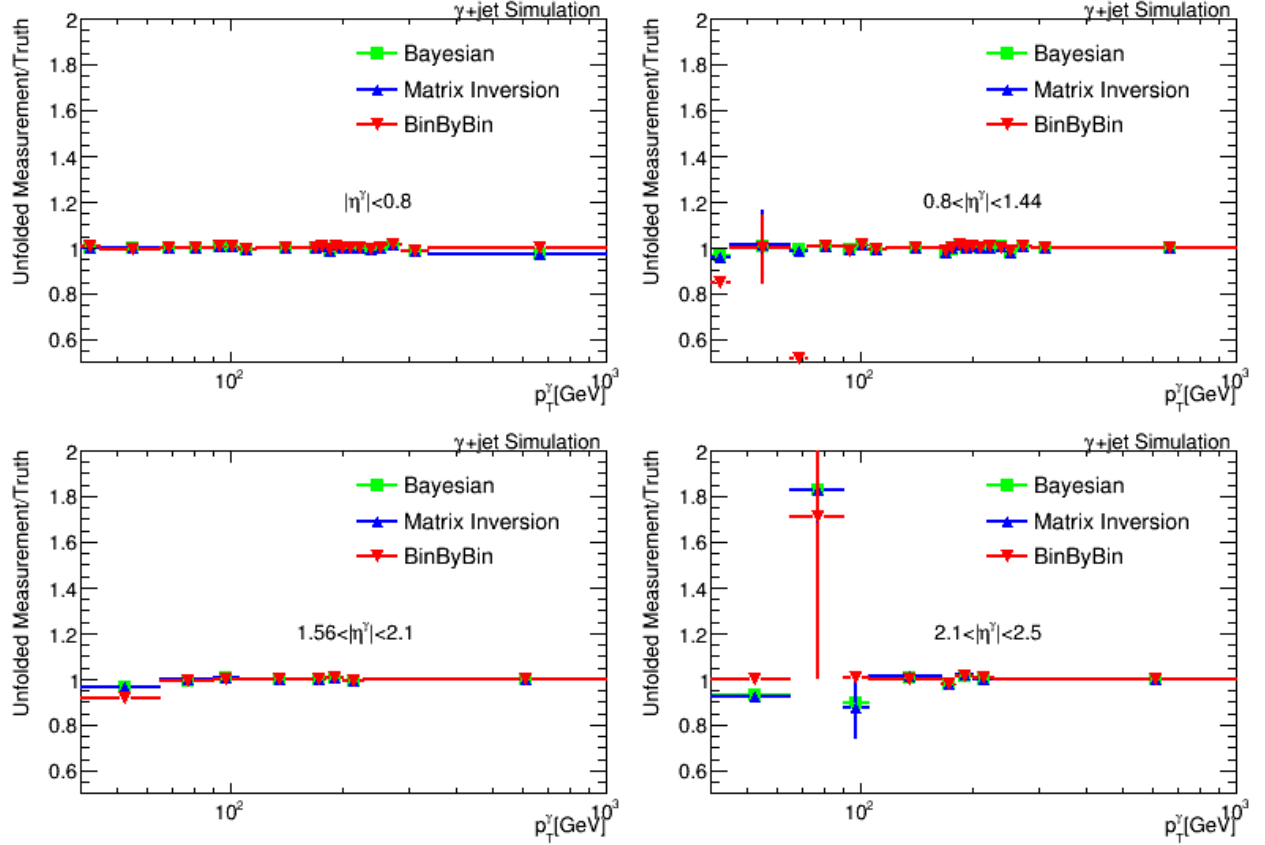


Figure 5.21: Ratio of the $\gamma + \text{jet}$ MC testing sample unfolded using the response matrices from Figure 5.20 to the truth distribution. The closure test is done using three different algorithms provided in the RooUnfold package.

with each other in most of the bins as shown in Figure 5.21. For $0.8 < |\eta^\gamma| < 1.44$ and $65 < p_T^\gamma < 72$, the bin-by-bin result does not give an unfolded result that agrees with the truth. However, as pointed out earlier, this method is presented for comparison purposes. For $2.1 < |\eta^\gamma| < 2.5$ and $65 < p_T^\gamma < 90$, the large uncertainty in the unfolding comes from low statistics in the training MC sample. Once the closure test is performed, the iterative Bayesian method is used to unfold the data distribution. The data distribution that is unfolded is obtained by using Equation 5.1 for each p_T^γ , $|\eta^\gamma|$ and $|\eta^{jet}|$ combination before dividing by the bin widths.

The systematic uncertainty on unfolding comes from the choice of MC model used to obtain the response matrix. To estimate this uncertainty, a data-driven technique is used. Since the truth

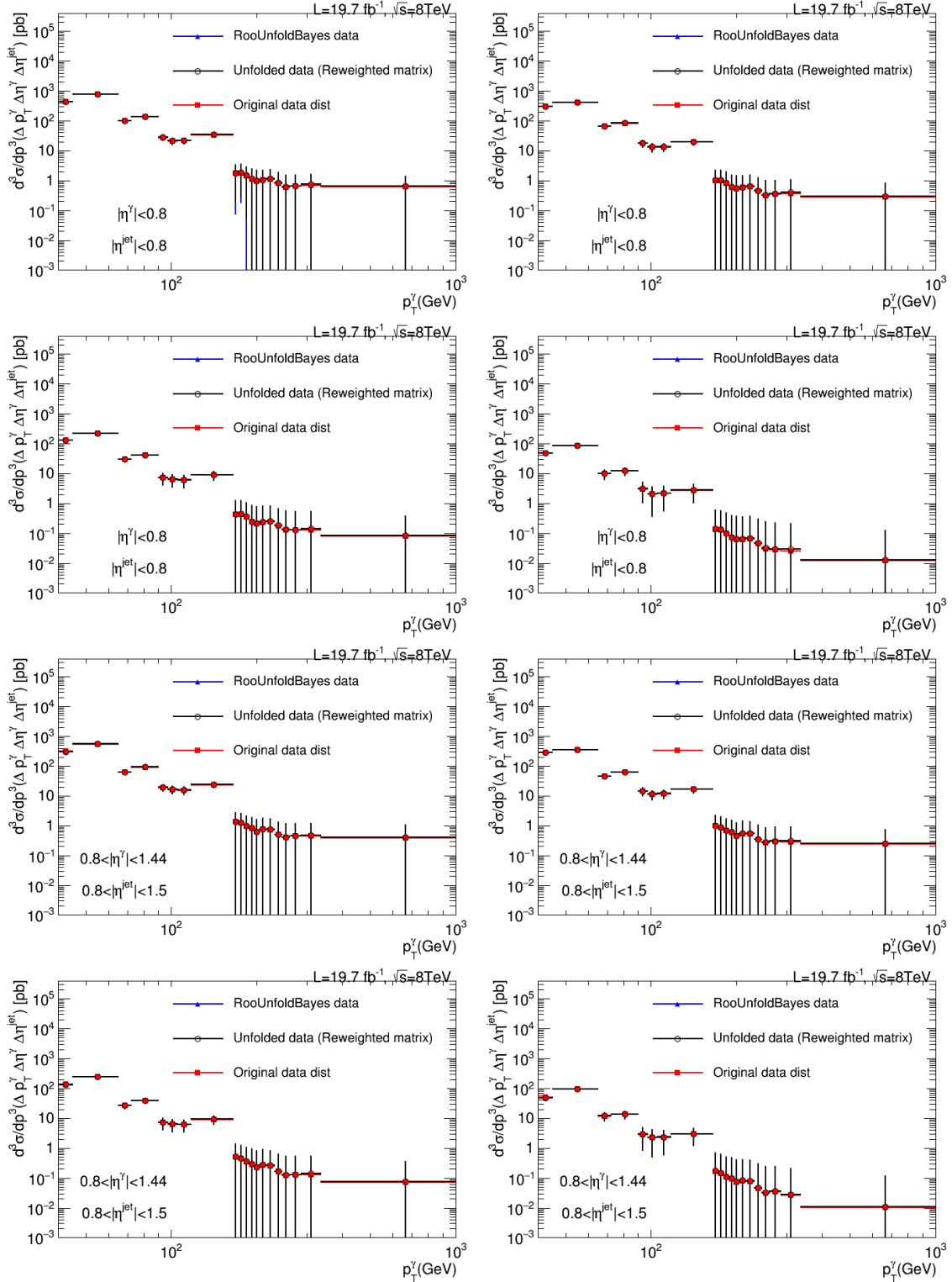


Figure 5.22: Original data distribution, the unfolded distribution, and the distribution unfolded using the dataweighted response matrix for uncertainty measurement for various bins with p_T^γ in the barrel region. The unfolded distributions do not differ much from the original distributions indicating a very small effect due to unfolding.

model is not known, the best estimate for true events is the data itself. The true MC events are weighted to match data, and the same weights are also applied to the reconstructed MC. A response matrix different from the nominal response matrix is constructed using the weighted MC. Data events are unfolded using the new response matrix. The deviation in the unfolded result from the nominal unfolded data is taken as the uncertainty due to the model dependence. The original data distribution, the unfolded distribution, and the distribution unfolded using the weighted response matrix are provided in Figures 5.22 and 5.23.

5.8 Results and Theoretical Comparison

The triple differential cross section for events with $\gamma + \text{jet}$ in the final state is evaluated from Equation 5.1 where the purity, number of candidate events, total efficiency, effective luminosity, and bin widths are used as inputs. The cross section is unfolded using the prescription described in Section 5.7. The uncertainties in purity, efficiency, and unfolding are discussed in their individual sections. To obtain these uncertainties along with the uncertainty in luminosity in the cross section, error propagation is applied to Equation 5.1. The uncertainty related to the luminosity is a flat uncertainty of 2.6%. Rewriting the equation where $X = d^3\sigma/(dp_T^\gamma d|\eta^\gamma| d|\eta^{jet}|)$ and $C = \Delta p_T^\gamma \Delta|\eta^\gamma| \Delta|\eta^{jet}|$,

$$X = \frac{1}{C} U \frac{Np}{\epsilon \mathcal{L}'}. \quad (5.17)$$

Error propagation of the above equation gives the uncertainty in the triple differential cross section.

$$dX^2 = \frac{1}{C^2} [(dU \frac{Np}{\epsilon \mathcal{L}'})^2 + (U \frac{Ndp}{\epsilon \mathcal{L}'})^2 + (U \frac{pdN}{\epsilon \mathcal{L}'})^2 + (U \frac{Npd\epsilon}{\epsilon^2 \mathcal{L}'})^2 + (U \frac{Npd\mathcal{L}'}{\epsilon \mathcal{L}'^2})^2]. \quad (5.18)$$

The triple differential cross sections for $\gamma + \text{jet}$ events are presented in this section in Figure 5.24. The values along with the statistical and systematic uncertainties are also provided in Tables 5.44, 5.45, 5.46 and 5.47 for different combinations of p_T^γ , $|\eta^\gamma|$, and $|\eta^{jet}|$. The statistical uncertainty comes from the purity estimate and from the selection of candidate sample. A breakdown of the systematic uncertainties is provided in Figures 5.25 and 5.26, which show the largest contribution to the total systematic uncertainty from the purity measurement in most of the bins, followed by the efficiency measurement, luminosity, and unfolding. The systematic uncertainties are observed

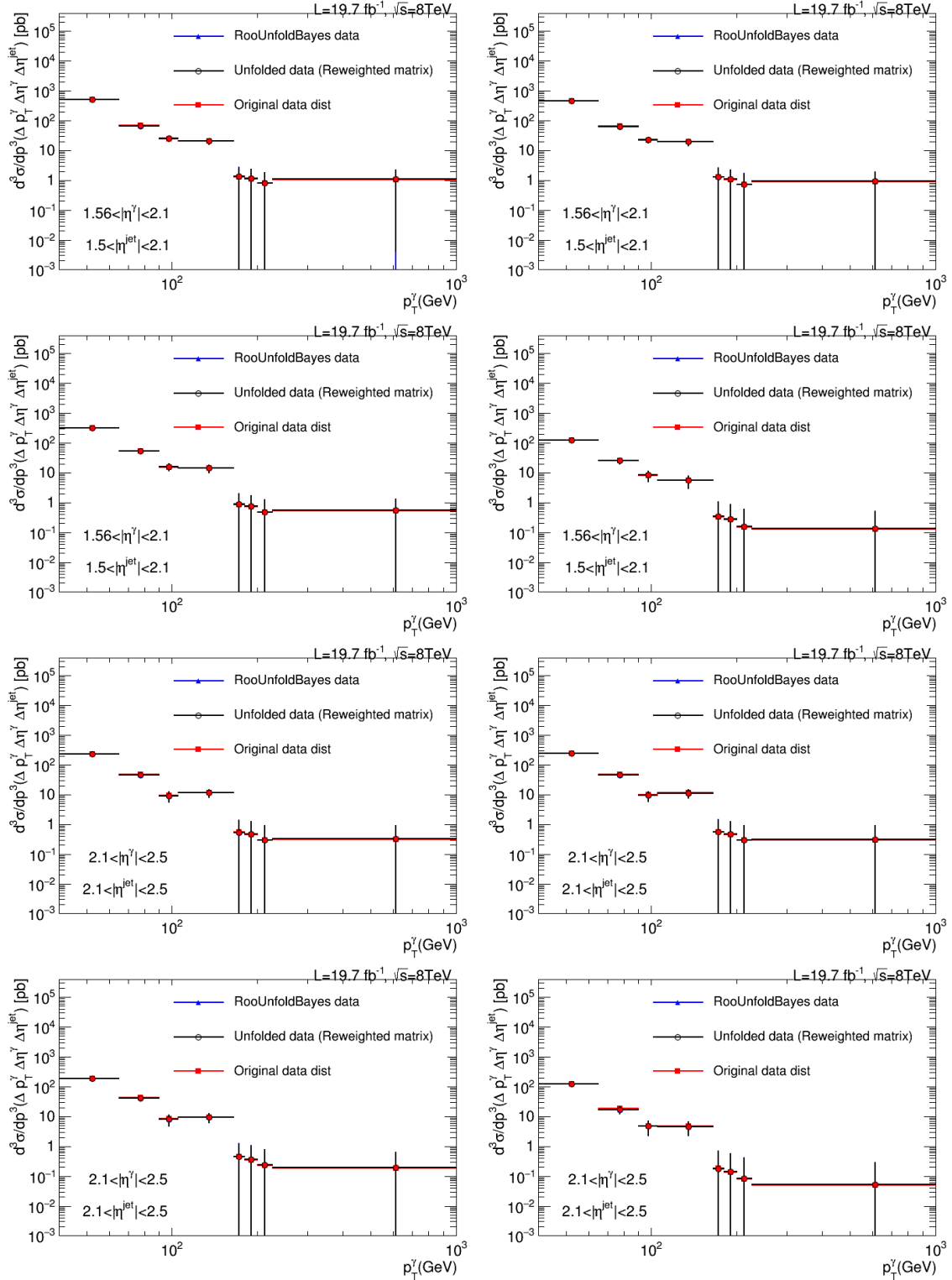


Figure 5.23: Original data distribution, the unfolded distribution, and the distribution unfolded using the dataweighted response matrix for uncertainty measurement for various bins with p_T^γ in the barrel region. The unfolded distributions do not differ much from the original distributions indicating a very small effect due to unfolding.

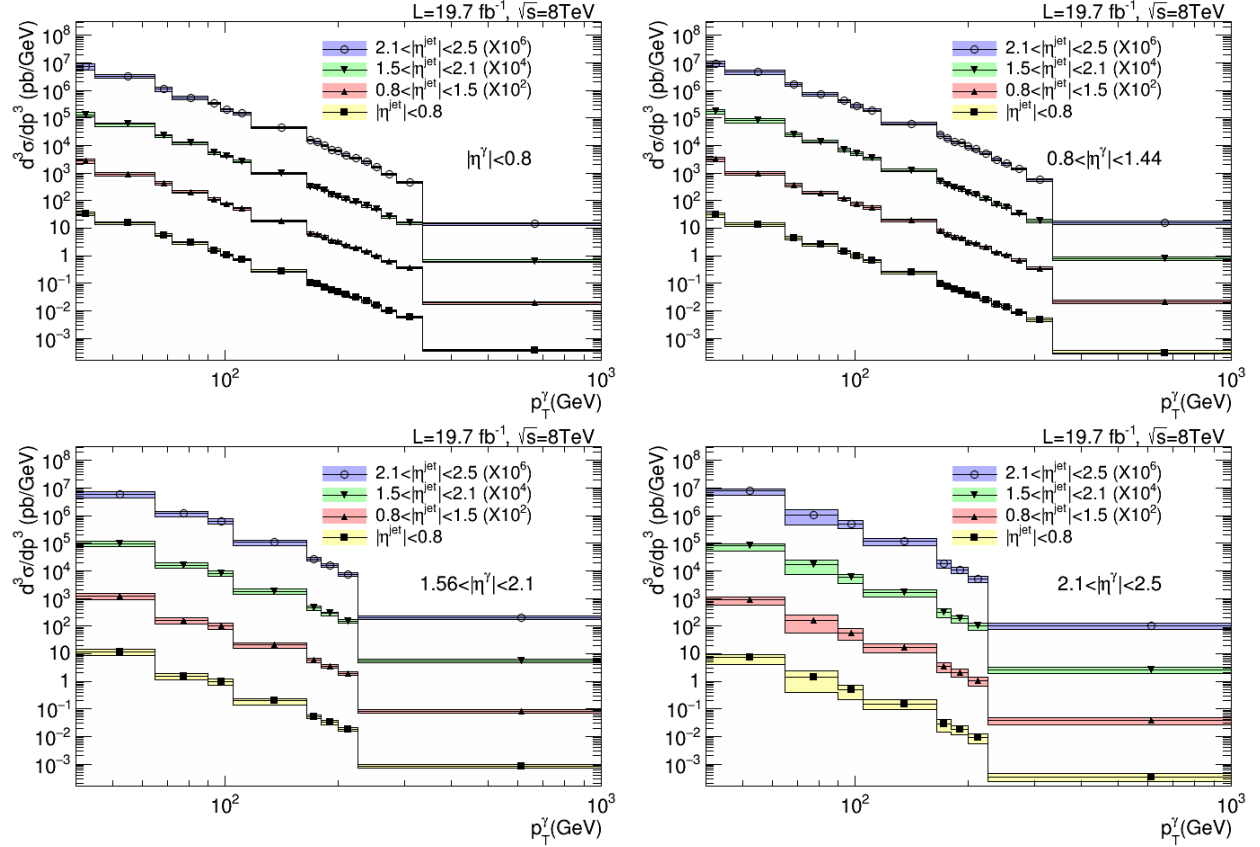


Figure 5.24: Measured triple differential cross section values along with the total uncertainty for different bins in p_T^γ , $|\eta^\gamma|$, and $|\eta^{jet}|$. Notice that the distributions are multiplied by a factor of 100, 10,000 and 1,000,000 for $0.8 < |\eta^{jet}| < 1.5$, $1.5 < |\eta^{jet}| < 2.1$, and $2.1 < |\eta^{jet}| < 2.5$, respectively. The systematic uncertainties are represented by colored bands around the central value, whereas the statistical uncertainties are represented by error bars that are not visible due to the extremely large vertical axis range.

to be larger for lower p_T^γ bins and larger values of $|\eta^\gamma|$.

Comparison with the theory predictions for the $\gamma + \text{jet}$ process is performed at next-to-leading order (NLO) in QCD using the GamJet package [39, 40]. The package uses the phase space slicing method which separates the phase space containing soft and collinear divergences from the rest of the non-divergent region, and absorbs them into non-perturbative quantities [41]. The theory calculation is made with the most recent set of CTEQ-JLab Collaboration Parton Distribution Functions called CJ15 [42]. The central values of the renormalization (μ_R), fragmentation (μ_F),

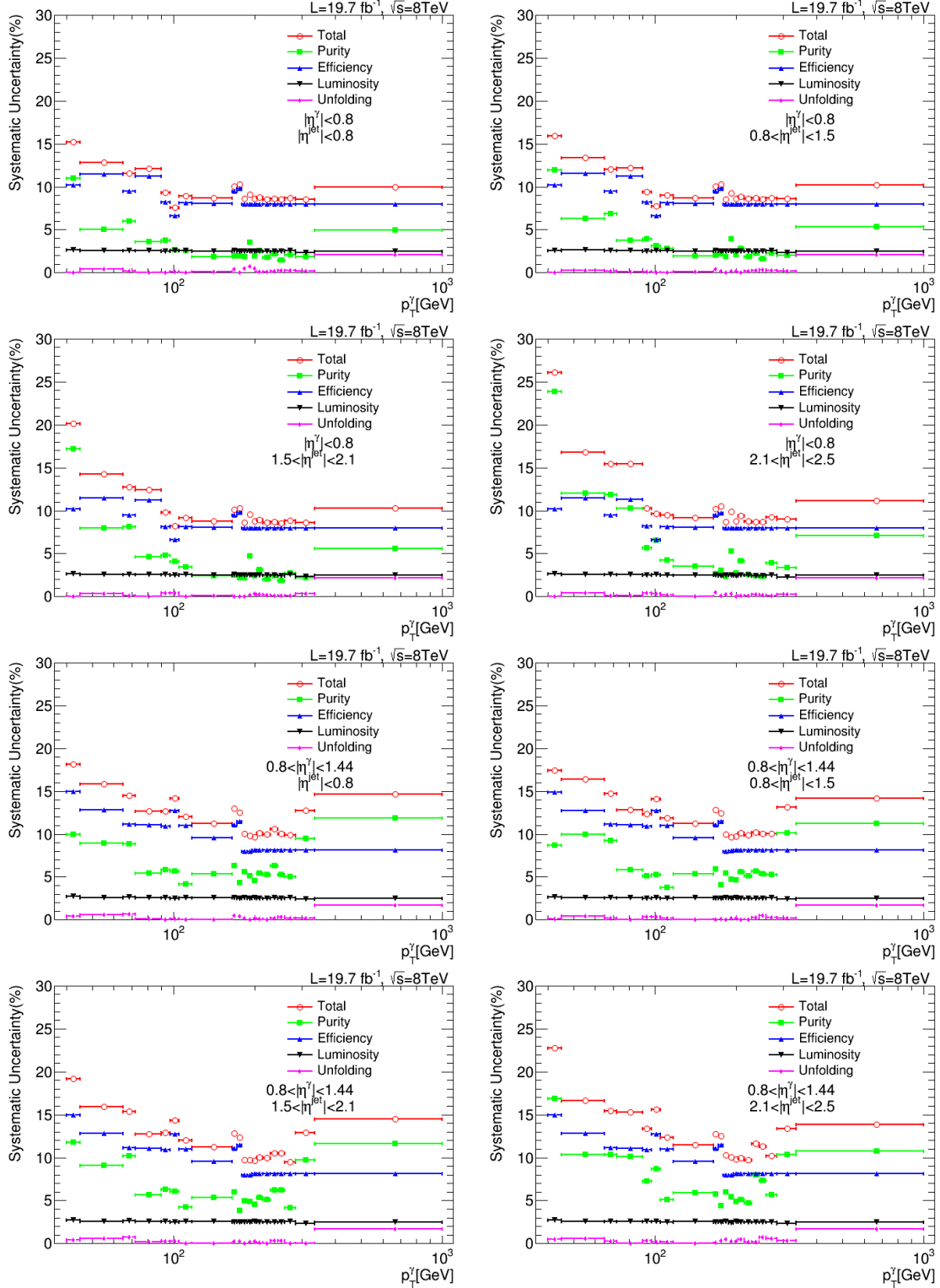


Figure 5.25: Breakdown of the systematic uncertainties on the cross section for different bins in p_T^γ , $|\eta^\gamma|$, and $|\eta^{jet}|$. The different colors/markers correspond to contributions from the purity measurement, efficiency measurement, luminosity measurement, unfolding, and the total uncertainty.

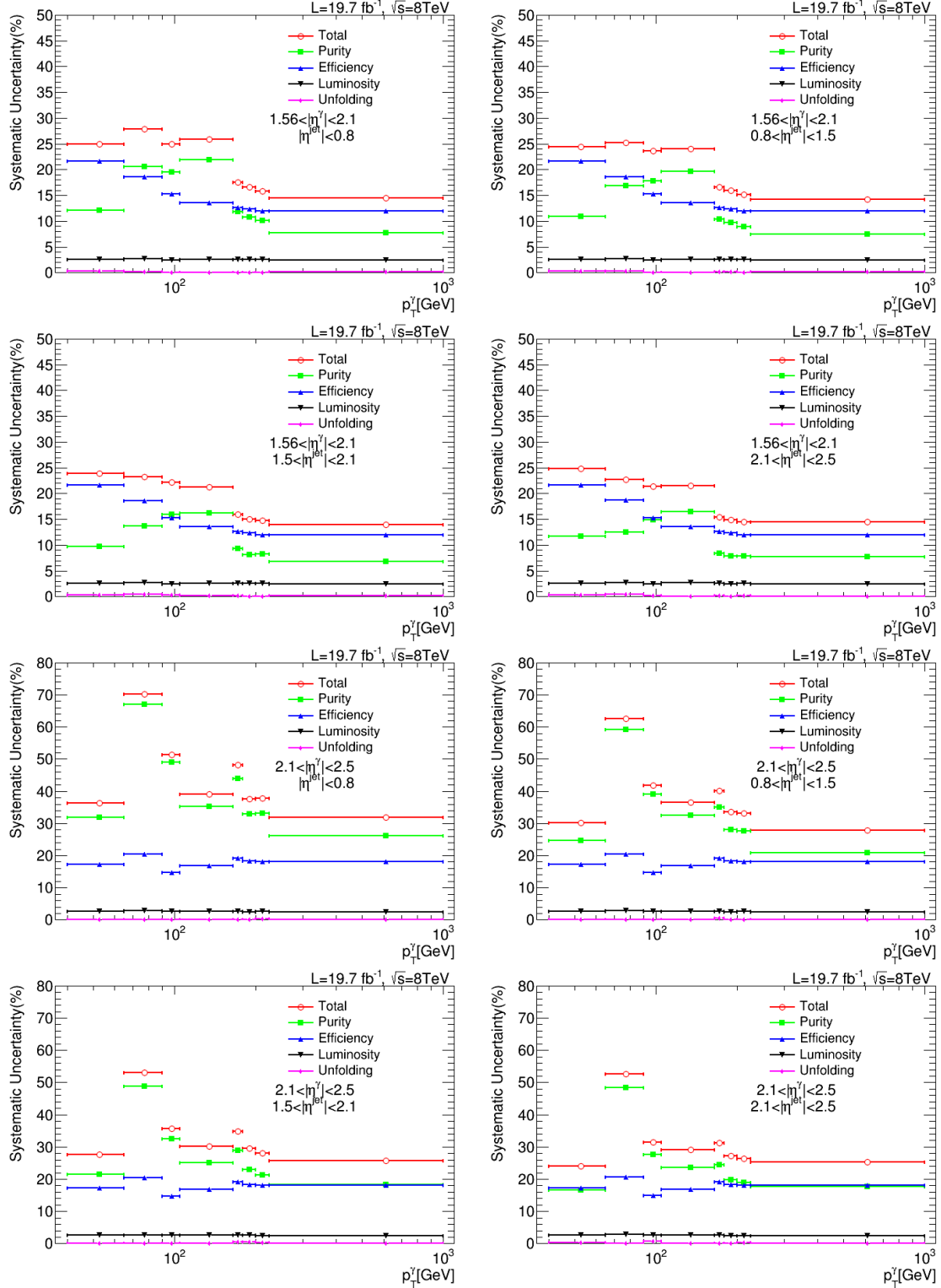


Figure 5.26: Breakdown of the systematic uncertainties on the cross section for different bins in p_T^γ , $|\eta^\gamma|$, and $|\eta^{jet}|$. The different colors/markers correspond to contributions from the purity measurement, efficiency measurement, luminosity measurement, unfolding, and the total uncertainty.

and PDF scales (μ_{PDF}) are set to p_T^γ . The following requirements are used to generate events:

- $40 < p_T^\gamma < 1000$ GeV
- $|\eta^{\gamma/\text{jet}}| < 2.5$
- $1.4442 < |\eta^\gamma| < 1.556$ excluded
- $p_T^{\text{jet}} > 25$ GeV
- $\Delta R(\gamma, \text{jet}) > 0.5$
- Isolation in a cone of radius $0.2 < 4.0$ GeV.

The above cuts are consistent with the cuts applied to the data or generator level MC signal in the previous sections. In total, approximately 30 million events are generated with a p_T^γ bin width of 1 GeV in the 40–500 GeV range and 5 GeV in the 500–1000 GeV range. The bins are then combined in the same bins as in the data measurement. The theory predictions for the $\gamma + \text{jet}$ differential cross section are shown in Figures 5.27 and 5.28. In order to determine the uncertainties in the theoretical prediction due to the PDF choice, the default value of the CJ15 PDF is set as the central value. There is a total of 49 PDFs, one central PDF set and 24 sets for uncertainty measurement. The PDF uncertainty $\delta\sigma$ is then estimated by generating events with the 24 PDF sets and using the following expression,

$$\delta\sigma = \frac{T}{2} \sqrt{\sum_{i=1}^{24} [\sigma(a_{2i-1}) - \sigma(a_{2i})]^2}. \quad (5.19)$$

Here, $\sigma(a_{2i})$ represents the differential cross section values as estimated by using the $2i^{\text{th}}$ PDF and $\sigma(a_{2i-1})$ represents the differential cross section values as estimated by using $2i-1^{\text{th}}$ set, and T is a tolerance factor. A tolerance factor of 1 assumes that all of the datasets used in the PDF calculation are statistically compatible and the experimental uncertainties are Gaussian. $T=1$ is used for the theory predictions presented in this analysis. This choice gives a thinnest uncertainty band for the theory prediction.

To estimate the uncertainty due to scale choice, μ_R and the PDF scale are varied and an asymmetric uncertainty for each bin is estimated based on the cross section value that is farthest from the central value. μ_F is not varied because the contribution from fragmentation photons is largely suppressed as a result of the isolation criterion that is applied. The following choices in the scales

are made for uncertainty estimation:

- $\mu_R = 0.5 p_T^\gamma, \mu_{PDF} = 0.5 p_T^\gamma$
- $\mu_R = 0.5 p_T^\gamma, \mu_{PDF} = p_T^\gamma$
- $\mu_R = p_T^\gamma, \mu_{PDF} = 0.5 p_T^\gamma$
- $\mu_R = p_T^\gamma, \mu_{PDF} = 2.0 p_T^\gamma$
- $\mu_R = 2.0 p_T^\gamma, \mu_{PDF} = p_T^\gamma$
- $\mu_R = 2.0 p_T^\gamma, \mu_{PDF} = 2.0 p_T^\gamma$

The theory prediction along with the uncertainties due to scale variation and PDF for various bins are provided in Figures 5.27 and 5.28.

Ratios of the data and theory triple differential cross section are given in Figures 5.29 and 5.30 for all bin combinations. Among the 224 bins in which the comparisons are made, the results are mostly in agreement within the estimated uncertainties. However, disagreement in data and theory predictions is seen for some large p_T^γ bins in $|\eta^\gamma| < 0.8$ and $0.8 < |\eta^{jet}| < 1.44$ or $1.56 < |\eta^{jet}| < 2.1$. Referring to Equation 5.19, the PDF uncertainty predicted by the theory uses a tolerance factor of 1.0. Though it is a reasonable choice to make, in the literature it is seen that a tolerance factor as high as 10 is also used. A slightly larger tolerance factor would reduce the data-theory disagreement in these regions; hence, attributing the current theory-measurement disagreement to a lack of proper understanding of the PDF uncertainty in these bins. The tightest theory uncertainty bands are approximately of the same order or larger as compared to the uncertainties in experimental result. Therefore, these experimental results can be utilized to gain better understanding of the PDFs in wide range of photon-jet kinematics.

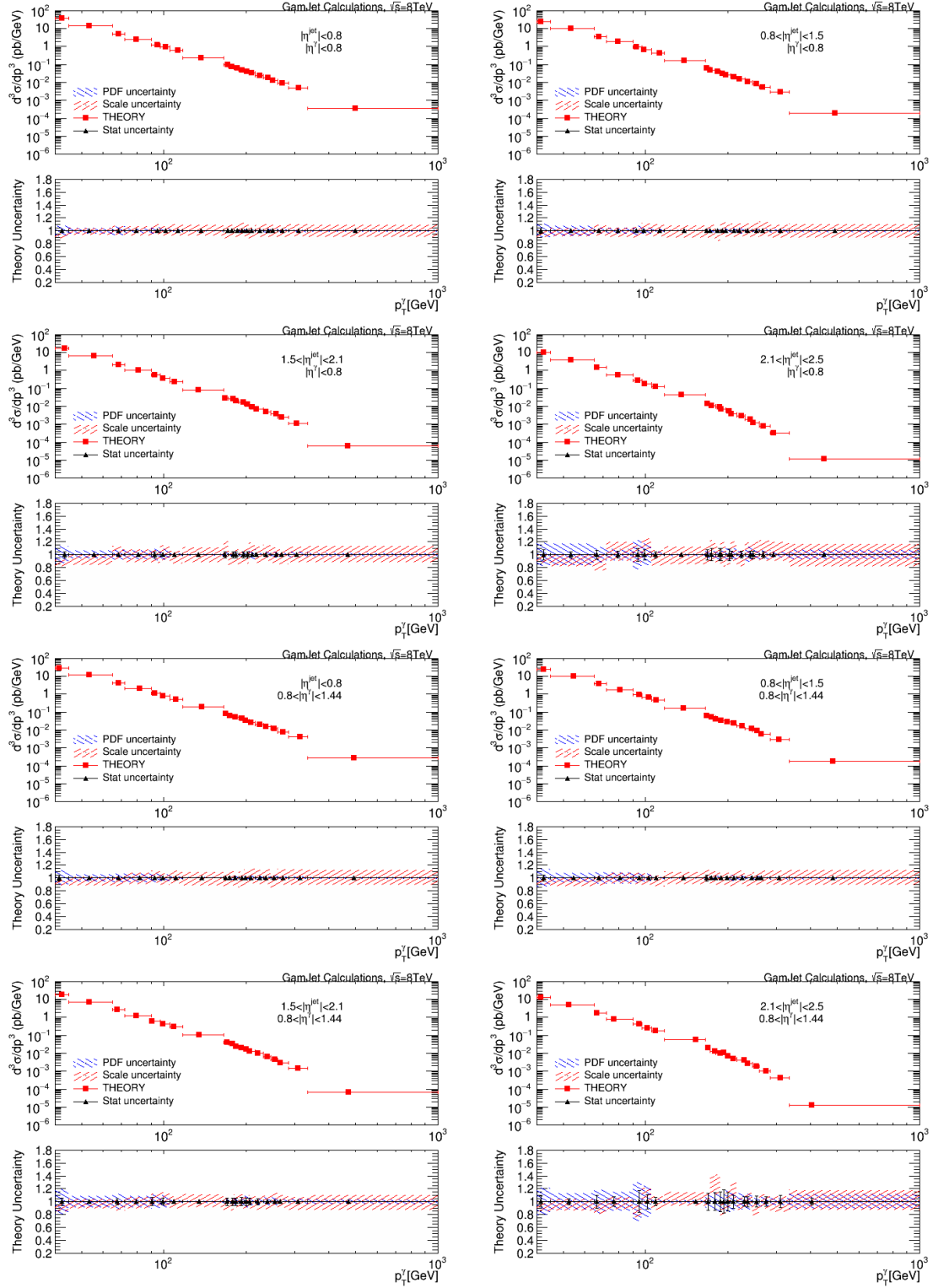


Figure 5.27: Theory prediction of γ +jet triple differential cross section using GamJet package at NLO.

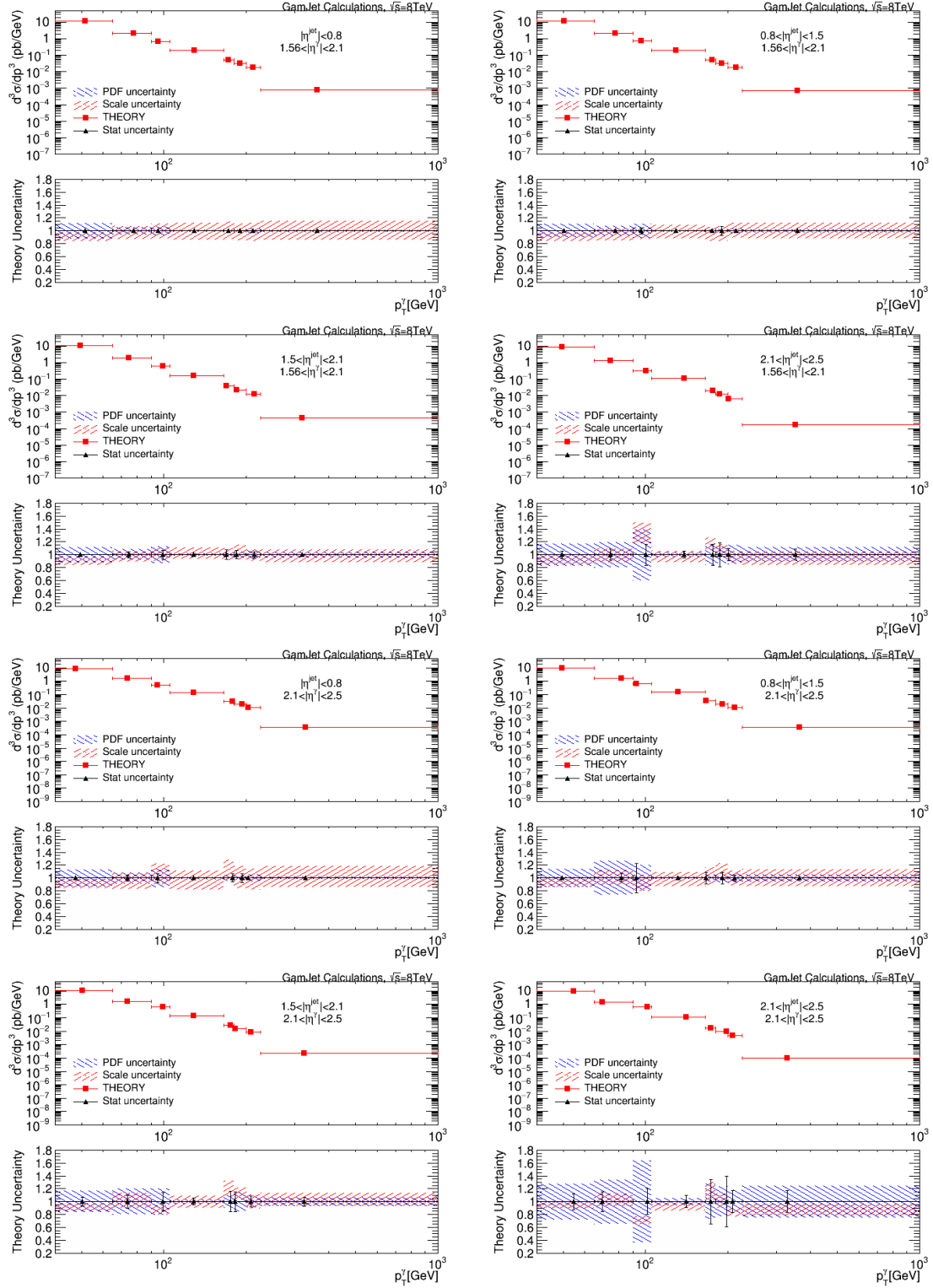


Figure 5.28: Theory prediction of γ +jet triple differential cross section using GamJet package at NLO.

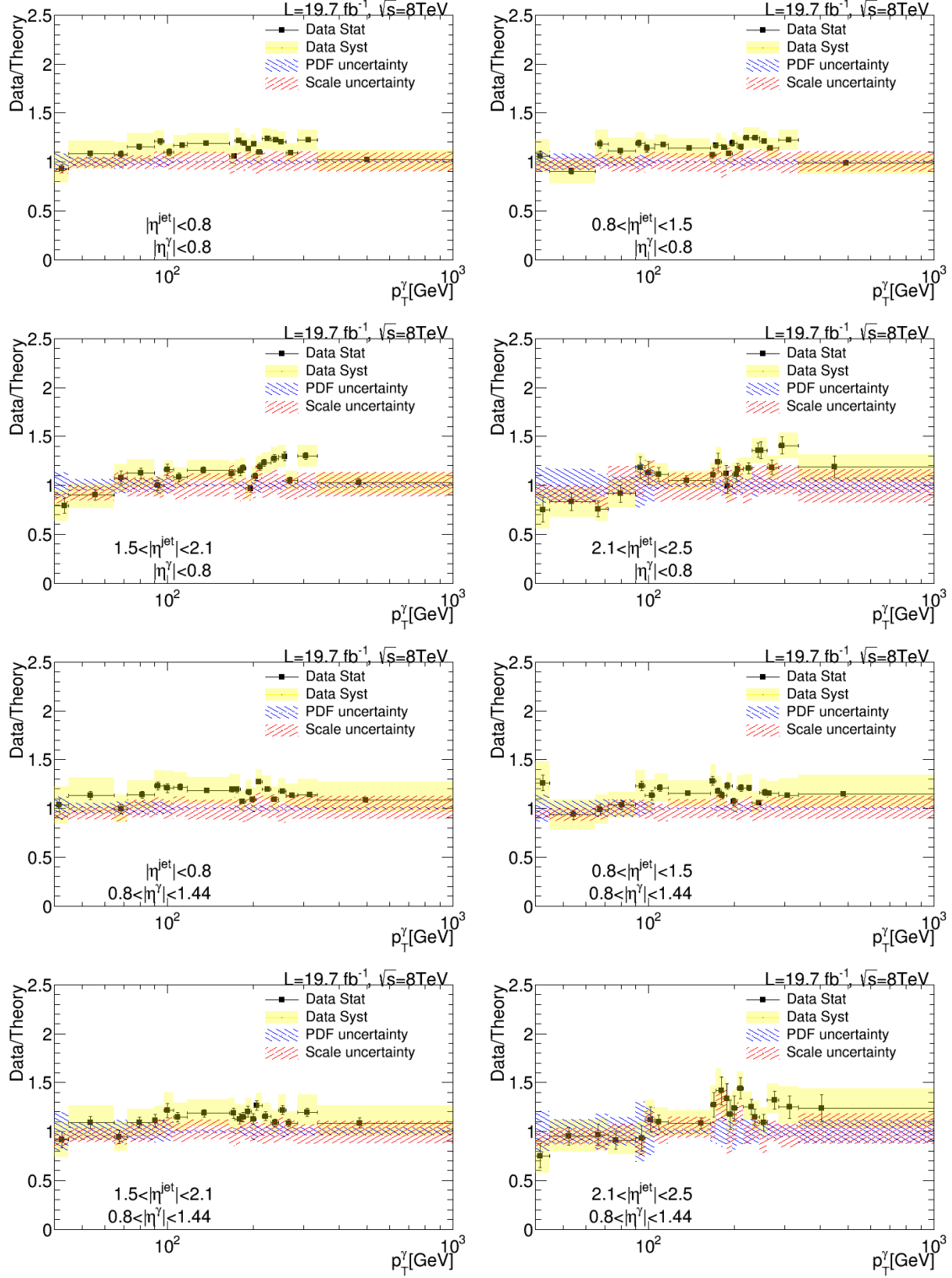


Figure 5.29: Ratios of measured triple differential cross section to the theory prediction using GamJet at NLO.

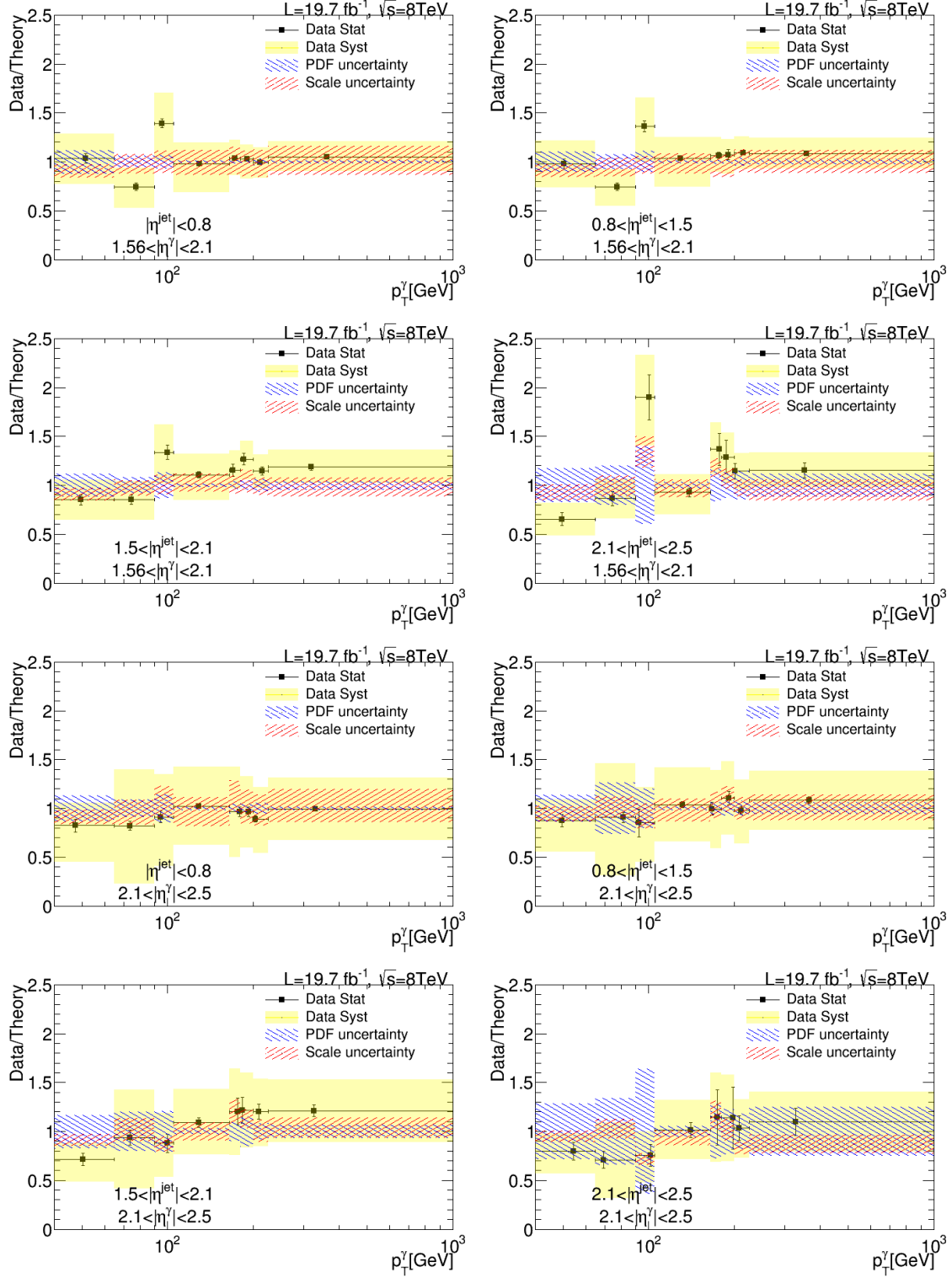


Figure 5.30: Ratios of measured triple differential cross section to the theory prediction using GamJet at NLO.

Table 5.44: Triple differential cross section values with the statistical and systematic (lower and higher) uncertainties for various p_T^γ regions and $|\eta^\gamma| < 0.8$.

p_T^γ (GeV)	$ \eta^\gamma < 0.8$	$0.8 < \eta^\gamma < 1.5$	$d^3\sigma/(dp_T^\gamma d \eta^\gamma d \eta^{jet})(\text{pb/GeV})$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40-45	$3.42\text{E}+01 \pm 1.39\text{E}+00$ $+1.26\text{E}+00$ $-1.16\text{E}+00$	$2.69\text{E}+01 \pm 1.16\text{E}+00$ $+4.33\text{E}+00$ $-1.13\text{E}+00$	$1.34\text{E}+01 \pm 1.24\text{E}+00$ $+2.74\text{E}+00$ $-1.02\text{E}+00$	$7.45\text{E}+00 \pm 1.18\text{E}+00$ $+1.97\text{E}+00$ $-1.09\text{E}+00$	$7.45\text{E}+00 \pm 1.18\text{E}+00$ $+1.97\text{E}+00$ $-1.09\text{E}+00$
45-65	$1.52\text{E}+01 \pm 4.10\text{E}-01$ $+2.06\text{E}+00$ $-2.06\text{E}+00$	$9.13\text{E}+00 \pm 3.04\text{E}-01$ $+1.31\text{E}+00$ $-1.31\text{E}+00$	$5.80\text{E}+00 \pm 3.35\text{E}-01$ $+8.94\text{E}-01$ $-8.94\text{E}-01$	$3.33\text{E}+00 \pm 3.43\text{E}-01$ $+6.17\text{E}-01$ $-6.17\text{E}-01$	$3.33\text{E}+00 \pm 3.43\text{E}-01$ $+6.17\text{E}-01$ $-6.17\text{E}-01$
65-72	$5.52\text{E}+00 \pm 1.46\text{E}-01$ $+6.27\text{E}-01$ $-6.27\text{E}-01$	$4.27\text{E}+00 \pm 1.16\text{E}-01$ $+5.04\text{E}-01$ $-5.04\text{E}-01$	$2.24\text{E}+00 \pm 1.18\text{E}-01$ $+2.83\text{E}-01$ $-2.83\text{E}-01$	$1.11\text{E}+00 \pm 1.07\text{E}-01$ $+3.10\text{E}-01$ $-3.10\text{E}-01$	$1.11\text{E}+00 \pm 1.07\text{E}-01$ $+3.10\text{E}-01$ $-3.10\text{E}-01$
72-90	$2.97\text{E}+00 \pm 6.55\text{E}-02$ $+3.61\text{E}-01$ $-3.61\text{E}-01$	$2.07\text{E}+00 \pm 4.85\text{E}-02$ $+2.54\text{E}-01$ $-2.54\text{E}-01$	$1.21\text{E}+00 \pm 5.16\text{E}-02$ $+5.17\text{E}-02$ $-5.17\text{E}-02$	$5.23\text{E}-01 \pm 4.56\text{E}-02$ $+8.07\text{E}-02$ $-8.07\text{E}-02$	$5.23\text{E}-01 \pm 4.56\text{E}-02$ $+8.07\text{E}-02$ $-8.07\text{E}-02$
90-97	$1.57\text{E}+00 \pm 2.96\text{E}-02$ $+1.43\text{E}-01$ $-1.43\text{E}-01$	$1.14\text{E}+00 \pm 2.22\text{E}-02$ $+1.13\text{E}-01$ $-1.13\text{E}-01$	$5.45\text{E}-01 \pm 2.06\text{E}-02$ $+5.59\text{E}-02$ $-5.59\text{E}-02$	$3.45\text{E}-01 \pm 2.05\text{E}-02$ $+3.74\text{E}-02$ $-3.74\text{E}-02$	$3.45\text{E}-01 \pm 2.05\text{E}-02$ $+3.74\text{E}-02$ $-3.74\text{E}-02$
97-105	$1.05\text{E}+00 \pm 2.47\text{E}-02$ $+1.78\text{E}-02$ $-1.78\text{E}-02$	$7.42\text{E}-01 \pm 1.85\text{E}-02$ $+4.38\text{E}-02$ $-4.38\text{E}-02$	$4.16\text{E}-01 \pm 1.86\text{E}-02$ $+2.41\text{E}-02$ $-2.41\text{E}-02$	$2.00\text{E}-01 \pm 1.72\text{E}-02$ $+1.86\text{E}-02$ $-1.86\text{E}-02$	$2.00\text{E}-01 \pm 1.72\text{E}-02$ $+1.86\text{E}-02$ $-1.86\text{E}-02$
105-117	$7.20\text{E}-01 \pm 1.03\text{E}-02$ $+6.45\text{E}-02$ $-6.45\text{E}-02$	$5.00\text{E}-01 \pm 7.42\text{E}-03$ $+4.51\text{E}-02$ $-4.51\text{E}-02$	$2.63\text{E}-01 \pm 7.79\text{E}-03$ $+2.43\text{E}-02$ $-2.43\text{E}-02$	$1.44\text{E}-01 \pm 7.22\text{E}-03$ $+1.38\text{E}-02$ $-1.38\text{E}-02$	$1.44\text{E}-01 \pm 7.22\text{E}-03$ $+1.38\text{E}-02$ $-1.38\text{E}-02$
117-165	$2.79\text{E}-01 \pm 3.25\text{E}-03$ $+2.41\text{E}-02$ $-2.41\text{E}-02$	$1.85\text{E}-01 \pm 2.24\text{E}-03$ $+1.59\text{E}-02$ $-1.59\text{E}-02$	$9.79\text{E}-02 \pm 2.39\text{E}-03$ $+8.56\text{E}-03$ $-8.56\text{E}-03$	$4.55\text{E}-02 \pm 2.04\text{E}-03$ $+4.14\text{E}-03$ $-4.14\text{E}-03$	$4.55\text{E}-02 \pm 2.04\text{E}-03$ $+4.14\text{E}-03$ $-4.14\text{E}-03$
165-172	$1.01\text{E}-01 \pm 6.90\text{E}-04$ $+1.02\text{E}-02$ $-1.02\text{E}-02$	$6.55\text{E}-02 \pm 4.69\text{E}-04$ $+6.05\text{E}-03$ $-6.05\text{E}-03$	$3.20\text{E}-02 \pm 5.05\text{E}-04$ $+3.28\text{E}-03$ $-3.28\text{E}-03$	$1.56\text{E}-02 \pm 4.42\text{E}-04$ $+1.39\text{E}-03$ $-1.39\text{E}-03$	$1.56\text{E}-02 \pm 4.42\text{E}-04$ $+1.39\text{E}-03$ $-1.39\text{E}-03$
172-180	$9.19\text{E}-02 \pm 7.34\text{E}-04$ $+9.45\text{E}-03$ $-9.45\text{E}-03$	$5.87\text{E}-02 \pm 4.86\text{E}-04$ $+6.07\text{E}-03$ $-6.07\text{E}-03$	$2.97\text{E}-02 \pm 4.65\text{E}-04$ $+3.07\text{E}-03$ $-3.07\text{E}-03$	$1.33\text{E}-02 \pm 3.73\text{E}-04$ $+1.40\text{E}-03$ $-1.40\text{E}-03$	$1.33\text{E}-02 \pm 3.73\text{E}-04$ $+1.40\text{E}-03$ $-1.40\text{E}-03$
180-188	$7.12\text{E}-02 \pm 4.62\text{E}-04$ $+6.07\text{E}-03$ $-6.07\text{E}-03$	$4.78\text{E}-02 \pm 3.17\text{E}-04$ $+4.03\text{E}-03$ $-4.03\text{E}-03$	$2.33\text{E}-02 \pm 3.19\text{E}-04$ $+1.98\text{E}-03$ $-1.98\text{E}-03$	$9.77\text{E}-03 \pm 2.59\text{E}-04$ $+8.36\text{E}-04$ $-8.36\text{E}-04$	$9.77\text{E}-03 \pm 2.59\text{E}-04$ $+8.36\text{E}-04$ $-8.36\text{E}-04$
188-196	$5.59\text{E}-02 \pm 4.51\text{E}-04$ $+5.85\text{E}-03$ $-5.85\text{E}-03$	$3.43\text{E}-02 \pm 2.93\text{E}-04$ $+3.43\text{E}-03$ $-3.43\text{E}-03$	$1.59\text{E}-02 \pm 3.06\text{E}-04$ $+1.68\text{E}-03$ $-1.68\text{E}-03$	$7.04\text{E}-03 \pm 2.64\text{E}-04$ $+6.40\text{E}-04$ $-6.40\text{E}-04$	$7.04\text{E}-03 \pm 2.64\text{E}-04$ $+6.40\text{E}-04$ $-6.40\text{E}-04$
196-204	$4.78\text{E}-02 \pm 4.52\text{E}-04$ $+4.44\text{E}-03$ $-4.44\text{E}-03$	$3.13\text{E}-02 \pm 3.04\text{E}-04$ $+2.68\text{E}-03$ $-2.68\text{E}-03$	$1.41\text{E}-02 \pm 2.71\text{E}-04$ $+1.22\text{E}-03$ $-1.22\text{E}-03$	$6.29\text{E}-03 \pm 2.28\text{E}-04$ $+5.47\text{E}-04$ $-5.47\text{E}-04$	$6.29\text{E}-03 \pm 2.28\text{E}-04$ $+5.47\text{E}-04$ $-5.47\text{E}-04$
204-215	$3.73\text{E}-02 \pm 3.02\text{E}-04$ $+3.37\text{E}-03$ $-3.37\text{E}-03$	$2.39\text{E}-02 \pm 1.99\text{E}-04$ $+2.74\text{E}-03$ $-2.74\text{E}-03$	$1.13\text{E}-02 \pm 1.94\text{E}-04$ $+1.02\text{E}-03$ $-1.02\text{E}-03$	$4.46\text{E}-03 \pm 1.60\text{E}-04$ $+4.21\text{E}-04$ $-4.21\text{E}-04$	$4.46\text{E}-03 \pm 1.60\text{E}-04$ $+4.21\text{E}-04$ $-4.21\text{E}-04$
215-230	$3.05\text{E}-02 \pm 2.24\text{E}-04$ $+2.60\text{E}-03$ $-2.60\text{E}-03$	$1.92\text{E}-02 \pm 1.44\text{E}-04$ $+1.66\text{E}-03$ $-1.66\text{E}-03$	$8.87\text{E}-03 \pm 1.48\text{E}-04$ $+7.71\text{E}-04$ $-7.71\text{E}-04$	$3.57\text{E}-03 \pm 1.07\text{E}-04$ $+3.16\text{E}-04$ $-3.16\text{E}-04$	$3.57\text{E}-03 \pm 1.07\text{E}-04$ $+3.16\text{E}-04$ $-3.16\text{E}-04$
230-245	$2.16\text{E}-02 \pm 1.81\text{E}-04$ $+2.64\text{E}-03$ $-2.64\text{E}-03$	$1.37\text{E}-02 \pm 1.17\text{E}-04$ $+1.18\text{E}-03$ $-1.18\text{E}-03$	$6.26\text{E}-03 \pm 1.18\text{E}-04$ $+5.37\text{E}-04$ $-5.37\text{E}-04$	$2.47\text{E}-03 \pm 9.32\text{E}-05$ $+2.12\text{E}-04$ $-2.12\text{E}-04$	$2.47\text{E}-03 \pm 9.32\text{E}-05$ $+2.12\text{E}-04$ $-2.12\text{E}-04$
245-260	$1.59\text{E}-02 \pm 1.75\text{E}-04$ $+1.35\text{E}-03$ $-1.35\text{E}-03$	$9.83\text{E}-03 \pm 1.10\text{E}-04$ $+8.44\text{E}-04$ $-8.44\text{E}-04$	$4.75\text{E}-03 \pm 1.07\text{E}-04$ $+4.05\text{E}-04$ $-4.05\text{E}-04$	$1.67\text{E}-03 \pm 7.94\text{E}-05$ $+1.46\text{E}-04$ $-1.46\text{E}-04$	$1.67\text{E}-03 \pm 7.94\text{E}-05$ $+1.46\text{E}-04$ $-1.46\text{E}-04$
260-285	$1.02\text{E}-02 \pm 1.02\text{E}-04$ $+8.78\text{E}-04$ $-8.78\text{E}-04$	$6.35\text{E}-03 \pm 6.46\text{E}-05$ $+5.02\text{E}-04$ $-5.02\text{E}-04$	$2.63\text{E}-03 \pm 6.10\text{E}-05$ $+2.29\text{E}-04$ $-2.29\text{E}-04$	$9.18\text{E}-04 \pm 4.63\text{E}-05$ $+8.69\text{E}-05$ $-8.69\text{E}-05$	$9.18\text{E}-04 \pm 4.63\text{E}-05$ $+8.69\text{E}-05$ $-8.69\text{E}-05$
285-335	$5.82\text{E}-03 \pm 5.81\text{E}-05$ $+3.06\text{E}-04$ $-3.06\text{E}-04$	$3.56\text{E}-03 \pm 3.66\text{E}-05$ $+1.84\text{E}-04$ $-1.84\text{E}-04$	$1.49\text{E}-03 \pm 3.64\text{E}-05$ $+1.39\text{E}-04$ $-1.39\text{E}-04$	$4.50\text{E}-04 \pm 2.47\text{E}-05$ $+1.51\text{E}-05$ $-1.51\text{E}-05$	$4.50\text{E}-04 \pm 2.47\text{E}-05$ $+1.51\text{E}-05$ $-1.51\text{E}-05$
335-1000	$3.67\text{E}-04 \pm 4.30\text{E}-06$ $+4.02\text{E}-05$ $-4.02\text{E}-05$	$1.94\text{E}-04 \pm 2.39\text{E}-06$ $+2.13\text{E}-05$ $-2.13\text{E}-05$	$6.34\text{E}-05 \pm 2.42\text{E}-06$ $+7.26\text{E}-06$ $-7.26\text{E}-06$	$1.45\text{E}-05 \pm 1.32\text{E}-06$ $+1.75\text{E}-06$ $-1.75\text{E}-06$	$1.45\text{E}-05 \pm 1.32\text{E}-06$ $+1.75\text{E}-06$ $-1.75\text{E}-06$

Table 5.45: Triple differential cross section values with the statistical and systematic (lower and higher) uncertainties for various p_T^γ regions and $|\eta^\gamma| < 1.44$.

p_T^γ (GeV)	$ \eta^\gamma < 0.8$	$0.8 < \eta^\gamma < 1.5$	$d^3\sigma/(dp_T^\gamma d \eta^\gamma d \eta^{jet})(\text{pb/GeV})$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40-45	$2.93\text{E}+01 \pm 1.61\text{E}+00 \pm 1.27\text{E}+00$	$3.09\text{E}+01 \pm 1.65\text{E}+00 \pm 3.37\text{E}+00$	$1.72\text{E}+01 \pm 1.44\text{E}+00 \pm 3.31\text{E}+00$	$9.45\text{E}+00 \pm 1.47\text{E}+00 \pm 2.12\text{E}+00$	$9.45\text{E}+00 \pm 1.47\text{E}+00 \pm 2.12\text{E}+00$
45-65	$1.35\text{E}+01 \pm 4.89\text{E}-01 \pm 2.13\text{E}+00$	$9.81\text{E}+00 \pm 4.62\text{E}-01 \pm 1.61\text{E}+00$	$7.97\text{E}+00 \pm 4.30\text{E}-01 \pm 1.37\text{E}+00$	$4.67\text{E}+00 \pm 4.24\text{E}-01 \pm 7.85\text{E}-01$	$4.67\text{E}+00 \pm 4.24\text{E}-01 \pm 7.85\text{E}-01$
65-72	$4.34\text{E}+00 \pm 1.78\text{E}-01 \pm 6.42\text{E}-01$	$3.66\text{E}+00 \pm 1.72\text{E}-01 \pm 5.51\text{E}-01$	$2.49\text{E}+00 \pm 1.57\text{E}-01 \pm 3.94\text{E}-01$	$1.67\text{E}+00 \pm 1.49\text{E}-01 \pm 2.15\text{E}-01$	$1.67\text{E}+00 \pm 1.49\text{E}-01 \pm 2.15\text{E}-01$
72-90	$2.50\text{E}+00 \pm 7.51\text{E}-02 \pm 6.17\text{E}-01$	$1.91\text{E}+00 \pm 7.38\text{E}-02 \pm 5.27\text{E}-01$	$1.40\text{E}+00 \pm 6.30\text{E}-02 \pm 3.70\text{E}-01$	$7.34\text{E}-01 \pm 6.35\text{E}-02 \pm 1.11\text{E}-01$	$7.34\text{E}-01 \pm 6.35\text{E}-02 \pm 1.11\text{E}-01$
90-97	$1.34\text{E}+00 \pm 3.36\text{E}-02 \pm 1.72\text{E}-01$	$1.16\text{E}+00 \pm 3.20\text{E}-02 \pm 1.40\text{E}-01$	$6.80\text{E}-01 \pm 2.74\text{E}-02 \pm 8.72\text{E}-02$	$4.10\text{E}-01 \pm 2.60\text{E}-02 \pm 5.69\text{E}-02$	$4.10\text{E}-01 \pm 2.60\text{E}-02 \pm 5.69\text{E}-02$
97-105	$1.00\text{E}+00 \pm 2.90\text{E}-02 \pm 1.43\text{E}-01$	$7.90\text{E}-01 \pm 2.73\text{E}-02 \pm 1.12\text{E}-01$	$5.31\text{E}-01 \pm 2.44\text{E}-02 \pm 7.67\text{E}-02$	$2.84\text{E}-01 \pm 2.31\text{E}-02 \pm 4.55\text{E}-02$	$2.84\text{E}-01 \pm 2.31\text{E}-02 \pm 4.55\text{E}-02$
105-117	$6.37\text{E}-01 \pm 1.19\text{E}-02 \pm 7.76\text{E}-02$	$5.58\text{E}-01 \pm 1.14\text{E}-02 \pm 6.70\text{E}-02$	$3.40\text{E}-01 \pm 9.99\text{E}-03 \pm 4.13\text{E}-02$	$1.90\text{E}-01 \pm 9.44\text{E}-03 \pm 2.39\text{E}-02$	$1.90\text{E}-01 \pm 9.44\text{E}-03 \pm 2.39\text{E}-02$
117-165	$2.43\text{E}-01 \pm 3.87\text{E}-03 \pm 2.82\text{E}-02$	$1.94\text{E}-01 \pm 3.56\text{E}-03 \pm 2.25\text{E}-02$	$1.25\text{E}-01 \pm 3.17\text{E}-03 \pm 1.45\text{E}-02$	$6.13\text{E}-02 \pm 2.68\text{E}-03 \pm 7.27\text{E}-03$	$6.13\text{E}-02 \pm 2.68\text{E}-03 \pm 7.27\text{E}-03$
165-172	$9.44\text{E}-02 \pm 7.41\text{E}-04 \pm 1.26\text{E}-02$	$7.93\text{E}-02 \pm 8.07\text{E}-04 \pm 1.04\text{E}-02$	$4.85\text{E}-02 \pm 6.56\text{E}-04 \pm 6.37\text{E}-03$	$2.43\text{E}-02 \pm 5.42\text{E}-04 \pm 3.18\text{E}-03$	$2.43\text{E}-02 \pm 5.42\text{E}-04 \pm 3.18\text{E}-03$
172-180	$7.64\text{E}-02 \pm 7.76\text{E}-04 \pm 9.27\text{E}-03$	$6.04\text{E}-02 \pm 6.66\text{E}-04 \pm 7.30\text{E}-03$	$3.74\text{E}-02 \pm 5.96\text{E}-04 \pm 4.50\text{E}-03$	$1.83\text{E}-02 \pm 5.08\text{E}-04 \pm 2.23\text{E}-03$	$1.83\text{E}-02 \pm 5.08\text{E}-04 \pm 2.23\text{E}-03$
180-188	$5.86\text{E}-02 \pm 5.24\text{E}-04 \pm 6.26\text{E}-03$	$4.87\text{E}-02 \pm 5.52\text{E}-04 \pm 5.18\text{E}-03$	$2.91\text{E}-02 \pm 4.33\text{E}-04 \pm 2.98\text{E}-03$	$1.34\text{E}-02 \pm 3.58\text{E}-04 \pm 1.48\text{E}-03$	$1.34\text{E}-02 \pm 3.58\text{E}-04 \pm 1.48\text{E}-03$
188-196	$5.02\text{E}-02 \pm 4.79\text{E}-04 \pm 5.59\text{E}-03$	$4.28\text{E}-02 \pm 4.36\text{E}-04 \pm 4.49\text{E}-03$	$2.44\text{E}-02 \pm 3.70\text{E}-04 \pm 2.59\text{E}-03$	$1.23\text{E}-02 \pm 3.27\text{E}-04 \pm 1.11\text{E}-03$	$1.23\text{E}-02 \pm 3.27\text{E}-04 \pm 1.11\text{E}-03$
196-204	$3.83\text{E}-02 \pm 4.69\text{E}-04 \pm 3.98\text{E}-03$	$3.08\text{E}-02 \pm 4.49\text{E}-04 \pm 3.21\text{E}-03$	$1.89\text{E}-02 \pm 3.65\text{E}-04 \pm 1.99\text{E}-03$	$8.97\text{E}-03 \pm 2.93\text{E}-04 \pm 8.39\text{E}-04$	$8.97\text{E}-03 \pm 2.93\text{E}-04 \pm 8.39\text{E}-04$
204-215	$3.43\text{E}-02 \pm 3.16\text{E}-04 \pm 3.48\text{E}-03$	$2.81\text{E}-02 \pm 3.19\text{E}-04 \pm 2.52\text{E}-03$	$1.65\text{E}-02 \pm 2.59\text{E}-04 \pm 1.73\text{E}-03$	$7.21\text{E}-03 \pm 2.15\text{E}-04 \pm 7.55\text{E}-04$	$7.21\text{E}-03 \pm 2.15\text{E}-04 \pm 7.55\text{E}-04$
215-230	$2.46\text{E}-02 \pm 2.29\text{E}-04 \pm 2.22\text{E}-03$	$2.01\text{E}-02 \pm 2.32\text{E}-04 \pm 1.80\text{E}-03$	$1.12\text{E}-02 \pm 1.92\text{E}-04 \pm 1.00\text{E}-03$	$5.16\text{E}-03 \pm 1.52\text{E}-04 \pm 4.62\text{E}-04$	$5.16\text{E}-03 \pm 1.52\text{E}-04 \pm 4.62\text{E}-04$
230-245	$1.63\text{E}-02 \pm 2.09\text{E}-04 \pm 1.92\text{E}-03$	$1.30\text{E}-02 \pm 1.97\text{E}-04 \pm 1.44\text{E}-03$	$7.28\text{E}-03 \pm 1.65\text{E}-04 \pm 8.44\text{E}-04$	$3.00\text{E}-03 \pm 1.32\text{E}-04 \pm 2.78\text{E}-04$	$3.00\text{E}-03 \pm 1.32\text{E}-04 \pm 2.78\text{E}-04$
245-260	$1.34\text{E}-02 \pm 1.96\text{E}-04 \pm 1.52\text{E}-03$	$1.01\text{E}-02 \pm 1.69\text{E}-04 \pm 1.16\text{E}-03$	$5.49\text{E}-03 \pm 1.39\text{E}-04 \pm 6.84\text{E}-04$	$2.10\text{E}-03 \pm 1.04\text{E}-04 \pm 2.06\text{E}-04$	$2.10\text{E}-03 \pm 1.04\text{E}-04 \pm 2.06\text{E}-04$
260-285	$8.58\text{E}-03 \pm 1.12\text{E}-04 \pm 1.20\text{E}-03$	$6.76\text{E}-03 \pm 1.11\text{E}-04 \pm 9.11\text{E}-04$	$3.30\text{E}-03 \pm 8.25\text{E}-05 \pm 5.01\text{E}-04$	$1.41\text{E}-03 \pm 6.30\text{E}-05 \pm 1.60\text{E}-04$	$1.41\text{E}-03 \pm 6.30\text{E}-05 \pm 1.60\text{E}-04$
285-335	$4.61\text{E}-03 \pm 7.05\text{E}-05 \pm 7.81\text{E}-04$	$3.45\text{E}-03 \pm 6.53\text{E}-05 \pm 6.18\text{E}-04$	$1.79\text{E}-03 \pm 4.70\text{E}-05 \pm 2.98\text{E}-04$	$5.53\text{E}-04 \pm 3.44\text{E}-05 \pm 1.32\text{E}-04$	$5.53\text{E}-04 \pm 3.44\text{E}-05 \pm 1.32\text{E}-04$
335-1000	$2.98\text{E}-04 \pm 5.39\text{E}-06 \pm 3.17\text{E}-05$	$2.08\text{E}-04 \pm 4.68\text{E}-06 \pm 3.54\text{E}-05$	$7.39\text{E}-05 \pm 3.29\text{E}-06 \pm 1.23\text{E}-05$	$1.57\text{E}-05 \pm 1.66\text{E}-06 \pm 2.51\text{E}-05$	$1.57\text{E}-05 \pm 1.66\text{E}-06 \pm 2.51\text{E}-05$

Table 5.46: Triple differential cross section values with the statistical and systematic (lower and higher) uncertainties for various p_T^γ regions and $|\eta^{\gamma et}|$ bins for $1.56 < |\eta^\gamma| < 2.1$.

p_T^γ (GeV)	$ \eta^{\gamma et} < 0.8$	$0.8 < \eta^{\gamma et} < 1.5$	$1.5 < \eta^{\gamma et} < 2.1$	$2.1 < \eta^{\gamma et} < 2.5$
40–65	$1.19\text{E}+01 \pm 5.37\text{E}-01$ $+2.89\text{E}+00$ $-3.05\text{E}+00$	$1.18\text{E}+01 \pm 5.36\text{E}-01$ $+2.83\text{E}+00$ $-2.95\text{E}+00$	$9.68\text{E}+00 \pm 5.12\text{E}-01$ $+2.28\text{E}+00$ $-2.36\text{E}+00$	$5.74\text{E}+00 \pm 4.95\text{E}-01$ $+1.40\text{E}+00$ $-1.45\text{E}+00$
65–90	$1.53\text{E}+00 \pm 7.24\text{E}-02$ $+4.03\text{E}-01$ $-4.02\text{E}-01$	$1.62\text{E}+00 \pm 7.35\text{E}-02$ $+3.90\text{E}-01$ $-3.78\text{E}-01$	$1.59\text{E}+00 \pm 7.37\text{E}-02$ $+3.55\text{E}-01$ $-3.67\text{E}-01$	$1.18\text{E}+00 \pm 7.43\text{E}-02$ $+2.60\text{E}-01$ $-2.63\text{E}-01$
90–105	$9.95\text{E}-01 \pm 2.03\text{E}-02$ $+2.32\text{E}-01$ $-2.76\text{E}-01$	$1.02\text{E}+00 \pm 2.33\text{E}-02$ $+2.67\text{E}-01$ $-2.67\text{E}-01$	$8.18\text{E}-01 \pm 2.21\text{E}-02$ $+1.98\text{E}-01$ $-1.98\text{E}-01$	$6.34\text{E}-01 \pm 2.35\text{E}-02$ $+1.47\text{E}-01$ $-1.47\text{E}-01$
105–165	$1.96\text{E}-01 \pm 3.53\text{E}-03$ $+5.91\text{E}-02$ $-5.91\text{E}-02$	$2.11\text{E}-01 \pm 3.67\text{E}-03$ $+4.30\text{E}-02$ $-5.92\text{E}-02$	$1.81\text{E}-01 \pm 3.44\text{E}-03$ $+3.48\text{E}-02$ $-4.23\text{E}-02$	$1.07\text{E}-01 \pm 3.35\text{E}-03$ $+2.04\text{E}-02$ $-2.58\text{E}-02$
165–180	$5.15\text{E}-02 \pm 5.37\text{E}-04$ $+9.52\text{E}-03$ $-8.85\text{E}-03$	$5.68\text{E}-02 \pm 5.99\text{E}-04$ $+9.31\text{E}-03$ $-6.06\text{E}-03$	$4.42\text{E}-02 \pm 5.67\text{E}-04$ $+7.32\text{E}-03$ $-4.01\text{E}-03$	$2.65\text{E}-02 \pm 4.90\text{E}-04$ $+4.23\text{E}-03$ $-2.22\text{E}-03$
180–200	$3.32\text{E}-02 \pm 3.22\text{E}-04$ $+4.75\text{E}-03$ $-6.36\text{E}-03$	$3.54\text{E}-02 \pm 3.56\text{E}-04$ $+5.05\text{E}-03$ $-2.84\text{E}-03$	$2.88\text{E}-02 \pm 2.94\text{E}-04$ $+4.01\text{E}-03$ $-4.68\text{E}-03$	$1.58\text{E}-02 \pm 2.77\text{E}-04$ $+3.22\text{E}-03$ $-2.50\text{E}-03$
200–225	$1.80\text{E}-02 \pm 2.04\text{E}-04$ $+2.78\text{E}-03$ $-2.91\text{E}-03$	$1.89\text{E}-02 \pm 2.18\text{E}-04$ $+2.89\text{E}-03$ $-1.12\text{E}-04$	$1.45\text{E}-02 \pm 1.93\text{E}-04$ $+2.13\text{E}-03$ $-2.14\text{E}-03$	$7.26\text{E}-03 \pm 1.68\text{E}-04$ $+3.05\text{E}-03$ $-3.44\text{E}-03$
225–1000	$8.29\text{E}-04 \pm 9.31\text{E}-06$ $+1.18\text{E}-04$ $-1.18\text{E}-04$	$7.95\text{E}-04 \pm 9.30\text{E}-06$ $+1.12\text{E}-04$ $-1.12\text{E}-04$	$5.41\text{E}-04 \pm 7.93\text{E}-06$ $+7.39\text{E}-05$ $-7.39\text{E}-05$	$2.02\text{E}-04 \pm 5.77\text{E}-06$ $+2.84\text{E}-05$ $-2.84\text{E}-05$

Table 5.47: Triple differential cross section values with the statistical and systematic (lower and higher) uncertainties for various p_T^γ regions for four bins of $|\eta^{jet}|$. These values are for $2.1 < |\eta^\gamma| < 2.5$.

p_T^γ (GeV)	$ \eta^{jet} < 0.8$	$0.8 < \eta^{jet} < 1.5$	$1.5 < \eta^{jet} < 2.1$	$2.1 < \eta^{jet} < 2.5$
40–65	$7.35E+00 \pm 5.70E-01$ $+2.04E+00$ $-3.34E+00$	$8.73E+00 \pm 5.93E-01$ $+2.13E+00$ $-3.21E+00$	$7.83E+00 \pm 5.52E-01$ $+1.82E+00$ $-2.53E+00$	$7.69E+00 \pm 6.00E-01$ $+1.63E+00$ $-2.11E+00$
65–90	$1.42E+00 \pm 5.63E-02$ $+9.79E-01$ $-1.02E+00$	$1.63E+00 \pm 6.41E-02$ $+9.80E-01$ $-1.07E+00$	$1.68E+00 \pm 6.41E-02$ $+8.57E-01$ $-9.82E-01$	$1.07E+00 \pm 6.39E-02$ $+5.55E-01$ $-5.72E-01$
90–105	$4.74E-01 \pm 2.11E-02$ $+2.60E-01$ $-2.20E-01$	$5.56E-01 \pm 2.06E-02$ $+2.40E-01$ $-2.17E-01$	$5.63E-01 \pm 2.03E-02$ $+2.15E-01$ $-1.92E-01$	$5.01E-01 \pm 2.32E-02$ $+1.68E-01$ $-1.72E-01$
105–165	$1.53E-01 \pm 3.06E-03$ $+5.95E-02$ $-5.95E-02$	$1.67E-01 \pm 3.28E-03$ $+6.17E-02$ $-6.08E-02$	$1.63E-01 \pm 3.48E-03$ $+5.06E-02$ $-4.82E-02$	$1.19E-01 \pm 3.56E-03$ $+3.46E-02$ $-3.46E-02$
165–180	$2.77E-02 \pm 4.21E-04$ $+1.32E-02$ $-1.32E-02$	$3.31E-02 \pm 4.74E-04$ $+1.31E-02$ $-1.31E-02$	$3.11E-02 \pm 4.67E-04$ $+1.07E-02$ $-1.07E-02$	$1.83E-02 \pm 4.11E-04$ $+3.00E-03$ $-3.00E-03$
180–200	$1.85E-02 \pm 2.48E-04$ $+6.99E-03$ $-6.99E-03$	$2.13E-02 \pm 2.79E-04$ $+7.13E-03$ $-7.13E-03$	$1.87E-02 \pm 2.66E-04$ $+5.48E-03$ $-5.48E-03$	$1.11E-02 \pm 2.47E-04$ $+3.03E-03$ $-3.03E-03$
200–225	$9.25E-03 \pm 1.68E-04$ $+3.65E-03$ $-3.65E-03$	$1.05E-02 \pm 1.78E-04$ $+3.69E-03$ $-3.69E-03$	$9.85E-03 \pm 1.77E-04$ $+2.67E-03$ $-2.67E-03$	$5.19E-03 \pm 1.53E-04$ $+1.33E-03$ $-1.33E-03$
225–1000	$3.37E-04 \pm 6.68E-06$ $+1.09E-04$ $-1.09E-04$	$3.67E-04 \pm 6.60E-06$ $+1.03E-04$ $-1.03E-04$	$2.60E-04 \pm 5.49E-06$ $+6.77E-05$ $-6.77E-05$	$1.04E-04 \pm 4.14E-06$ $+2.66E-05$ $-2.66E-05$

CHAPTER 6

CONCLUSION

The triple differential cross section $d^3\sigma/(dp_T\gamma d|\eta^\gamma|d|\eta^{jet}|)$ has been measured for the production of $\gamma + \text{jet}$ in proton-proton collisions at a center-of-mass energy of 8 TeV. The measurement is presented for 224 bins in p_T^γ , $|\eta^\gamma|$, and $|\eta^{jet}|$ for photon and jet pseudorapidities $|\eta^{\gamma/jet}| < 2.5$, and for $40 < p_T^\gamma < 1000$ GeV. The measured cross section probes the kinematic range of $Q^2 = (p_T^\gamma)^2 = [1.6 \times 10^3 - 10^6]$ GeV² and $x_T = 2p_T^\gamma/\sqrt{s} = [0.01 - 0.25]$, where x_T is an approximation for x when both the photon and jet pseudorapidity equal to zero. Based on the combination of the photon and jet pseudorapidity bins, wide range of x can be probed with the results from this measurement at a center-of-mass energy that was previously unexplored.

The differential cross section was measured by estimating signal purity using a template fitting method. The signal templates were obtained from Monte Carlo simulations, whereas the background templates were obtained from data sidebands. The inefficiencies in the measurement were estimated from various sources and folded into the final differential cross section values. The data distributions were unfolded in the bins of p_T^γ to correct for detector response. The uncertainties in the number of candidate events, purity, efficiency, unfolding, and luminosity measurements were estimated individually and then used to estimate the uncertainty in the measured differential cross section. The statistical uncertainty was found to be negligible in comparison to the systematic uncertainty for most bins. The most significant source of uncertainty came from efficiency measurements and purity estimates. In general, the data uncertainties in low p_T^γ and high $|\eta^\gamma|$ bins were found to be larger than in the higher values of p_T^γ and low $|\eta^\gamma|$ bins.

The measured differential cross section was compared to the QCD NLO prediction using the GamJet package with the CJ15 PDF set. Theory uncertainties related to the PDF set and the choices of scale were also calculated. The results for $\gamma + \text{jet}$ production show agreement with the theoretical predictions, within the uncertainties of the measurement and the theory for most bins. The

uncertainties in the measurement are found to be comparable or somewhat smaller to the tightest uncertainty bands for theory predictions. Therefore, the results from this measurement can be used to extract information on the gluon PDF. In the bins where some disagreement was observed, better understanding of the CJ15 PDF uncertainty is necessary.

APPENDIX A

MVA-ISOLATION CORRELATION STUDY

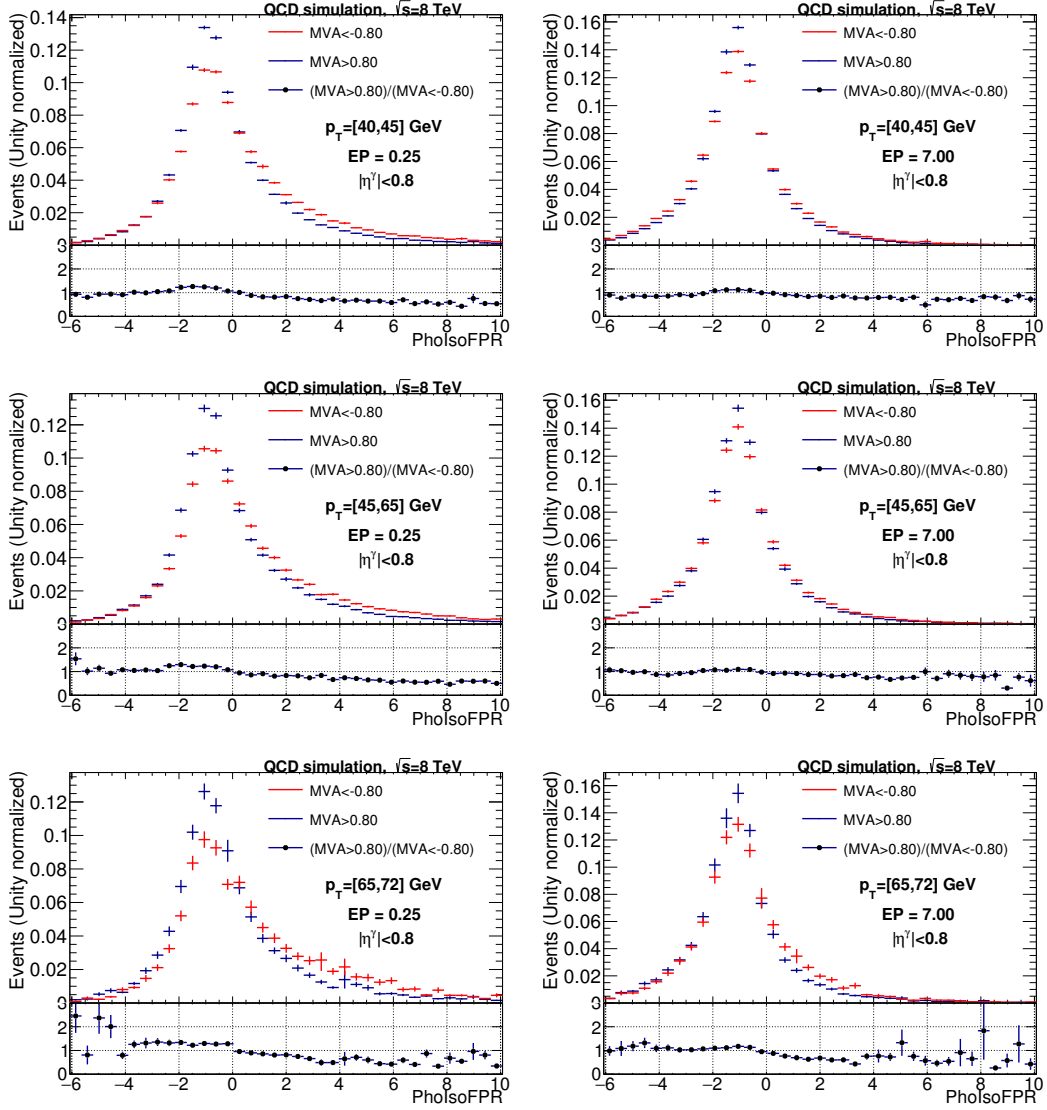


Figure A.1: Photon isolation with footprint removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

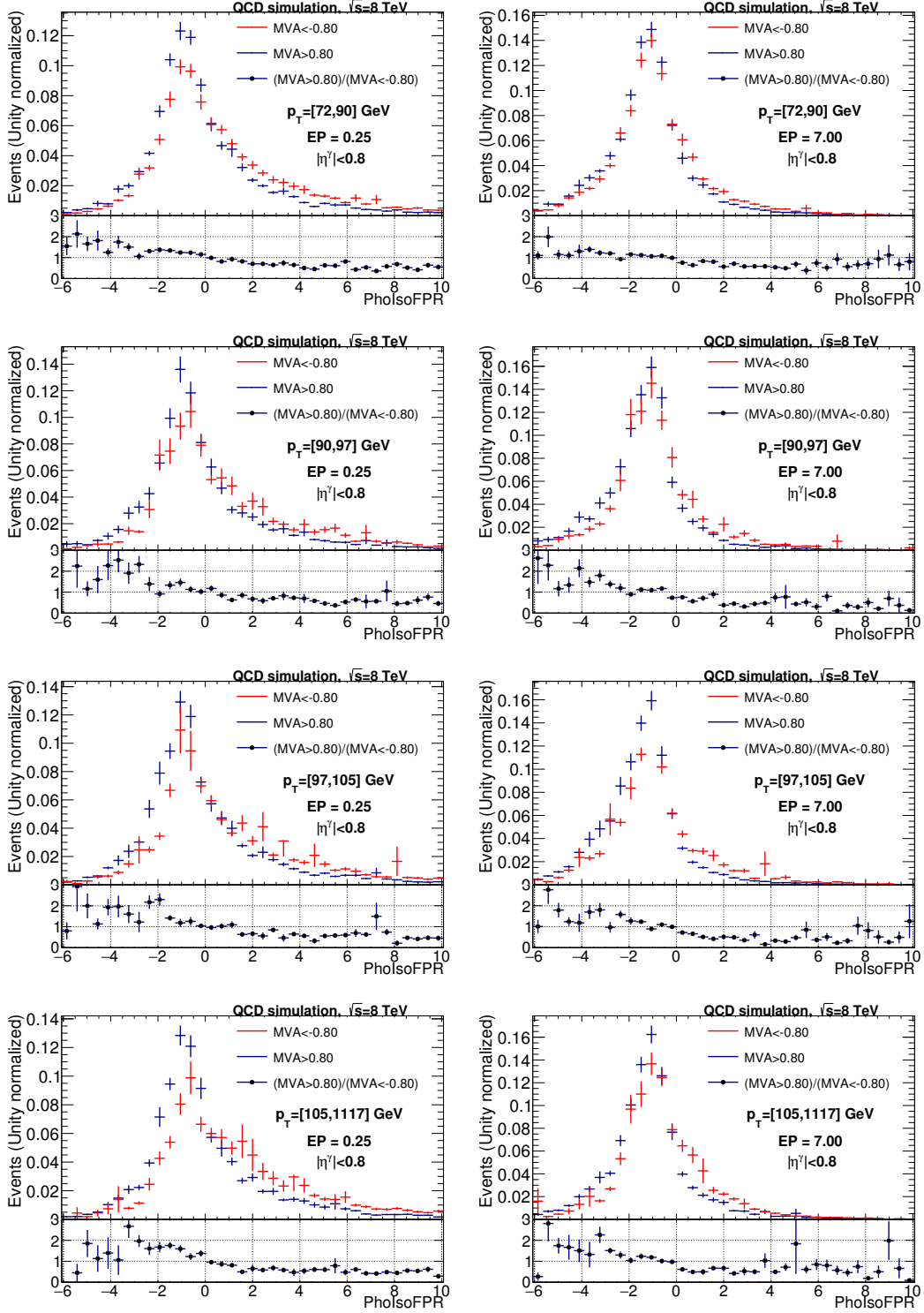


Figure A.2: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

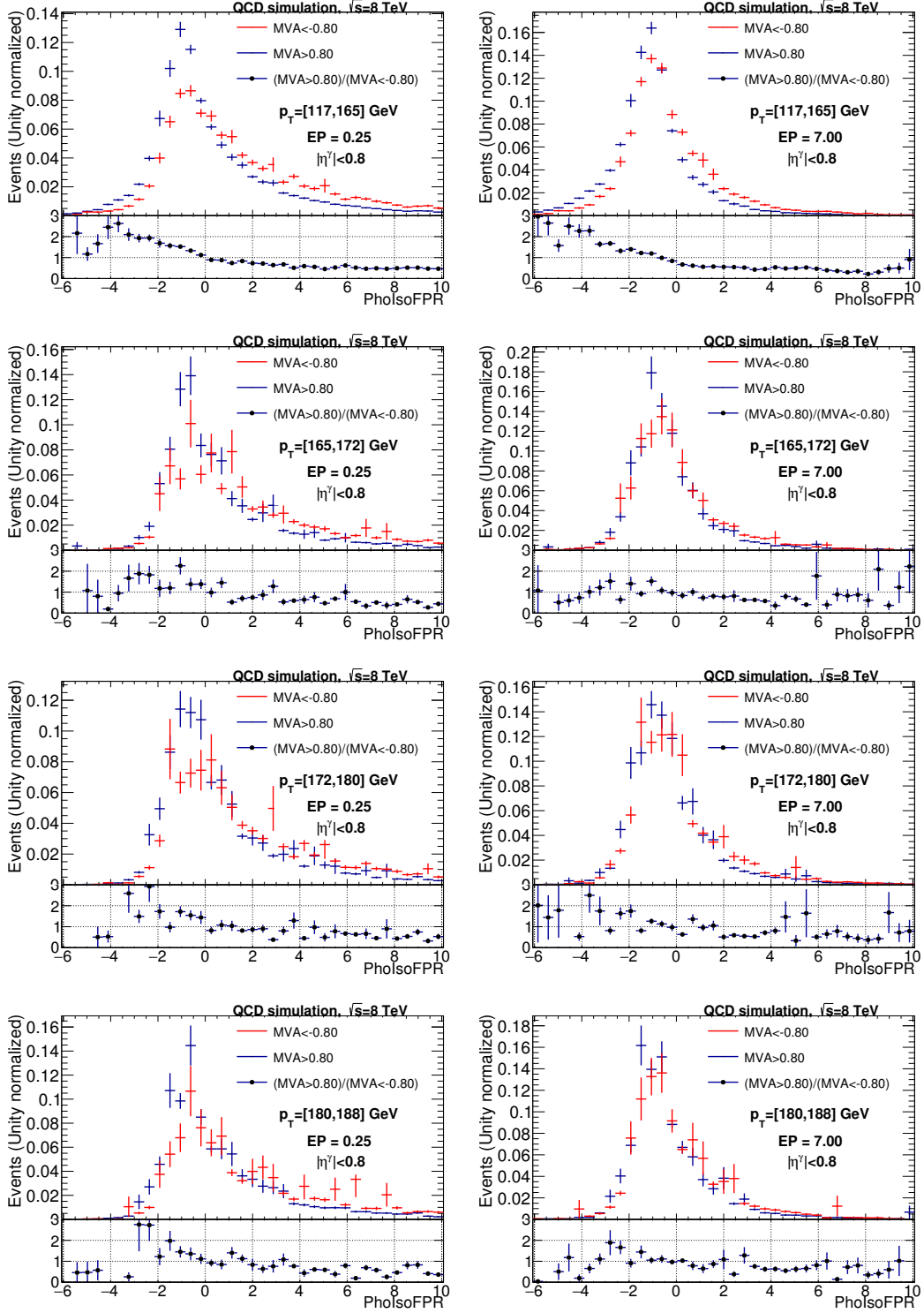


Figure A.3: Photon isolation with footprint removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

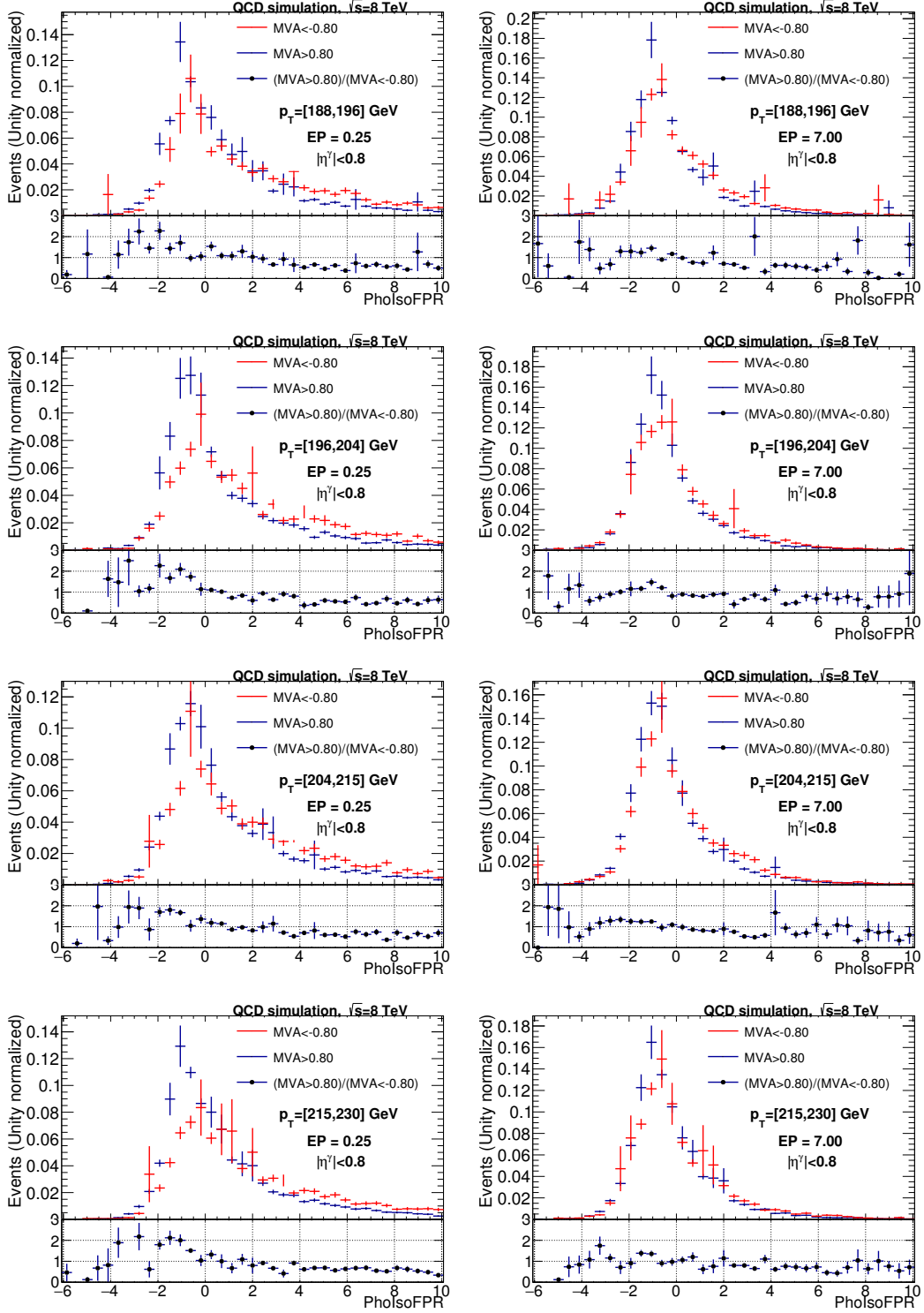


Figure A.4: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

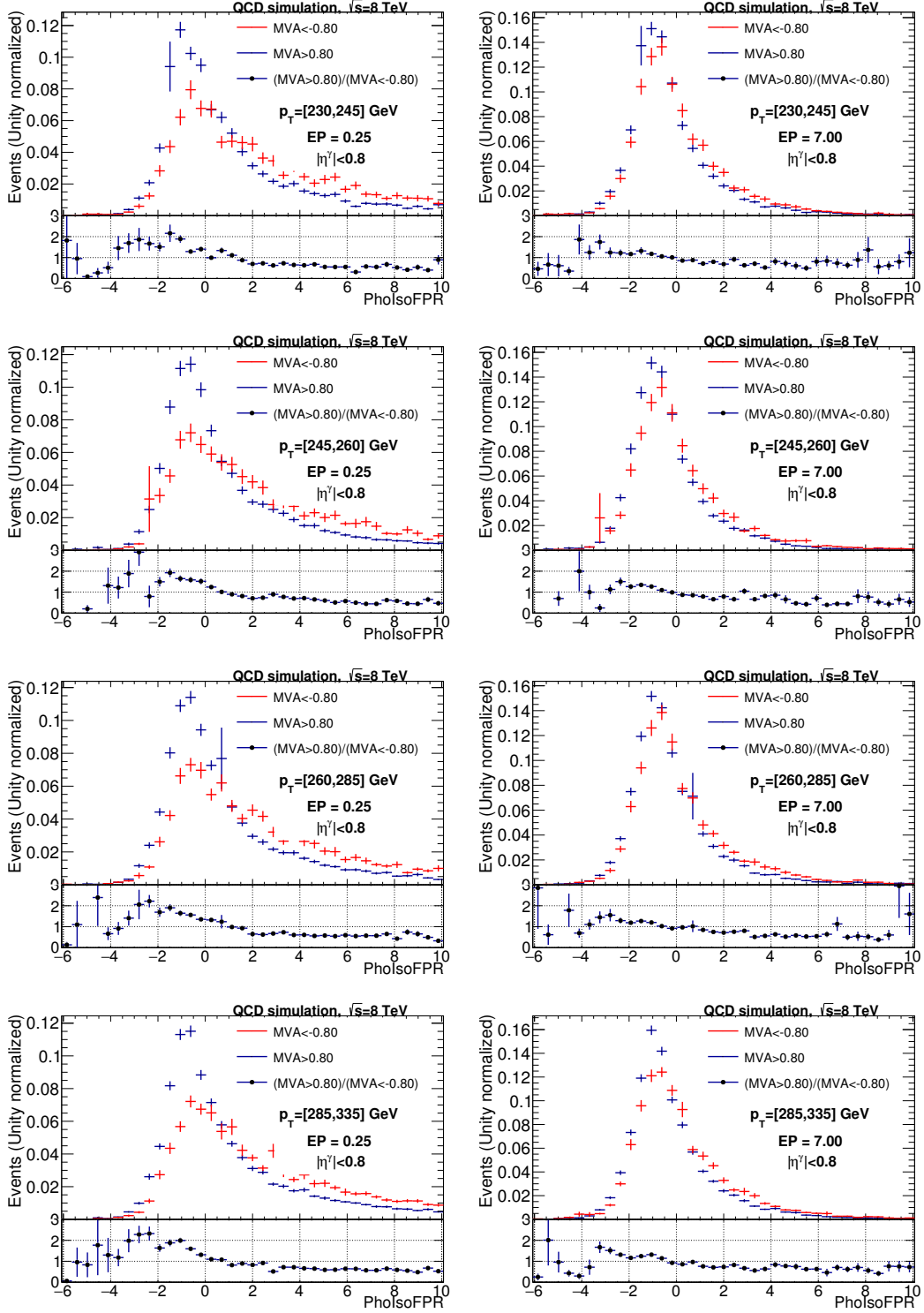


Figure A.5: Photon isolation with footprint removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

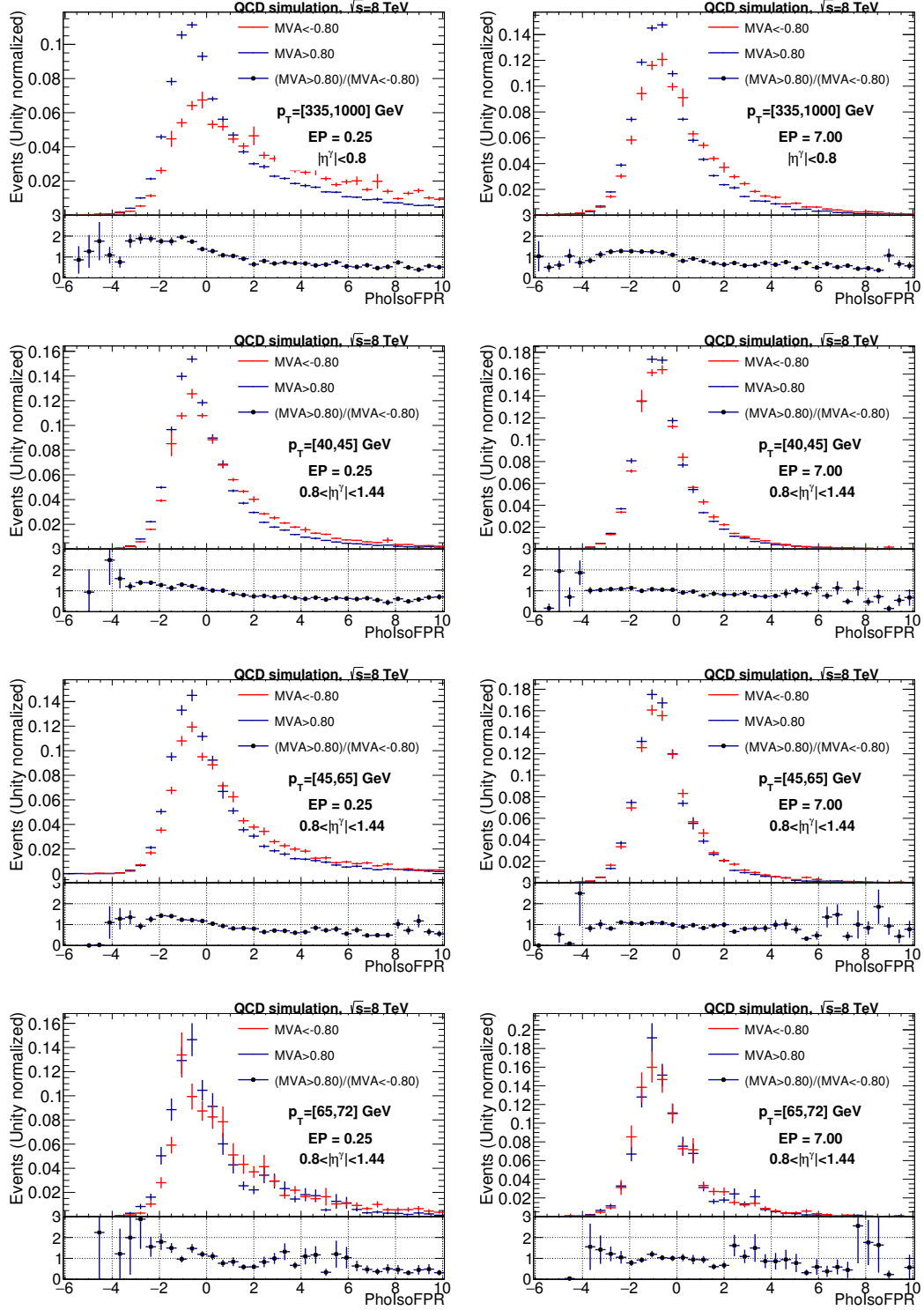


Figure A.6: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

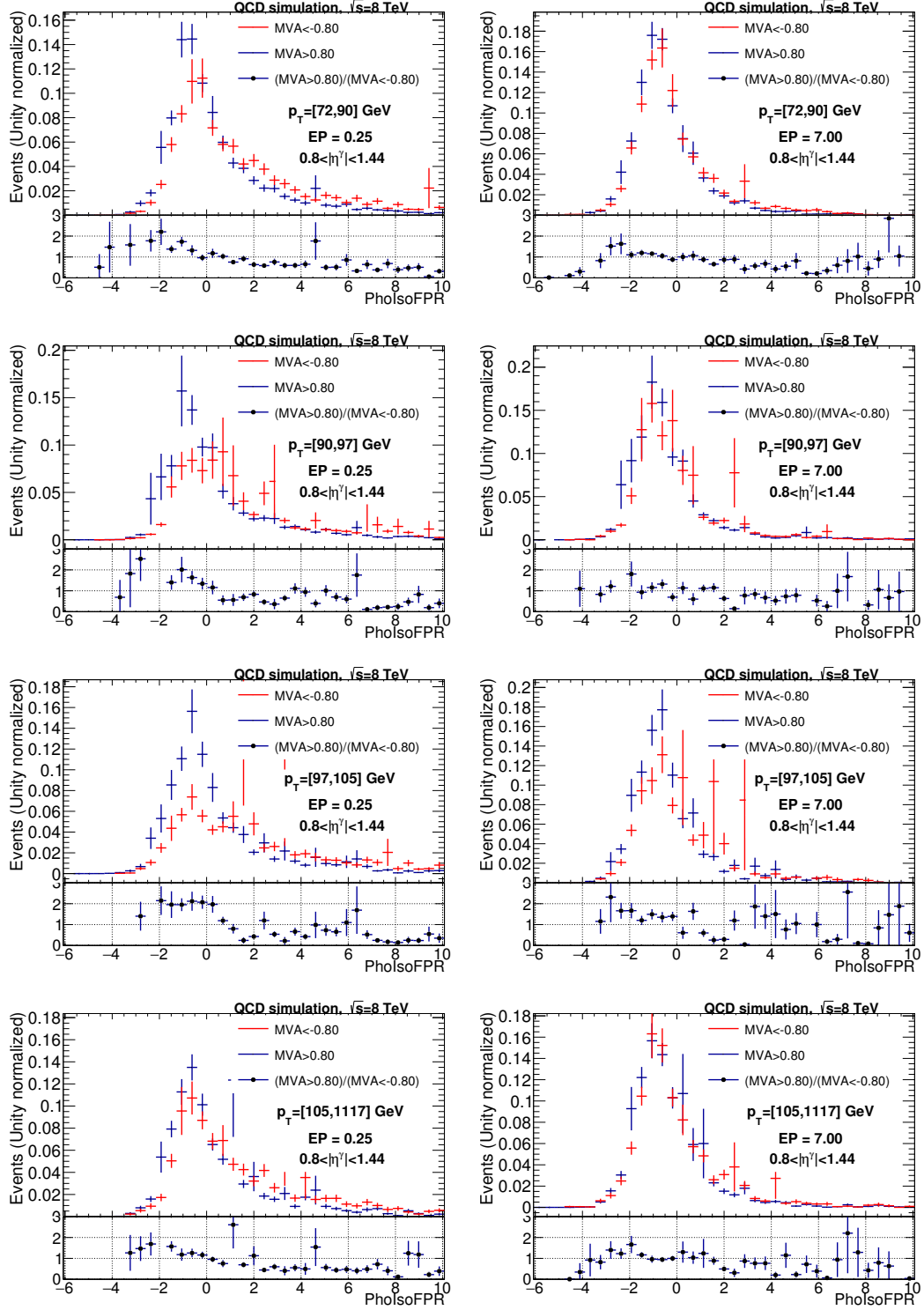


Figure A.7: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

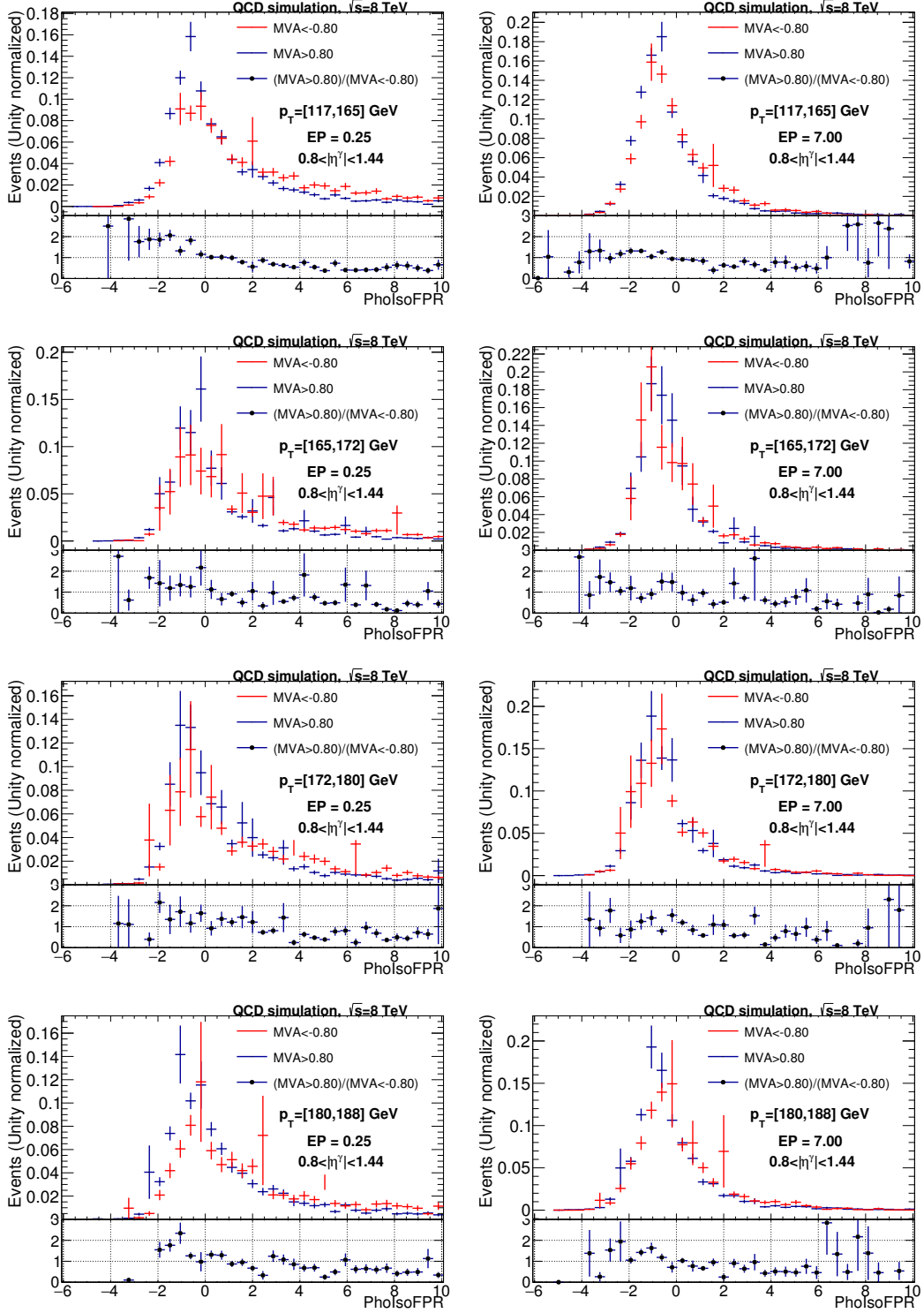


Figure A.8: Photon isolation with footprint removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

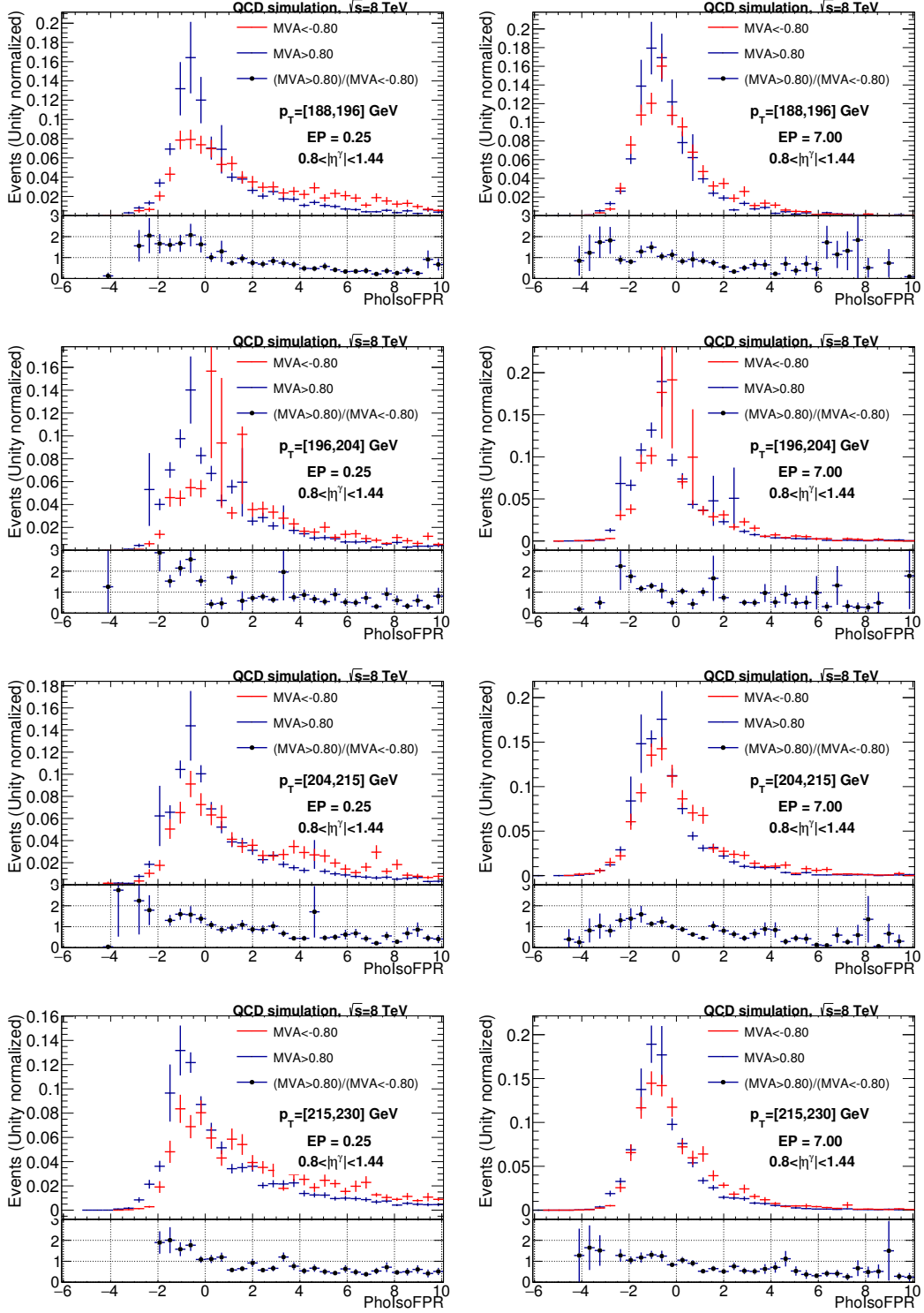


Figure A.9: Photon isolation with footprint removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

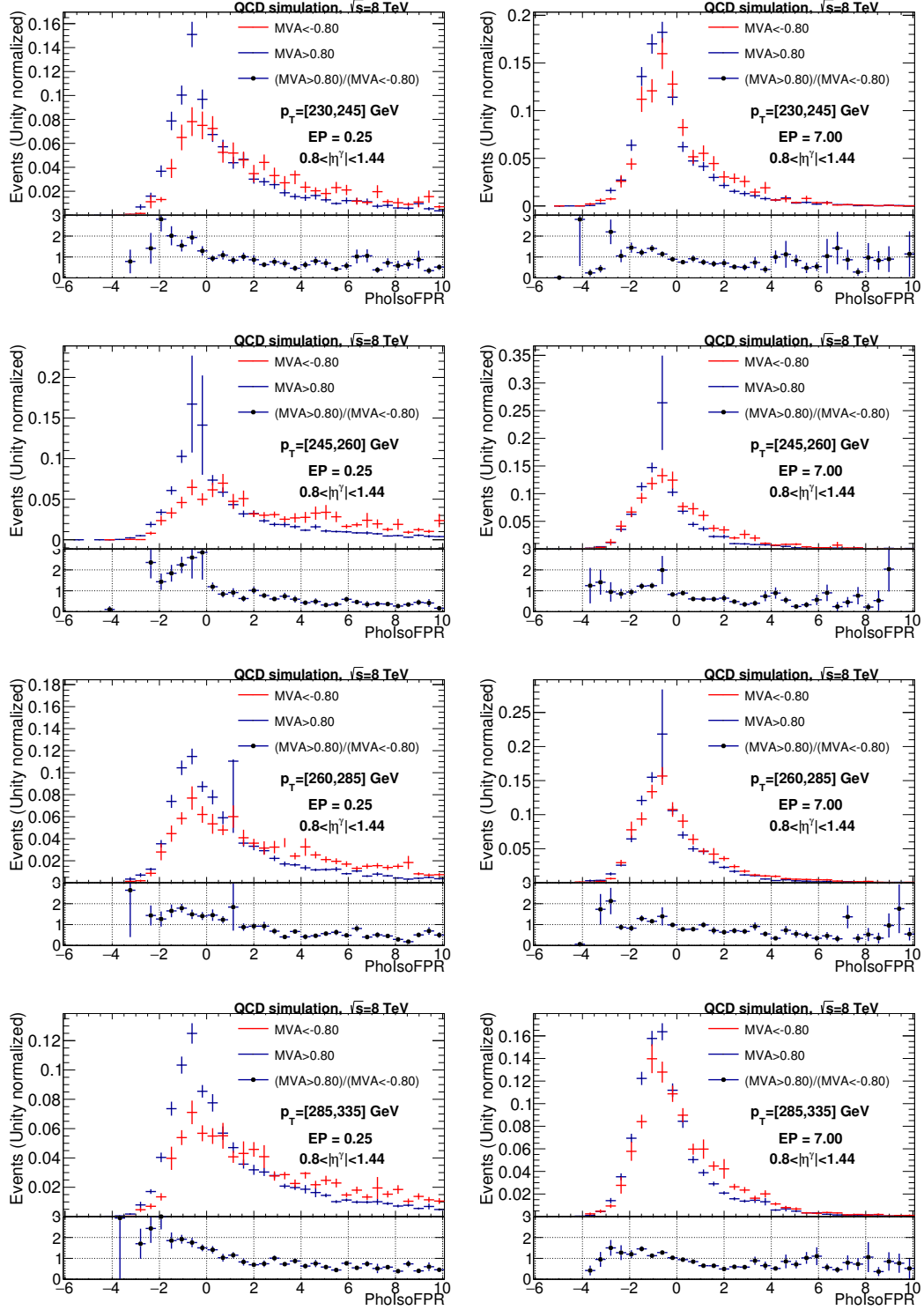


Figure A.10: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

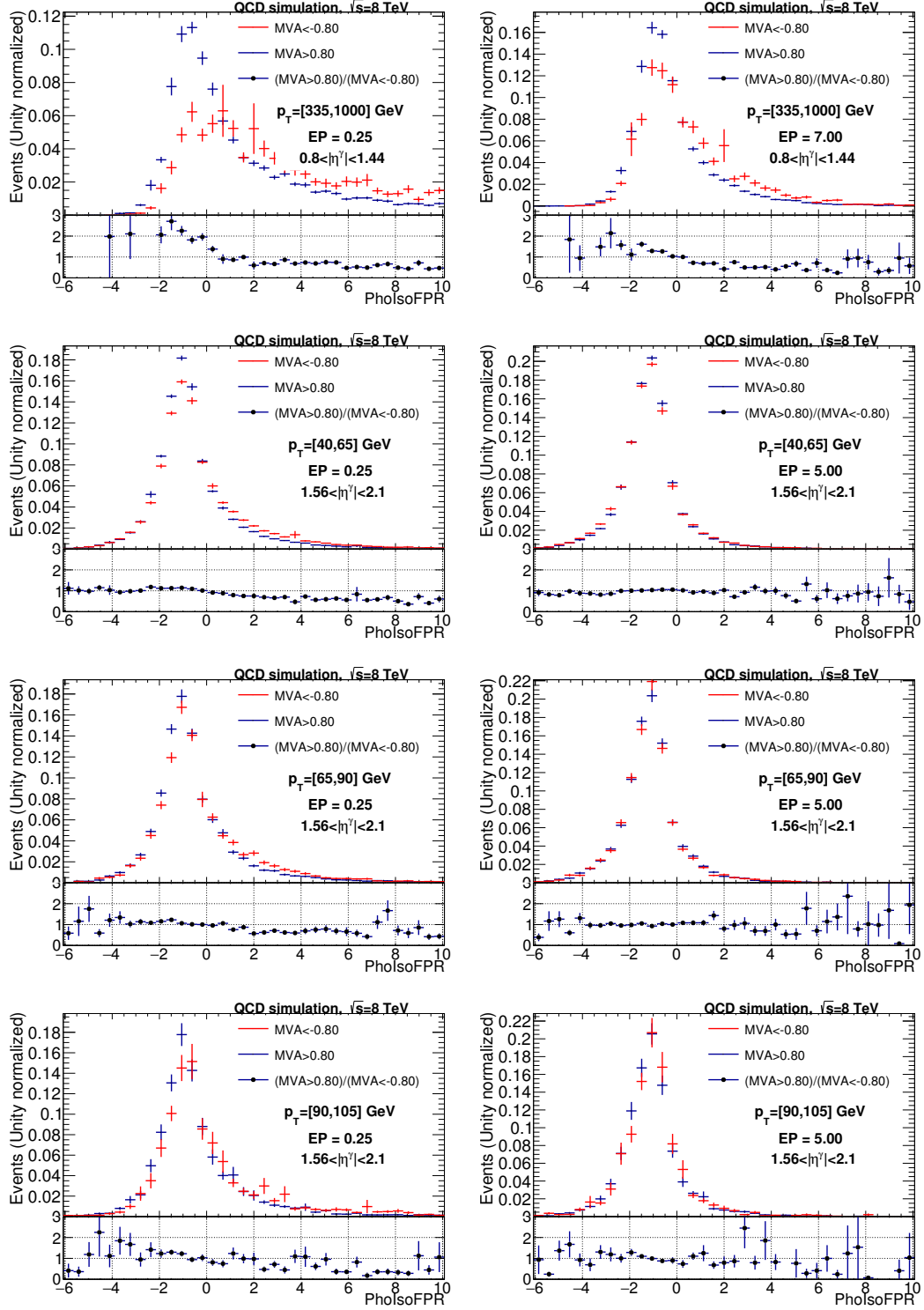


Figure A.11: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

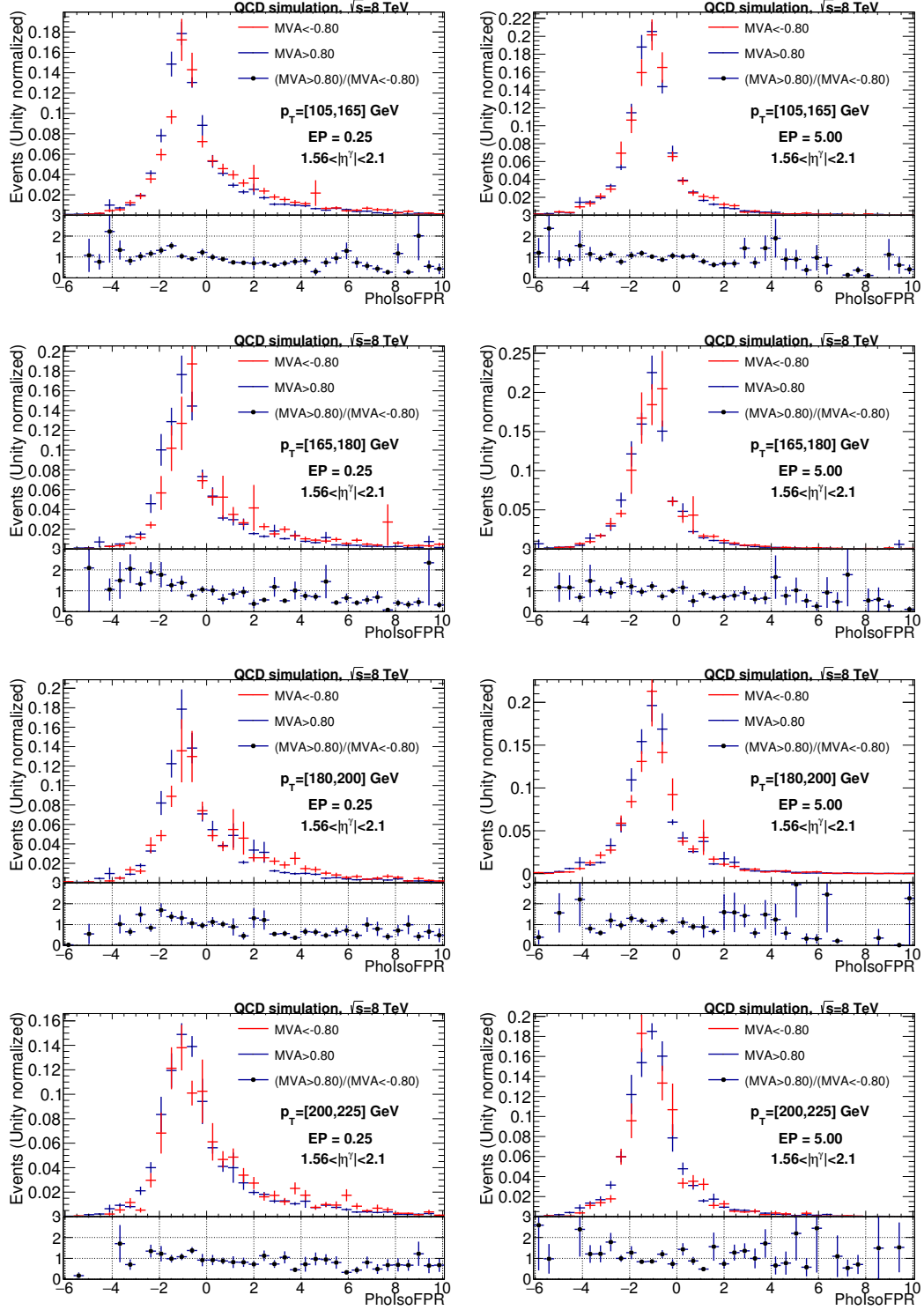


Figure A.12: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

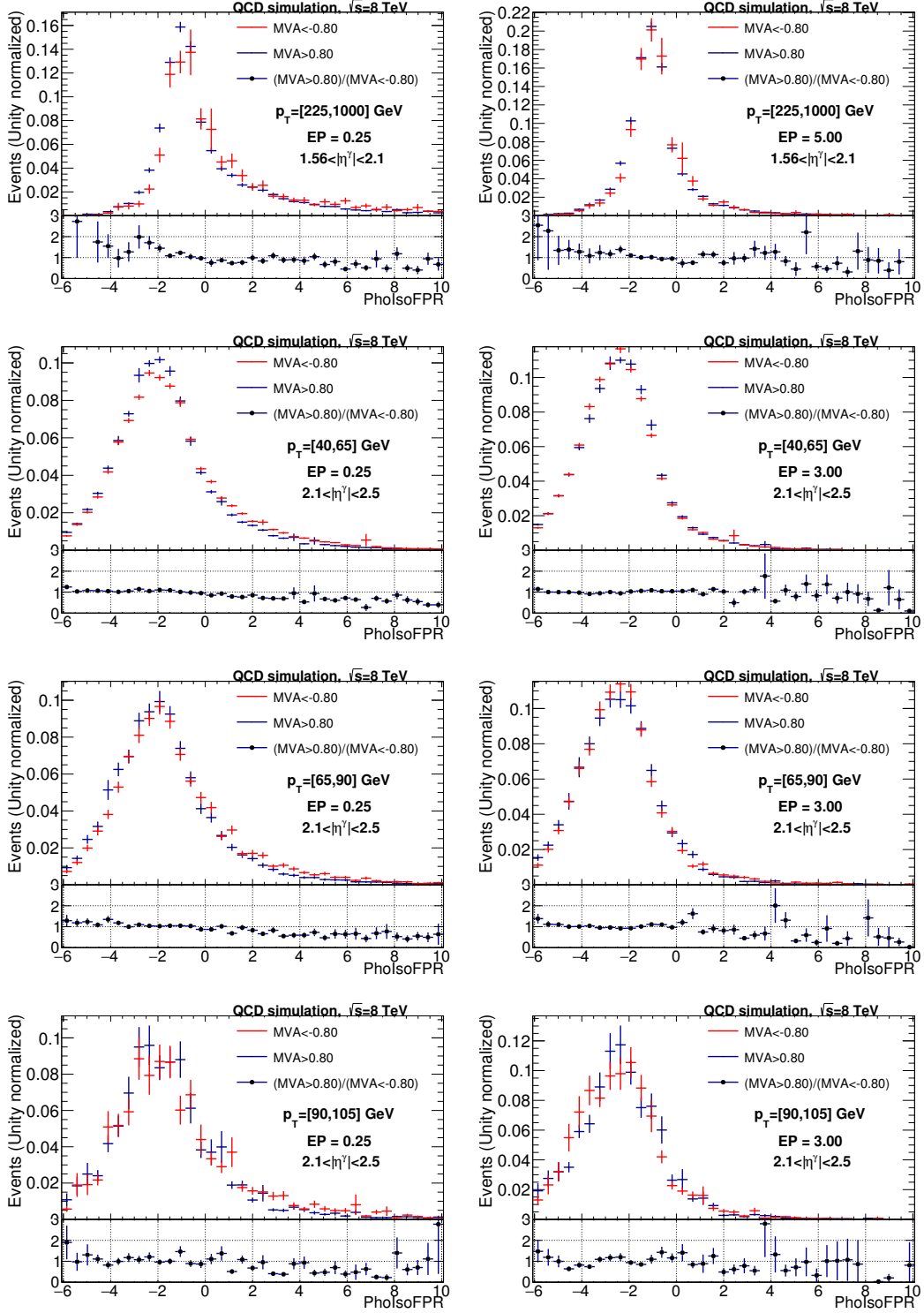


Figure A.13: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

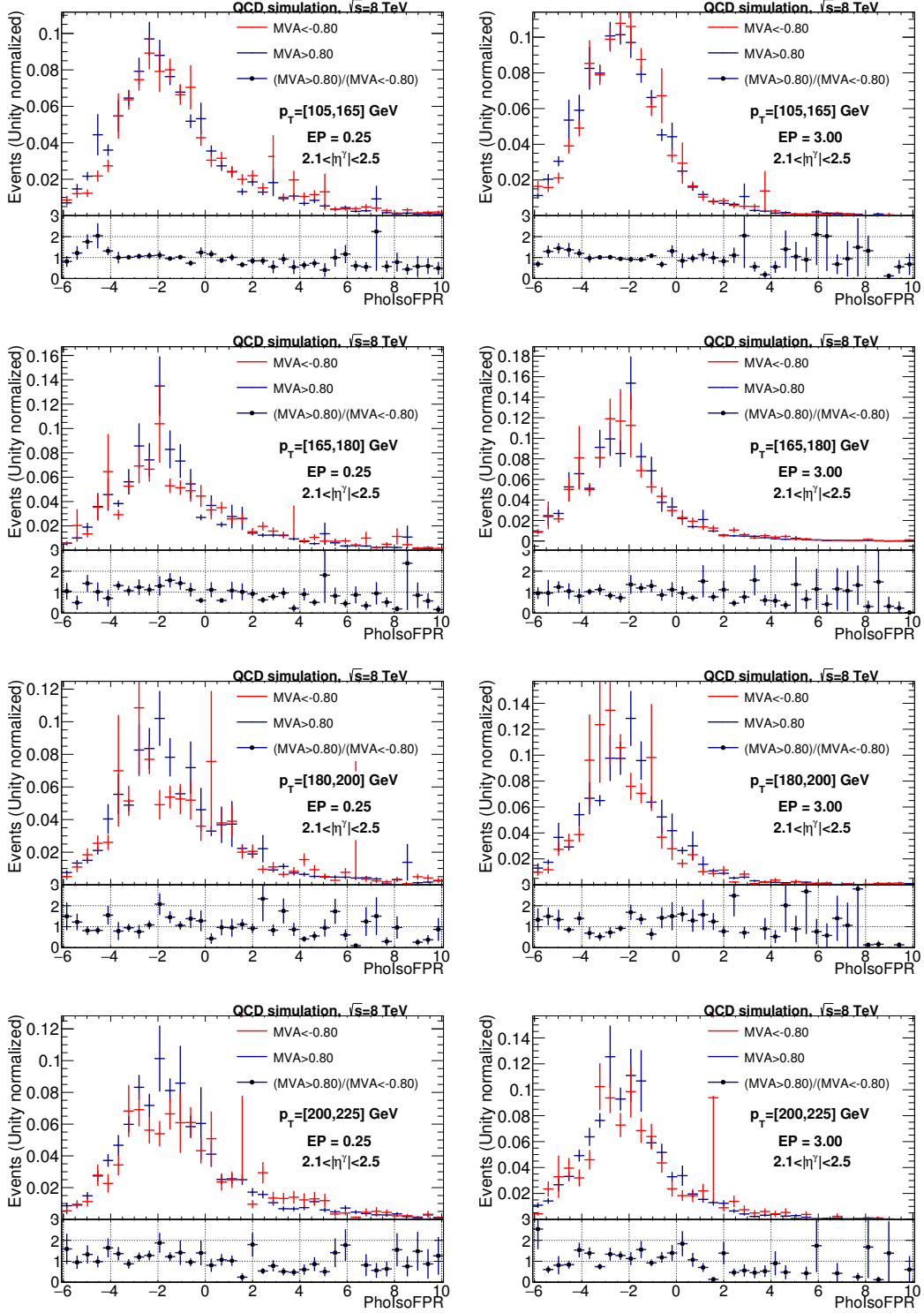


Figure A.14: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

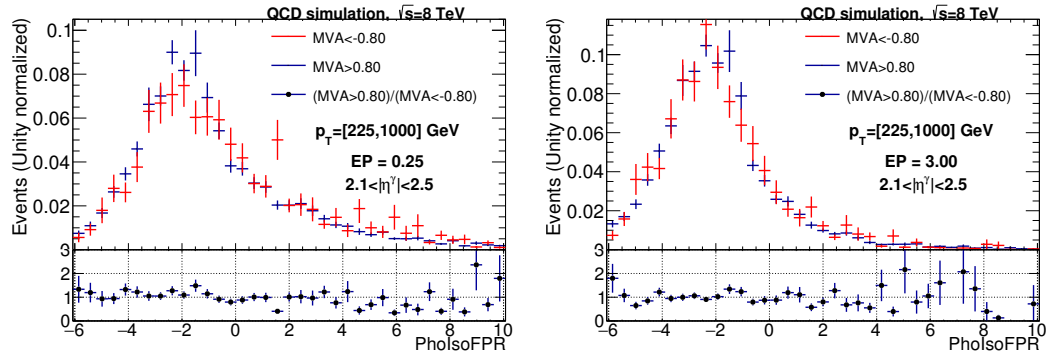


Figure A.15: Photon isolation with foot print removal using default setting [left] and tuned setting [right] for QCD MC fakes in various kinematic regions.

APPENDIX B

BIAS PLANES

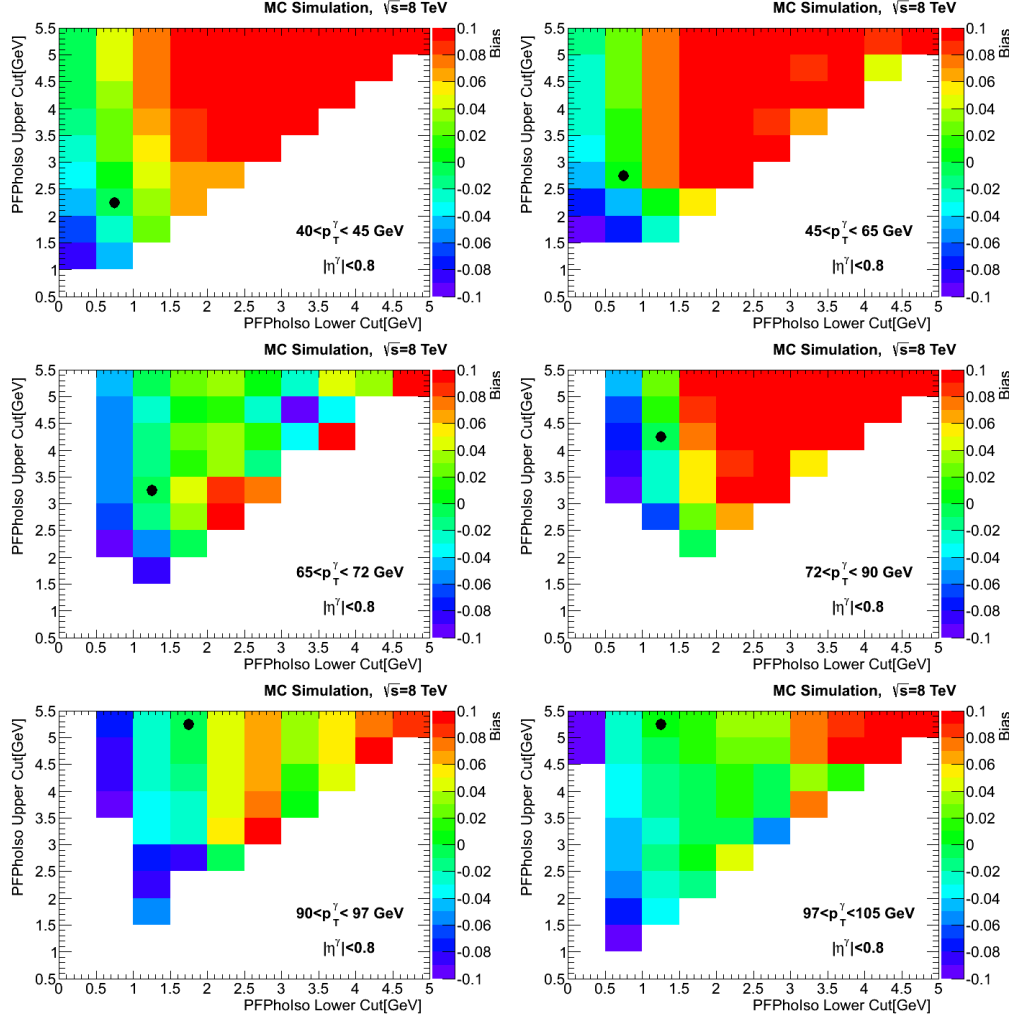


Figure B.1: Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.

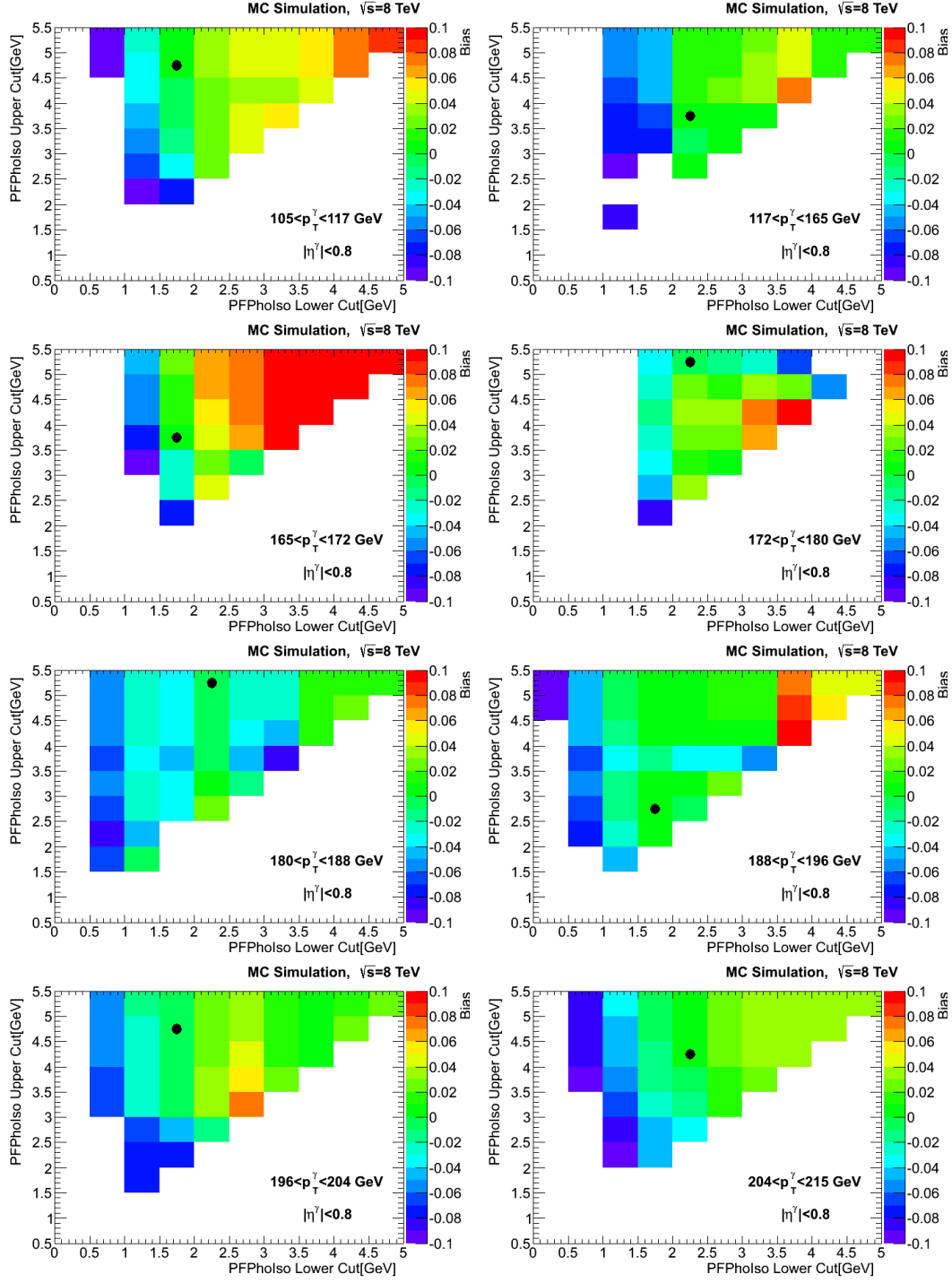


Figure B.2: Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.

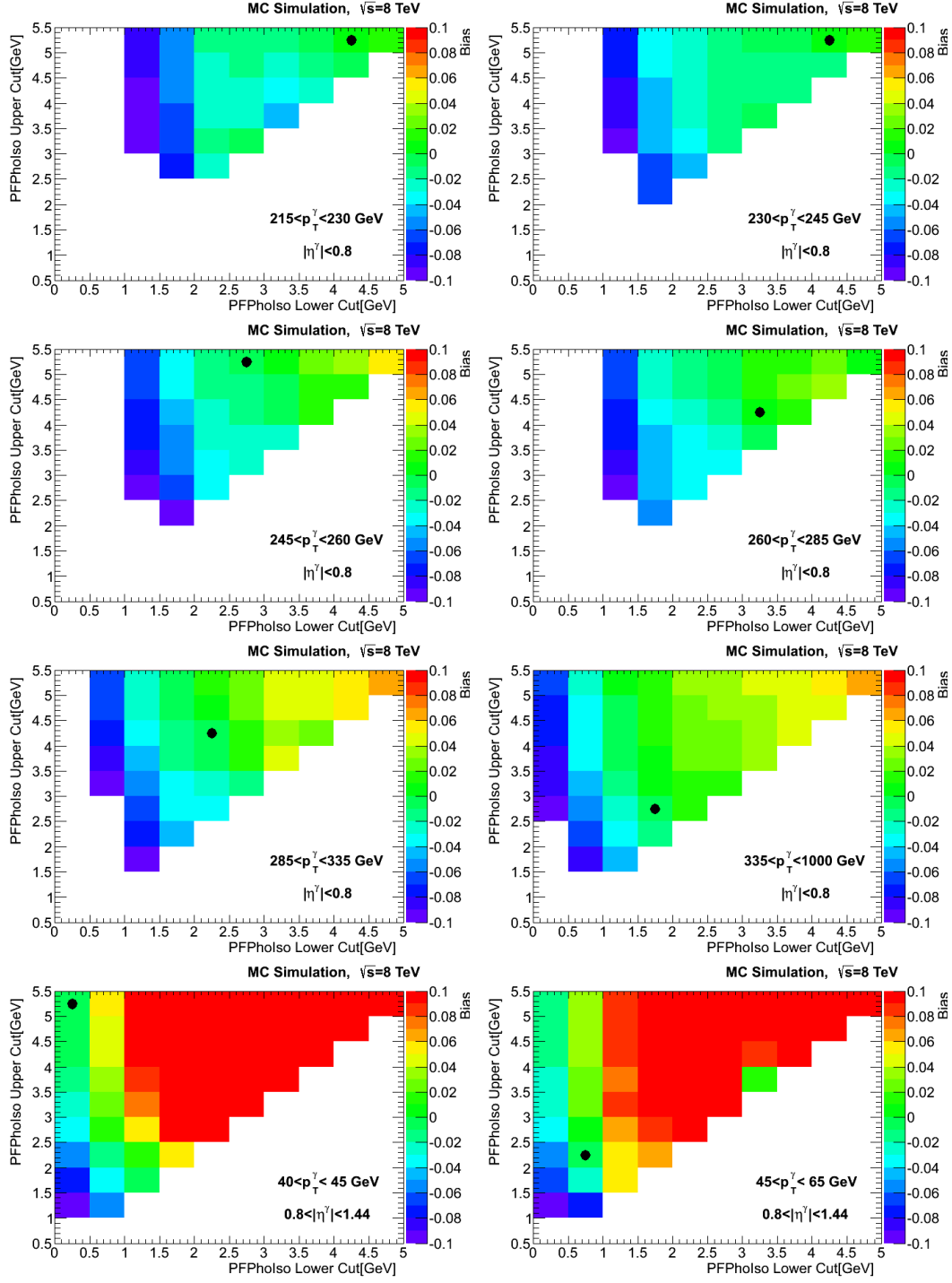


Figure B.3: Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.

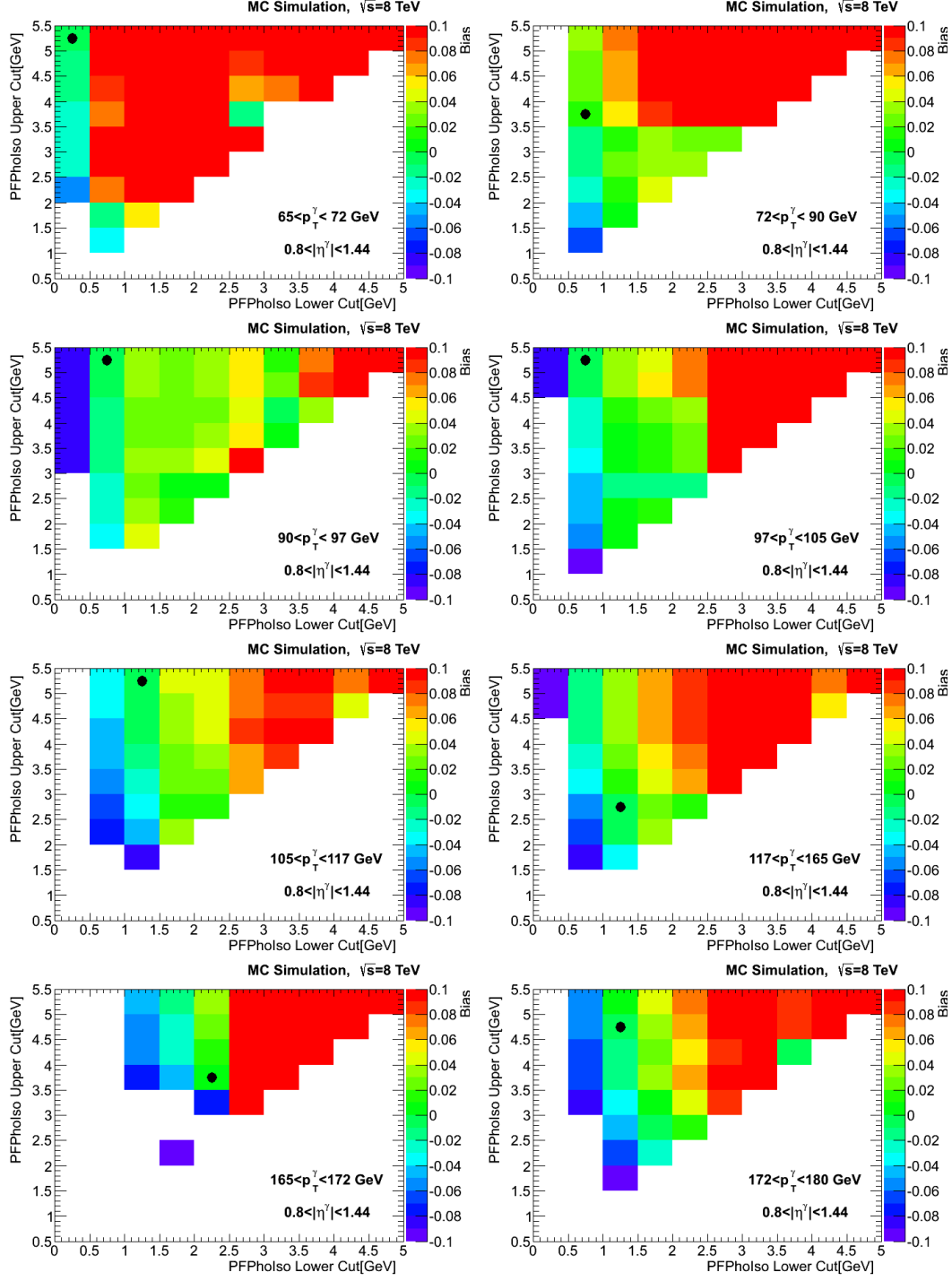


Figure B.4: Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.

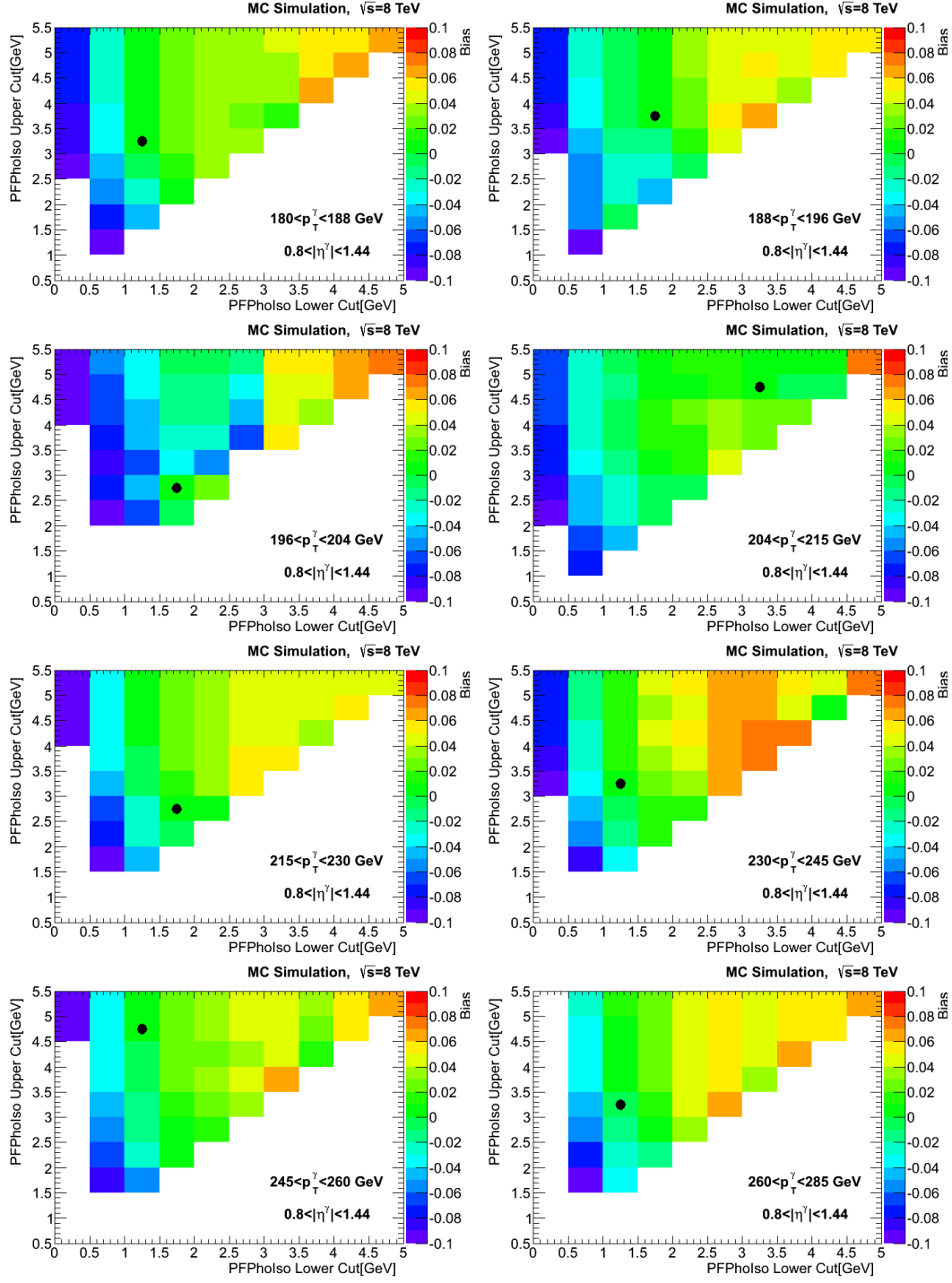


Figure B.5: Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.

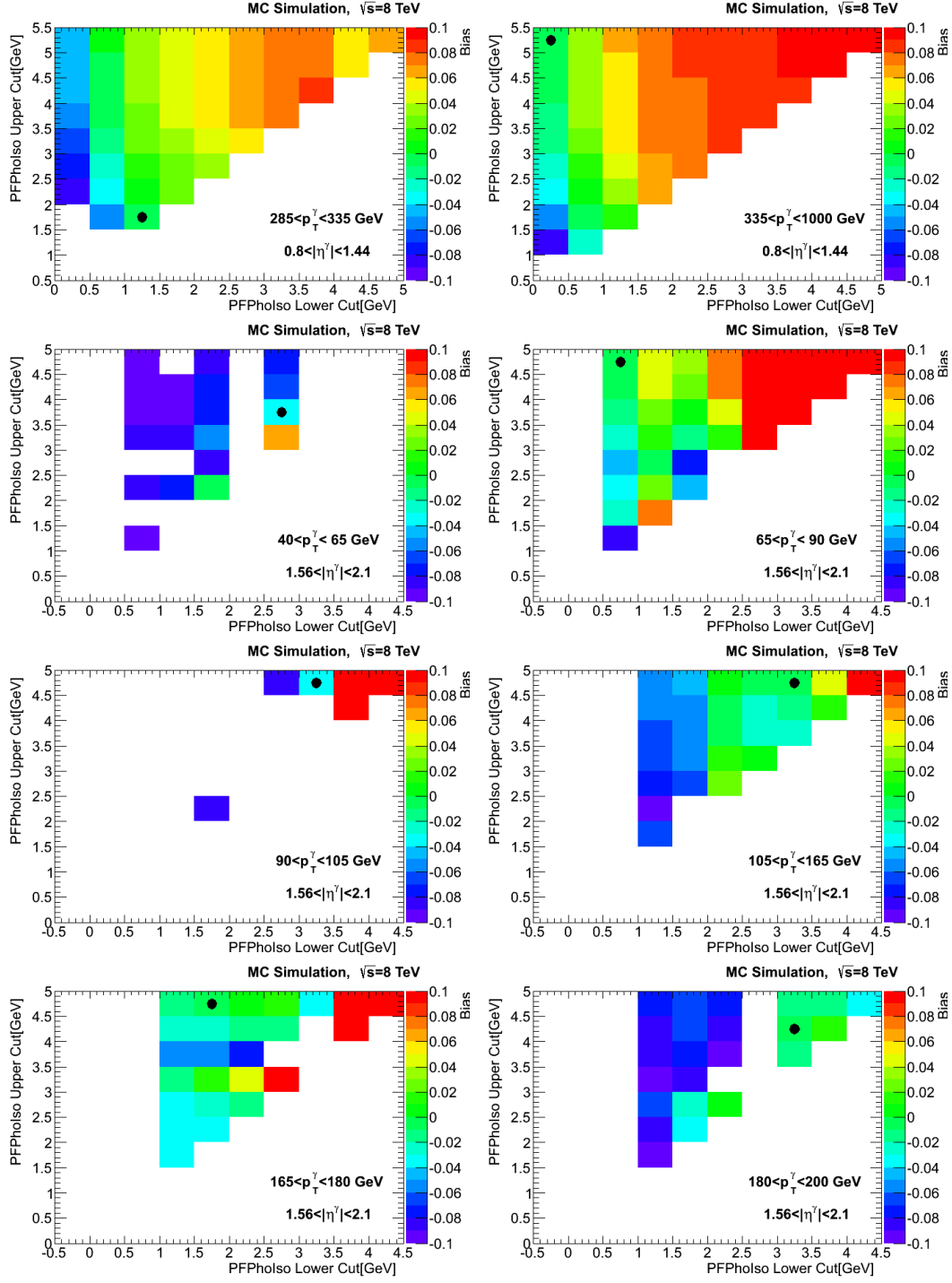


Figure B.6: Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.

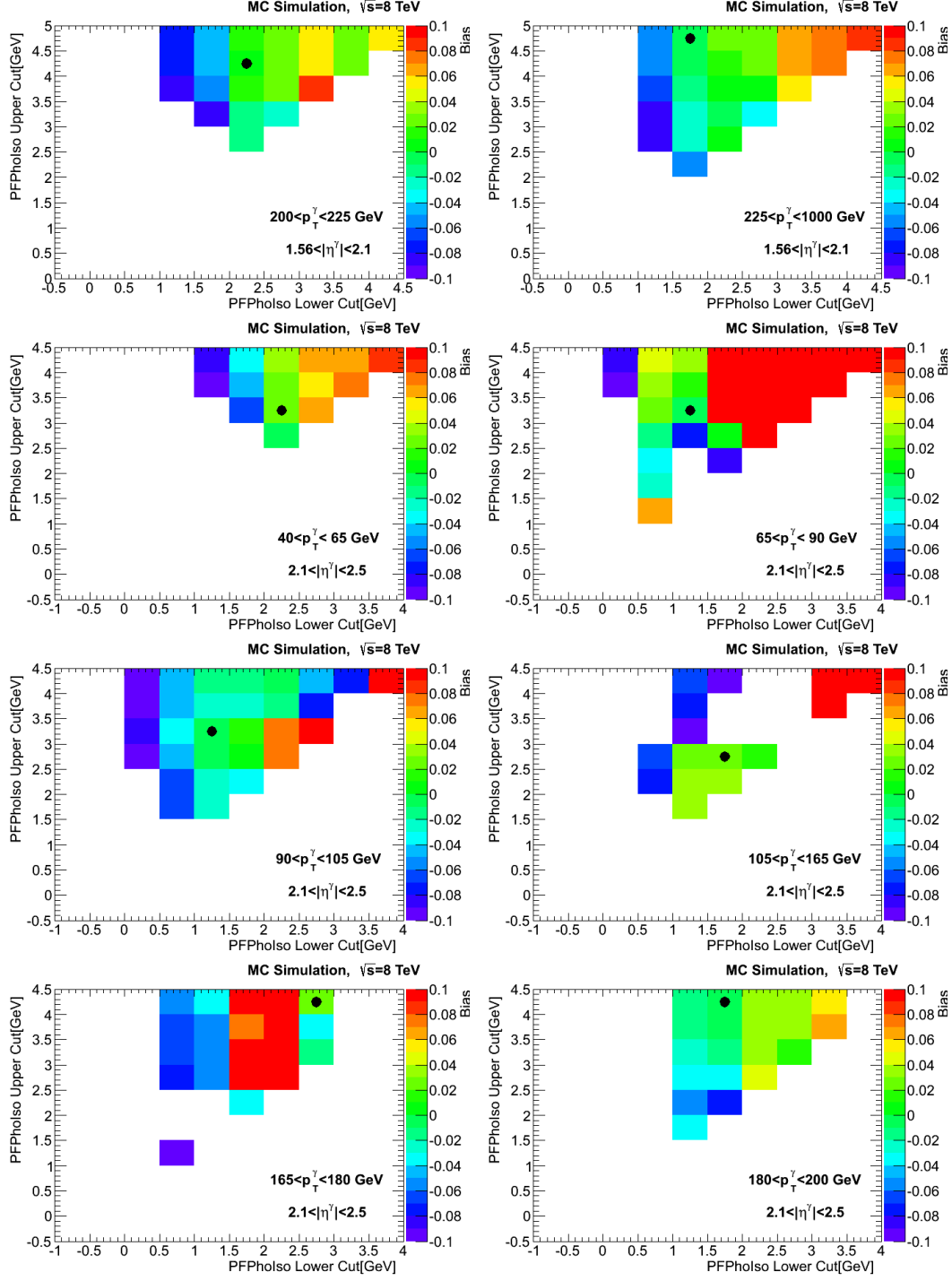


Figure B.7: Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.

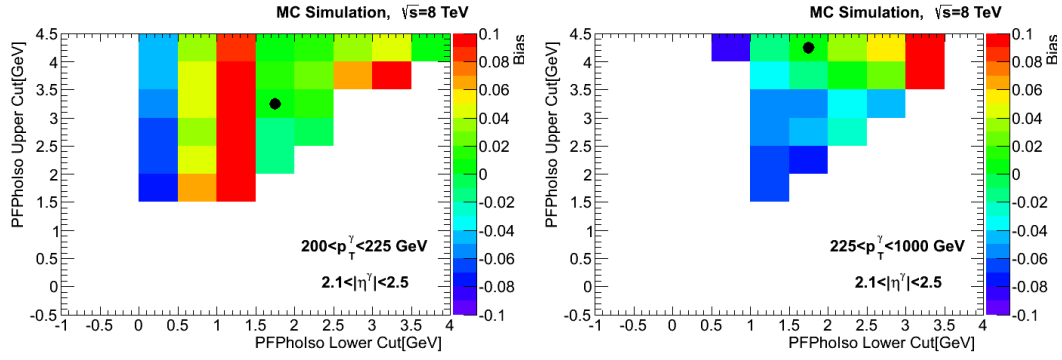


Figure B.8: Sideband variation plots. The z-axis shows bias values obtained for MC toy experiments done with tentative background templates with sideband window corresponding to the photon isolation values represented by x and y axis. Black dot represents the bin that gives the lowest bias. The photon isolation cuts corresponding to this bin is used to construct background template.

APPENDIX C

PURITY FITS

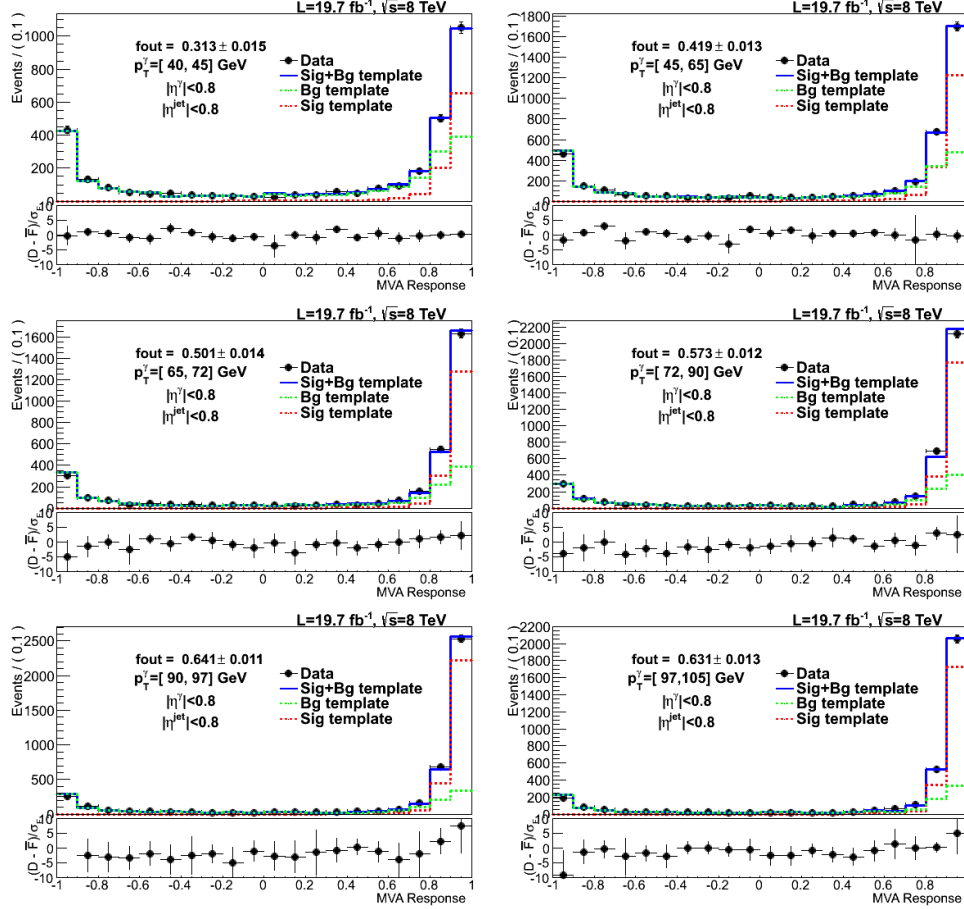


Figure C.1: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

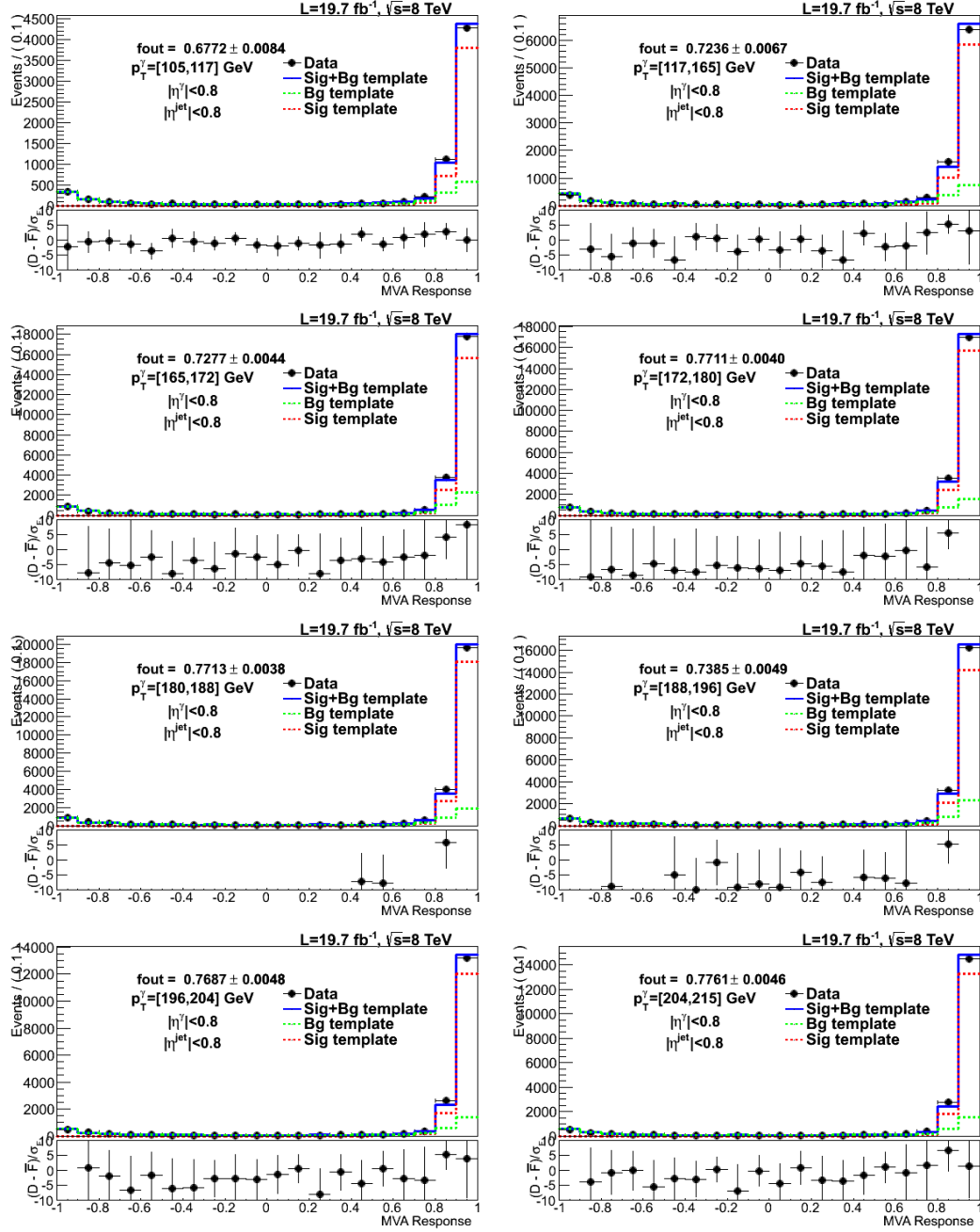


Figure C.2: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

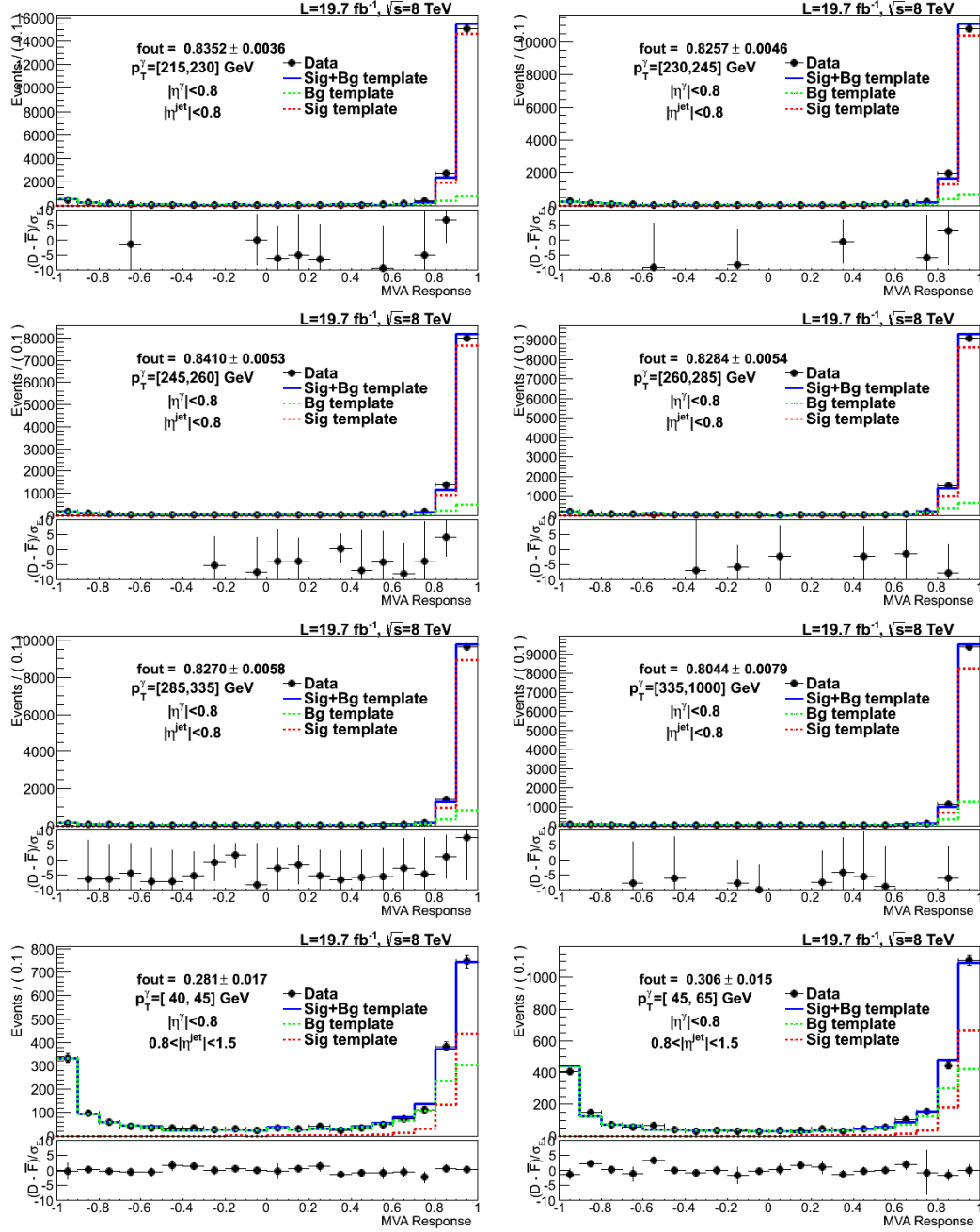


Figure C.3: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

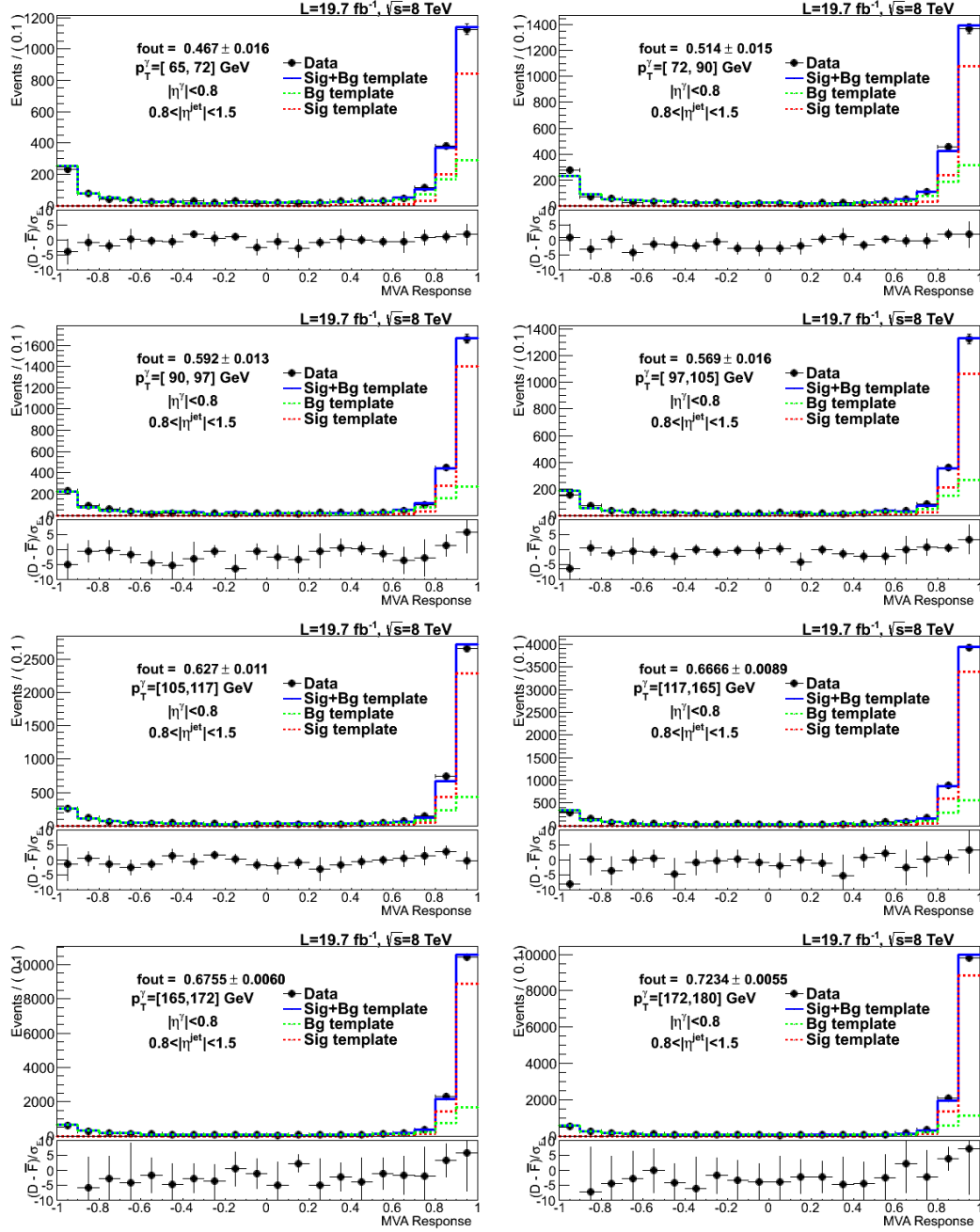


Figure C.4: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

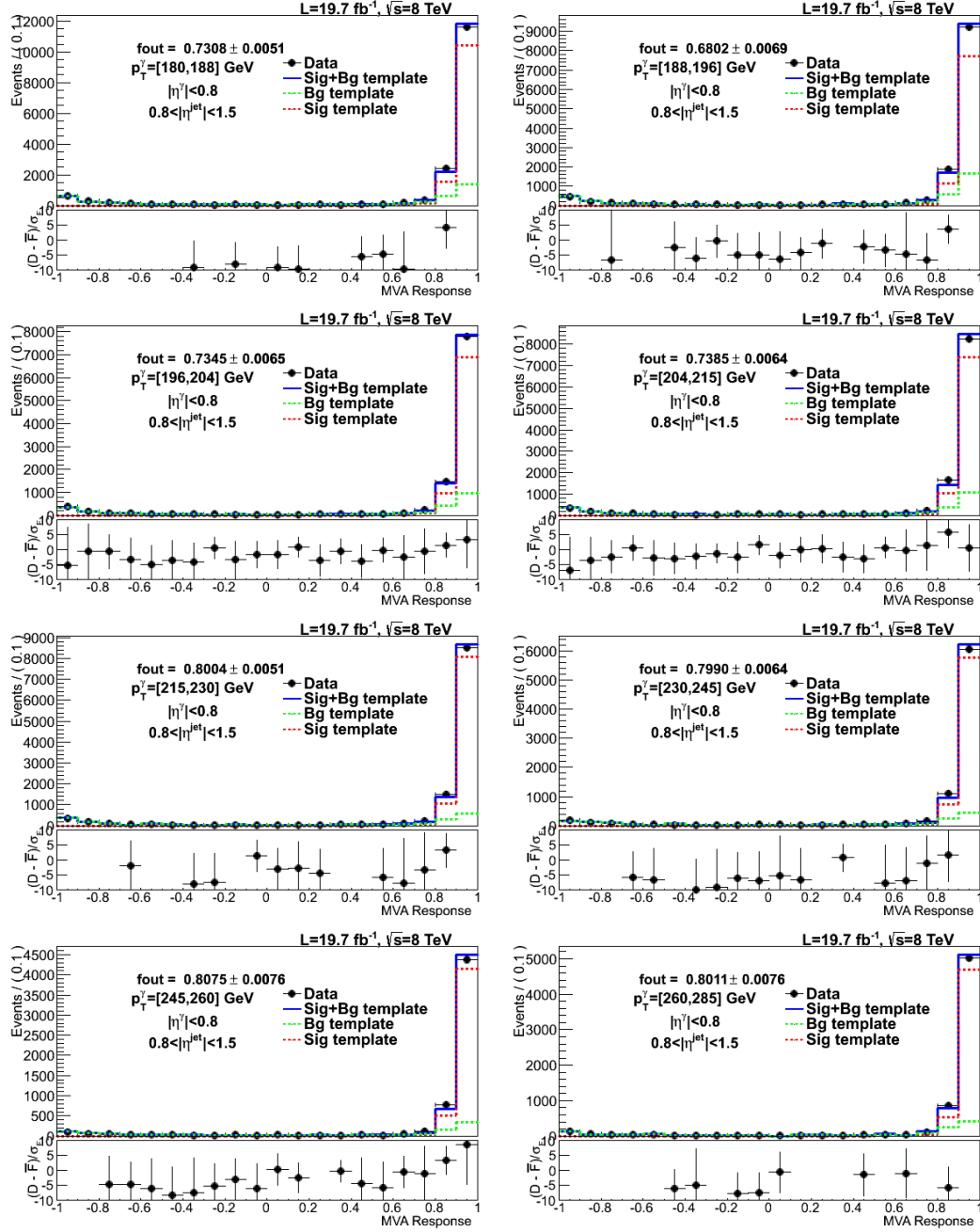


Figure C.5: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

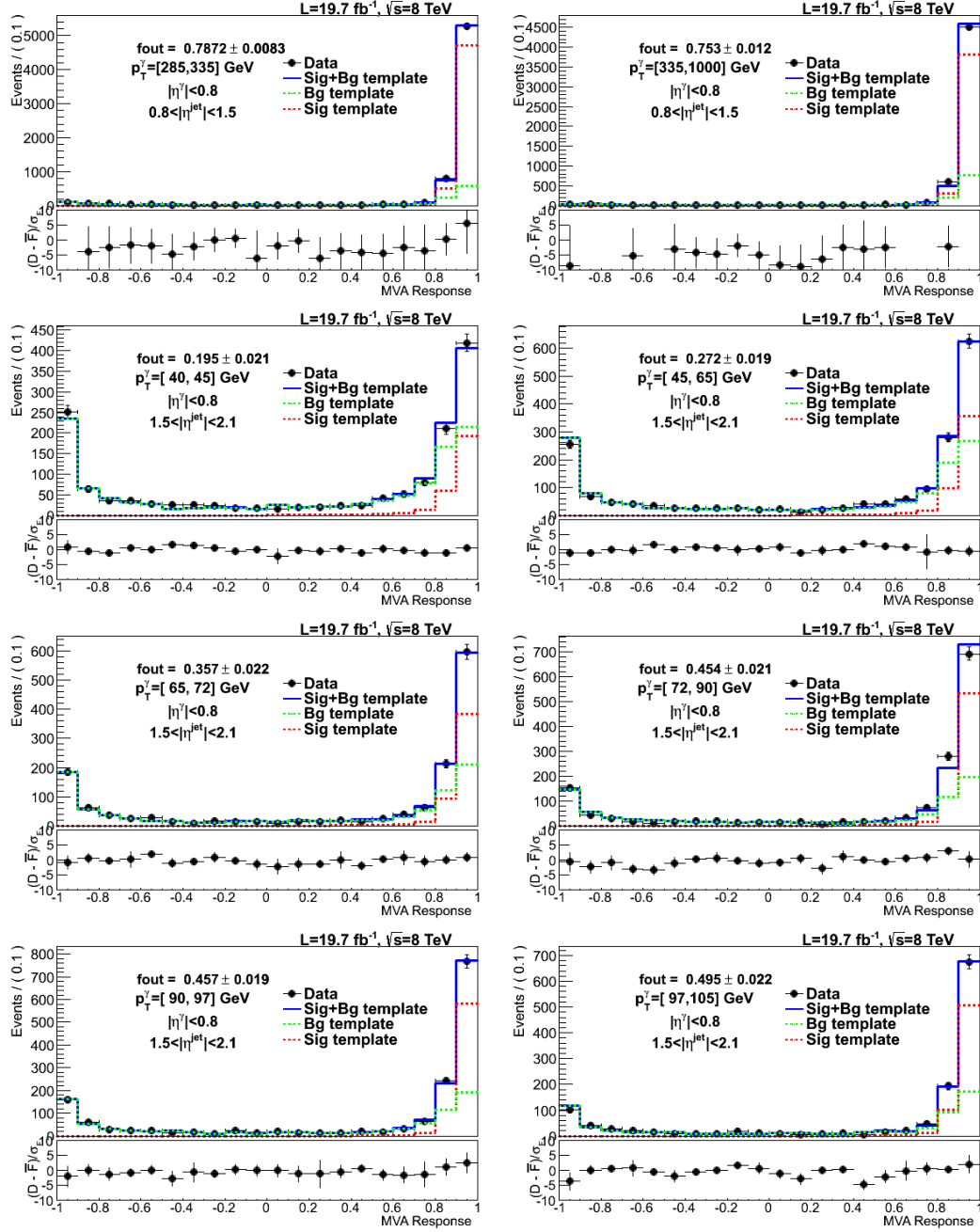


Figure C.6: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

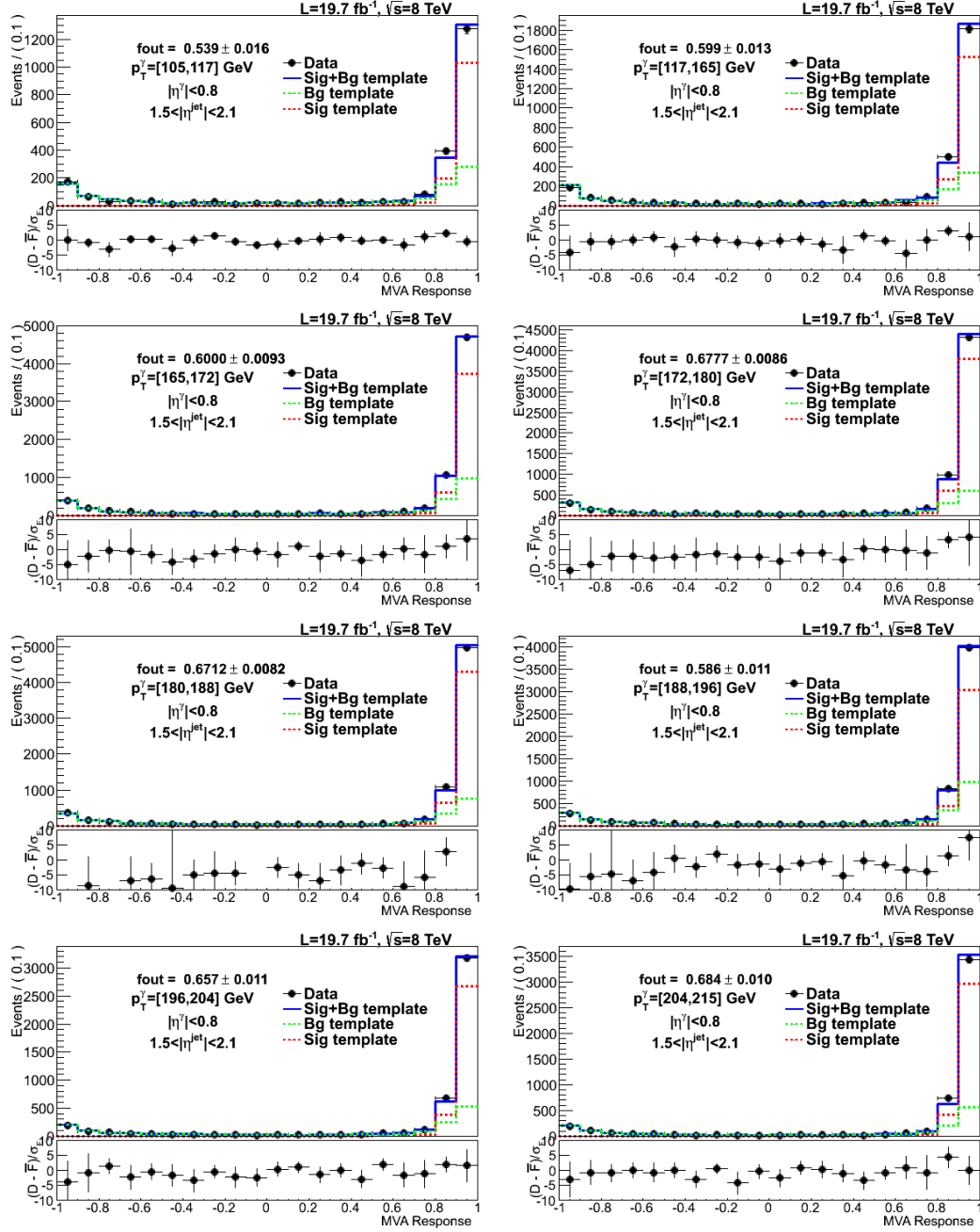


Figure C.7: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

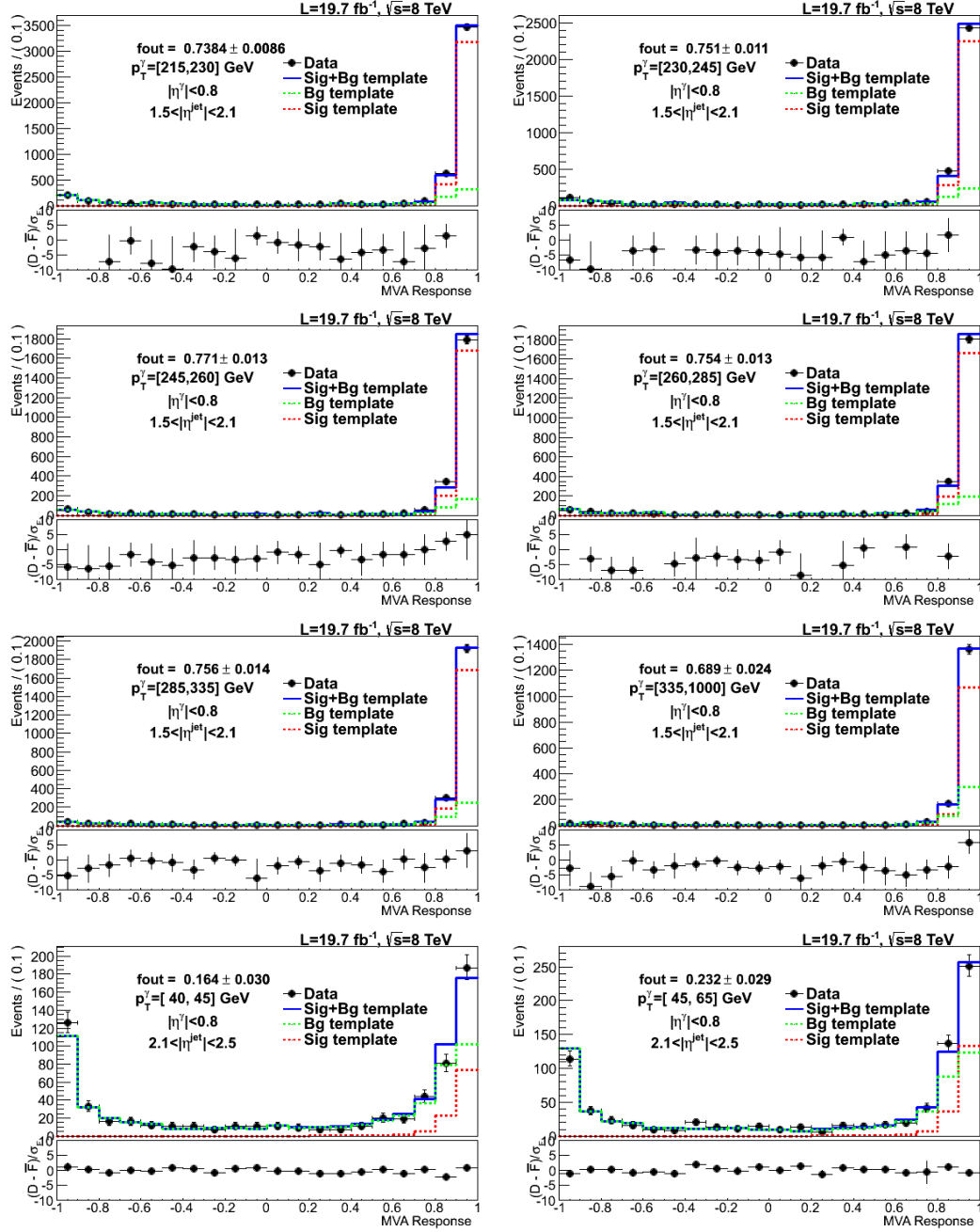


Figure C.8: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

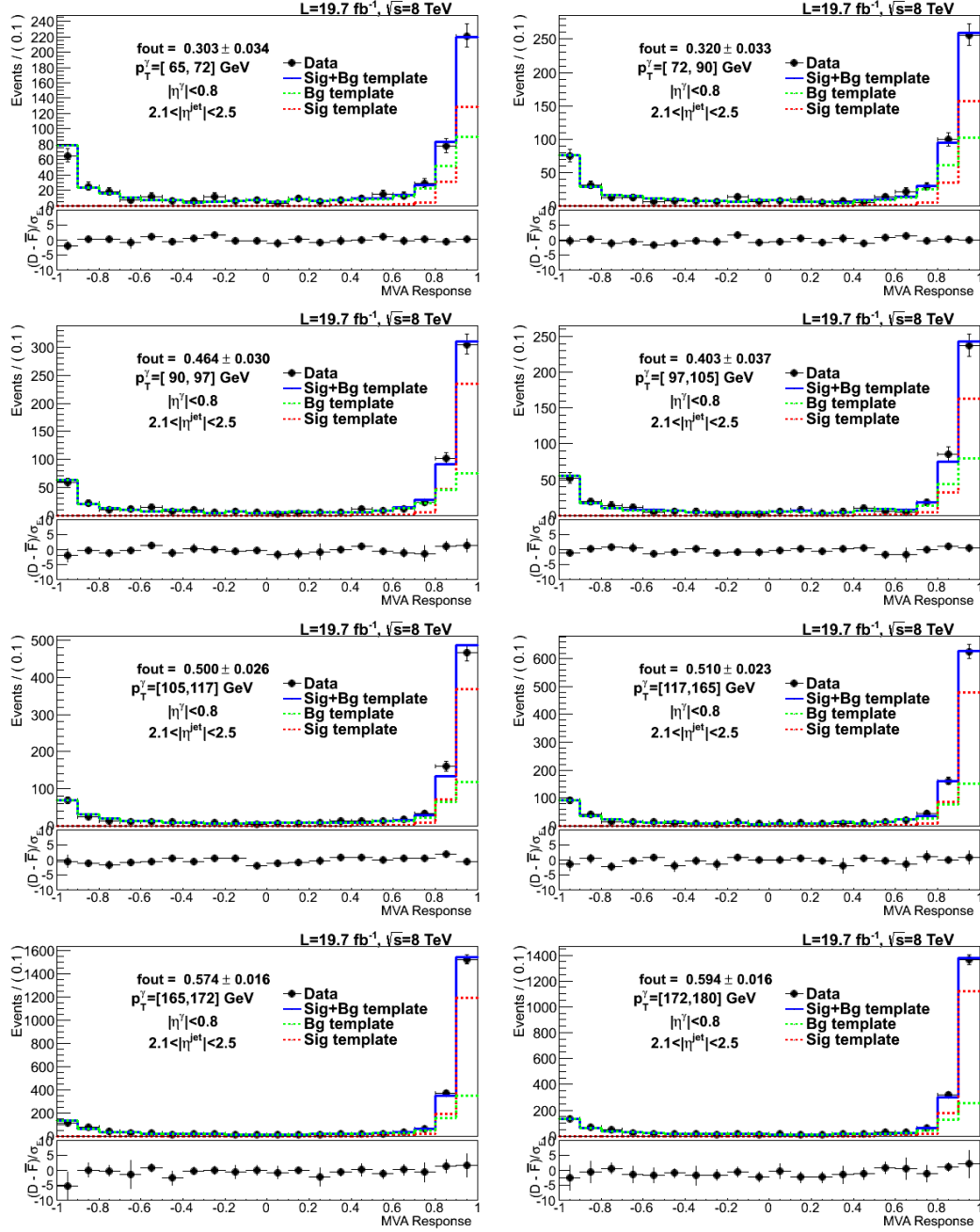


Figure C.9: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

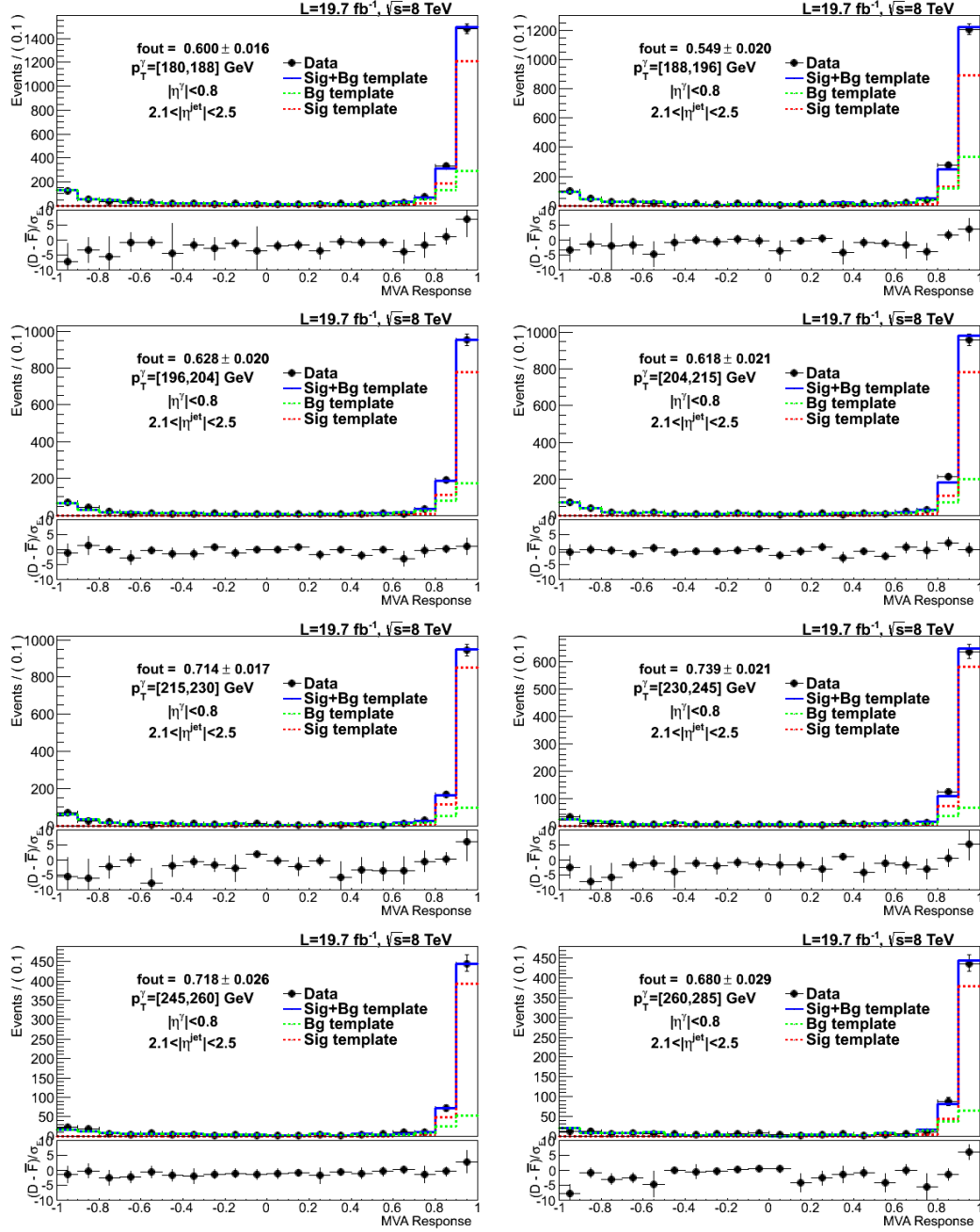


Figure C.10: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

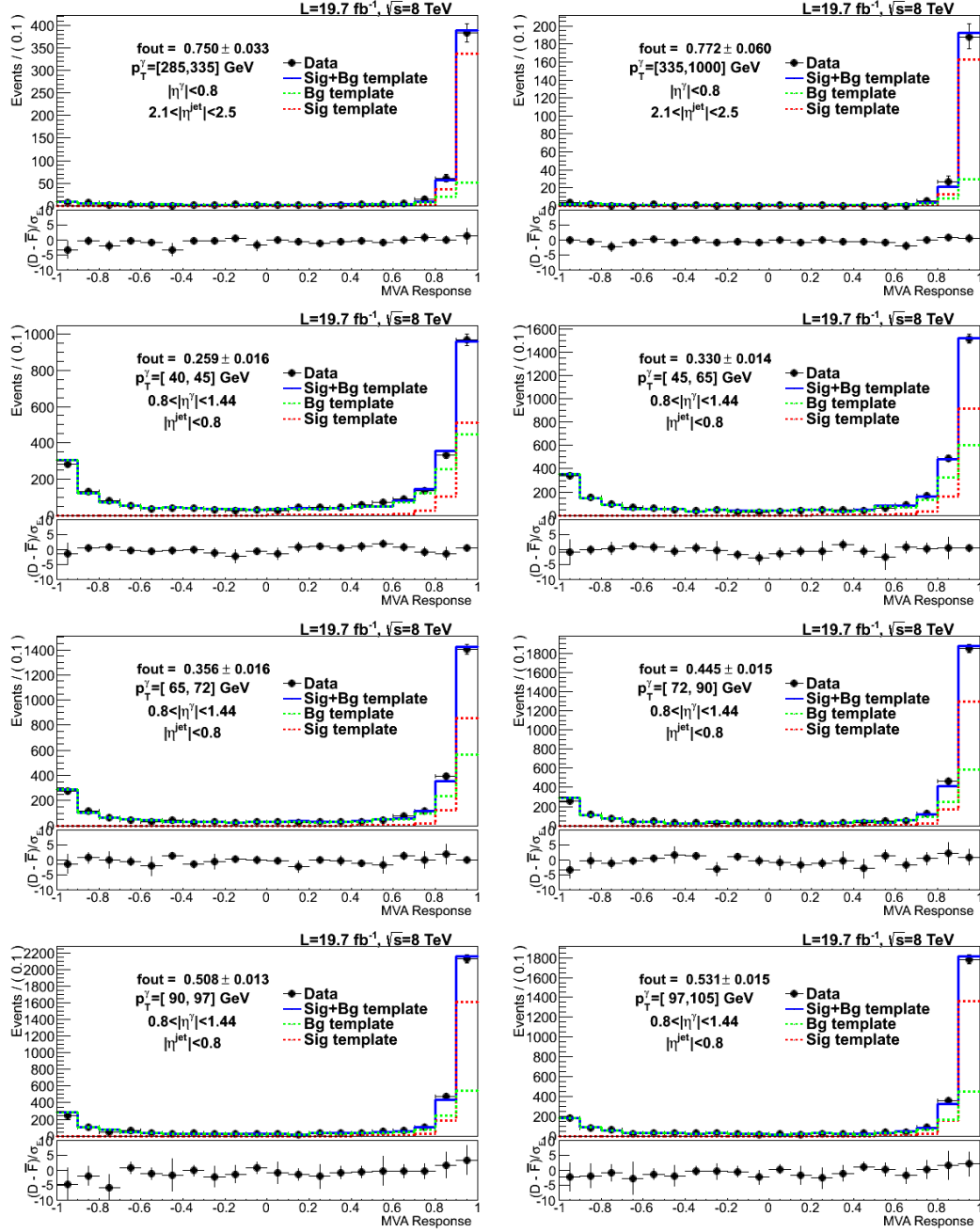


Figure C.11: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

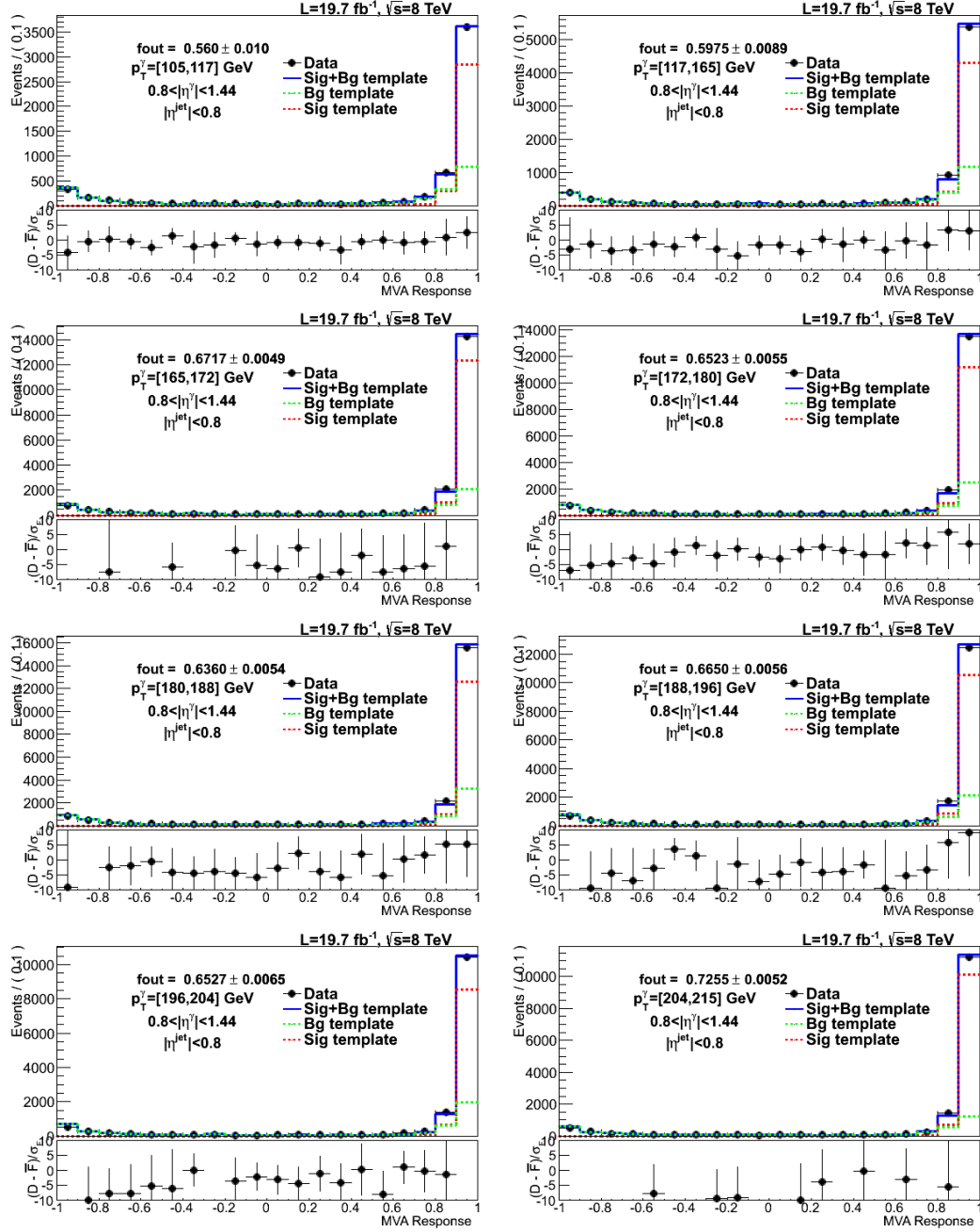


Figure C.12: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

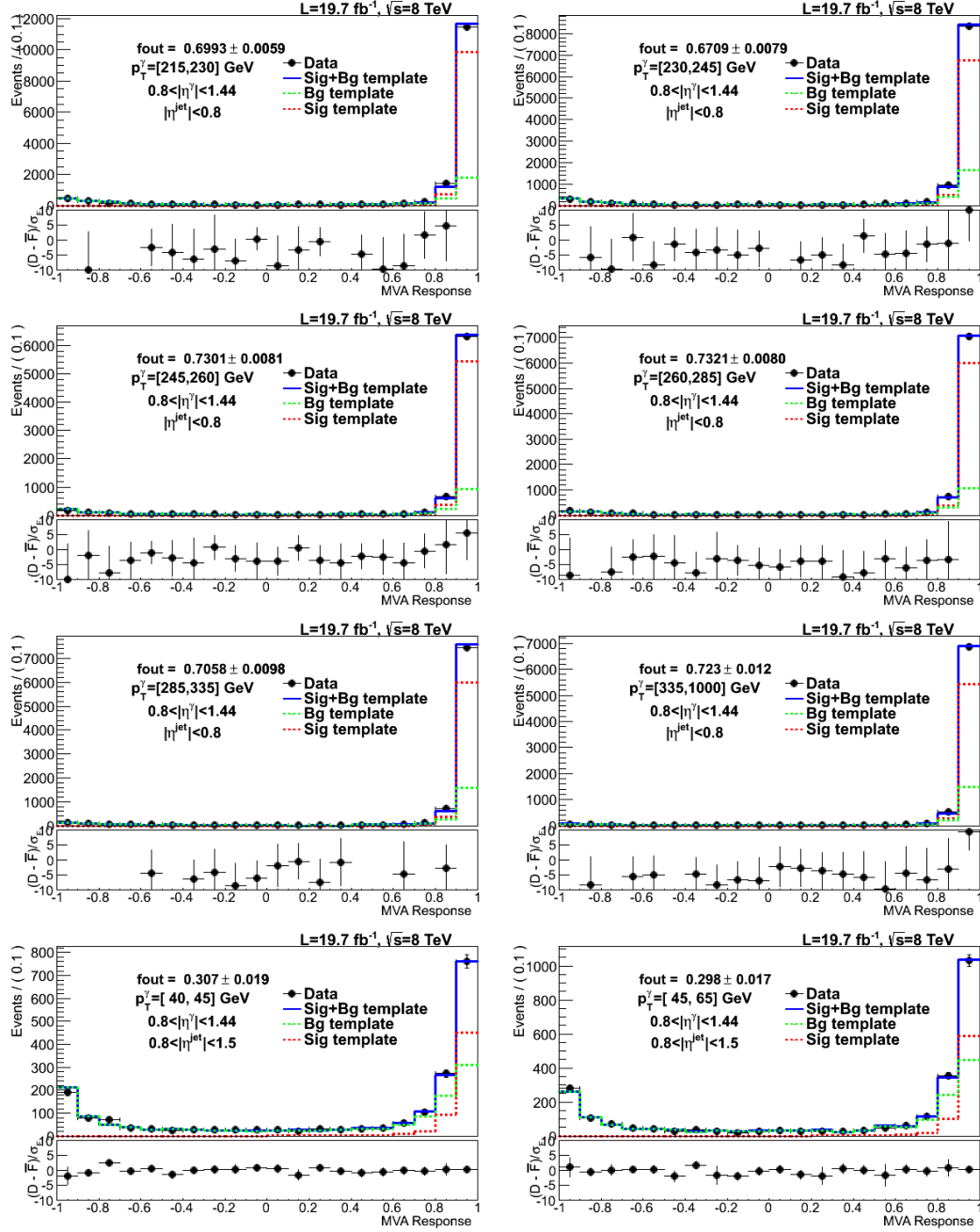


Figure C.13: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

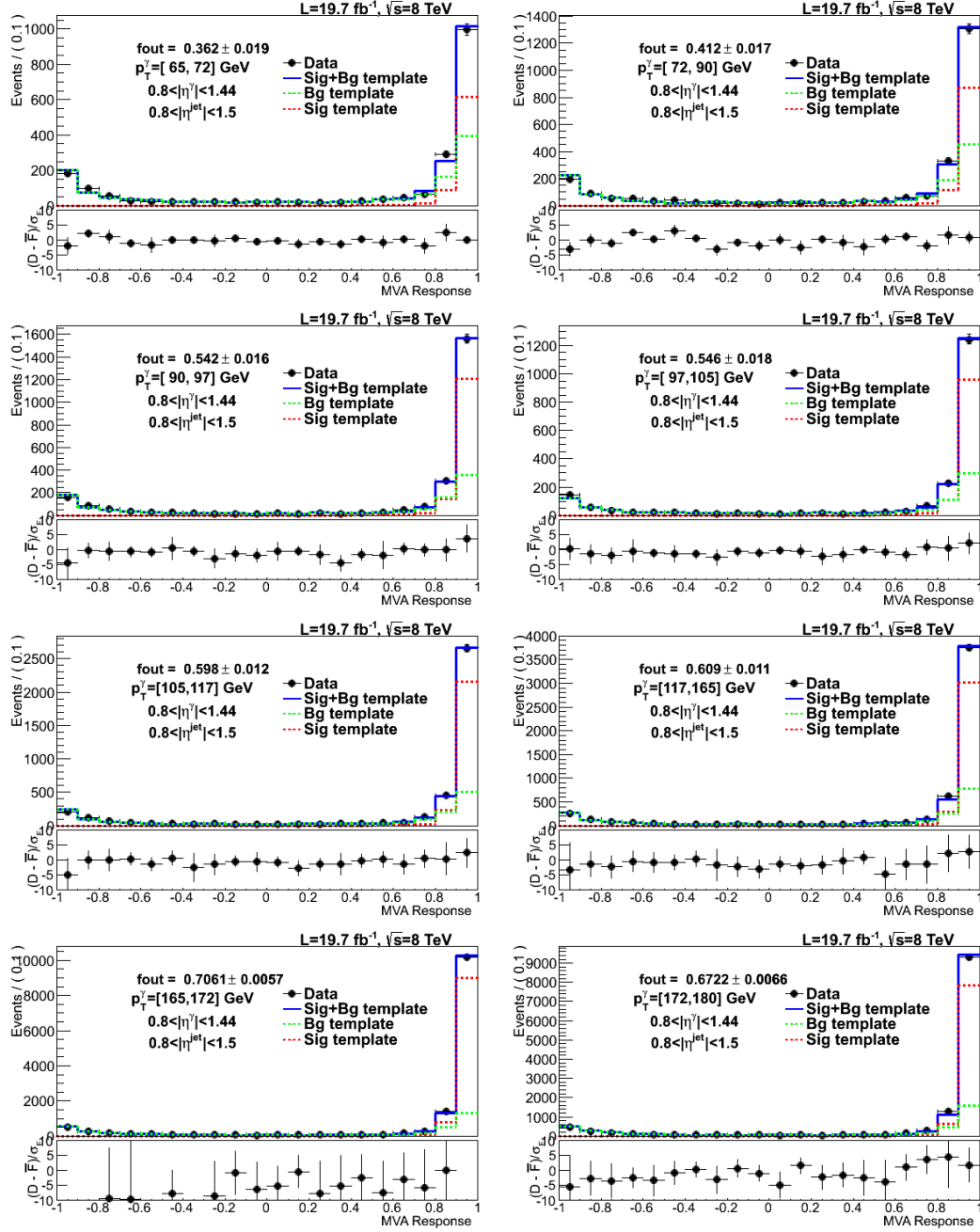


Figure C.14: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

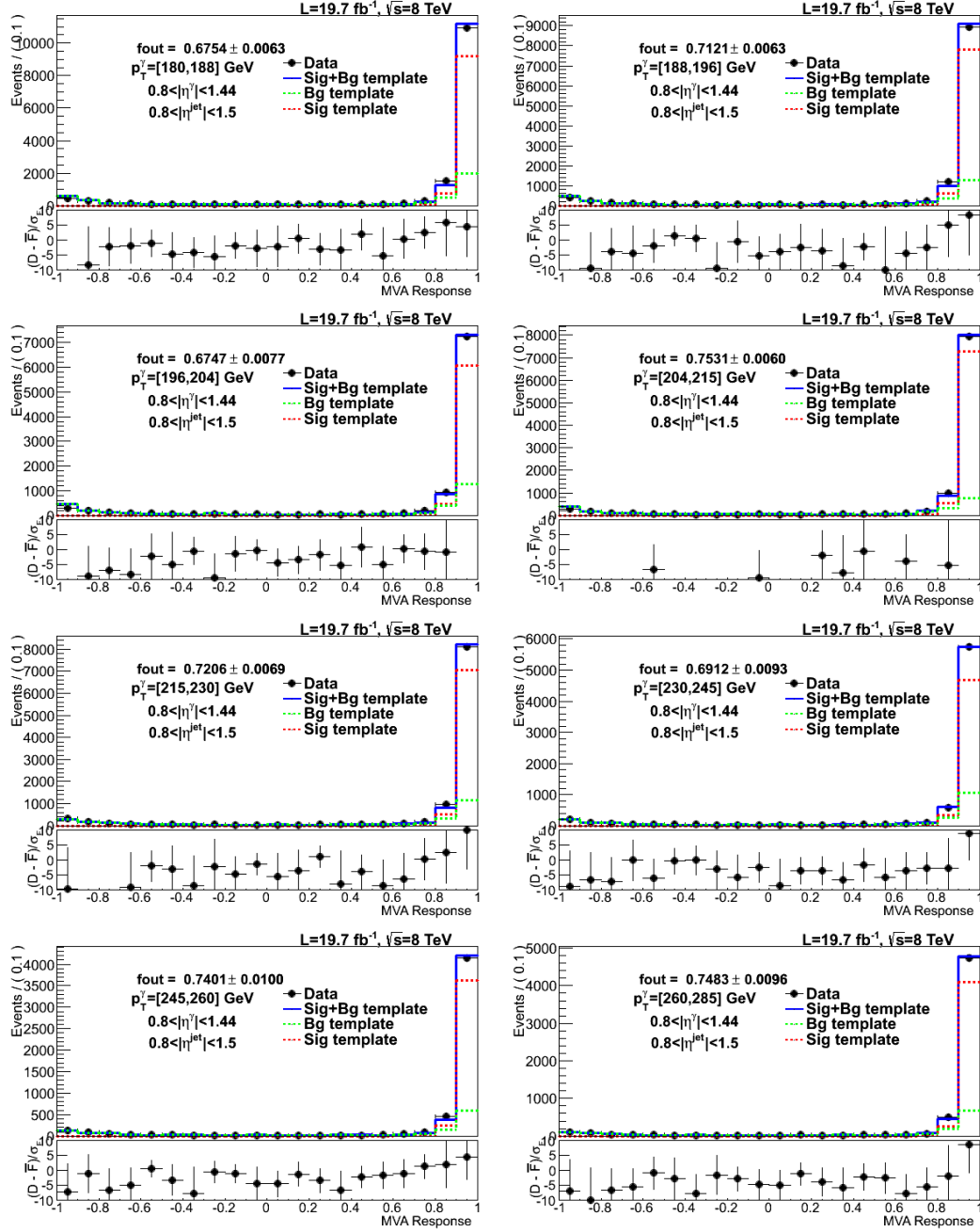


Figure C.15: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

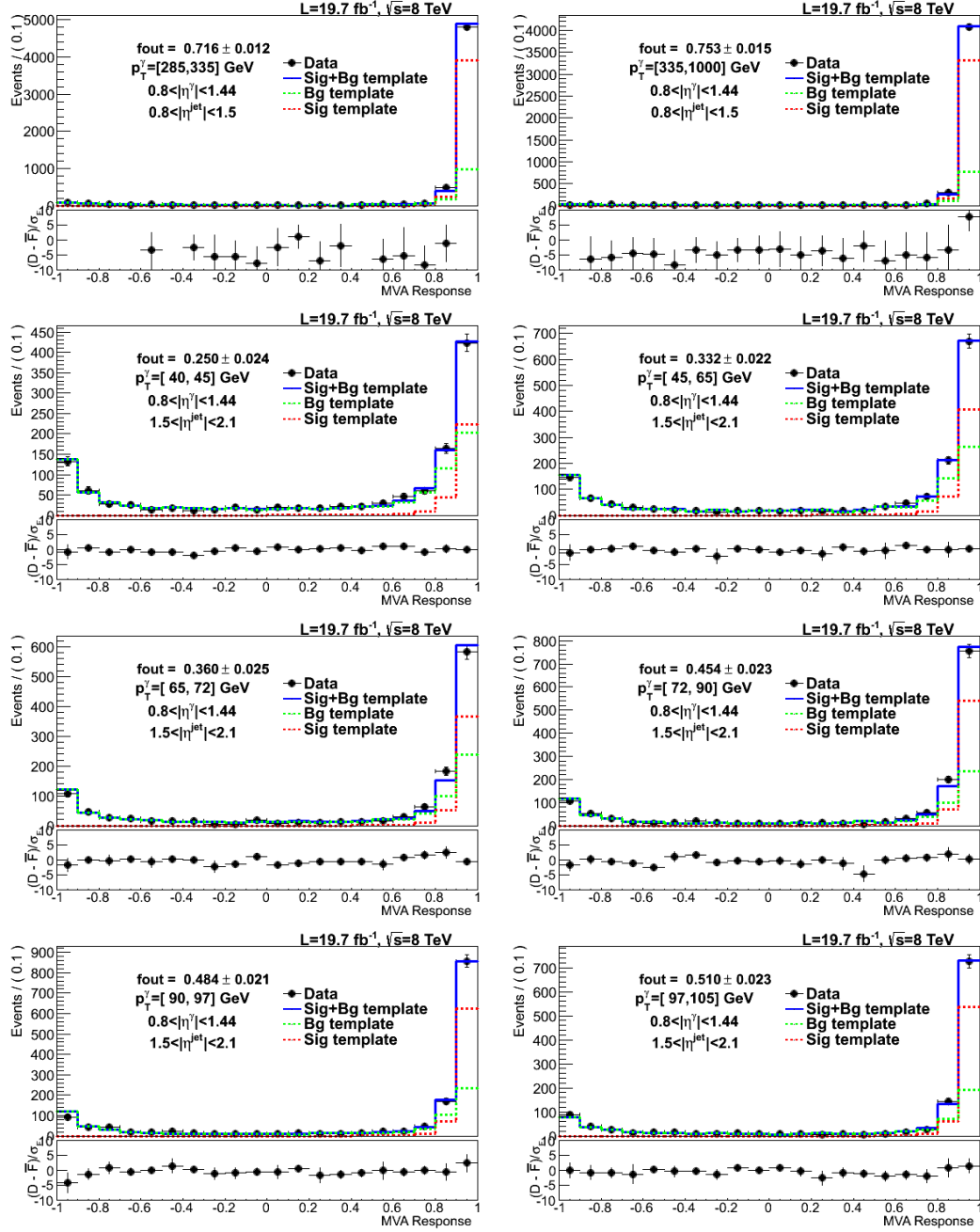


Figure C.16: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

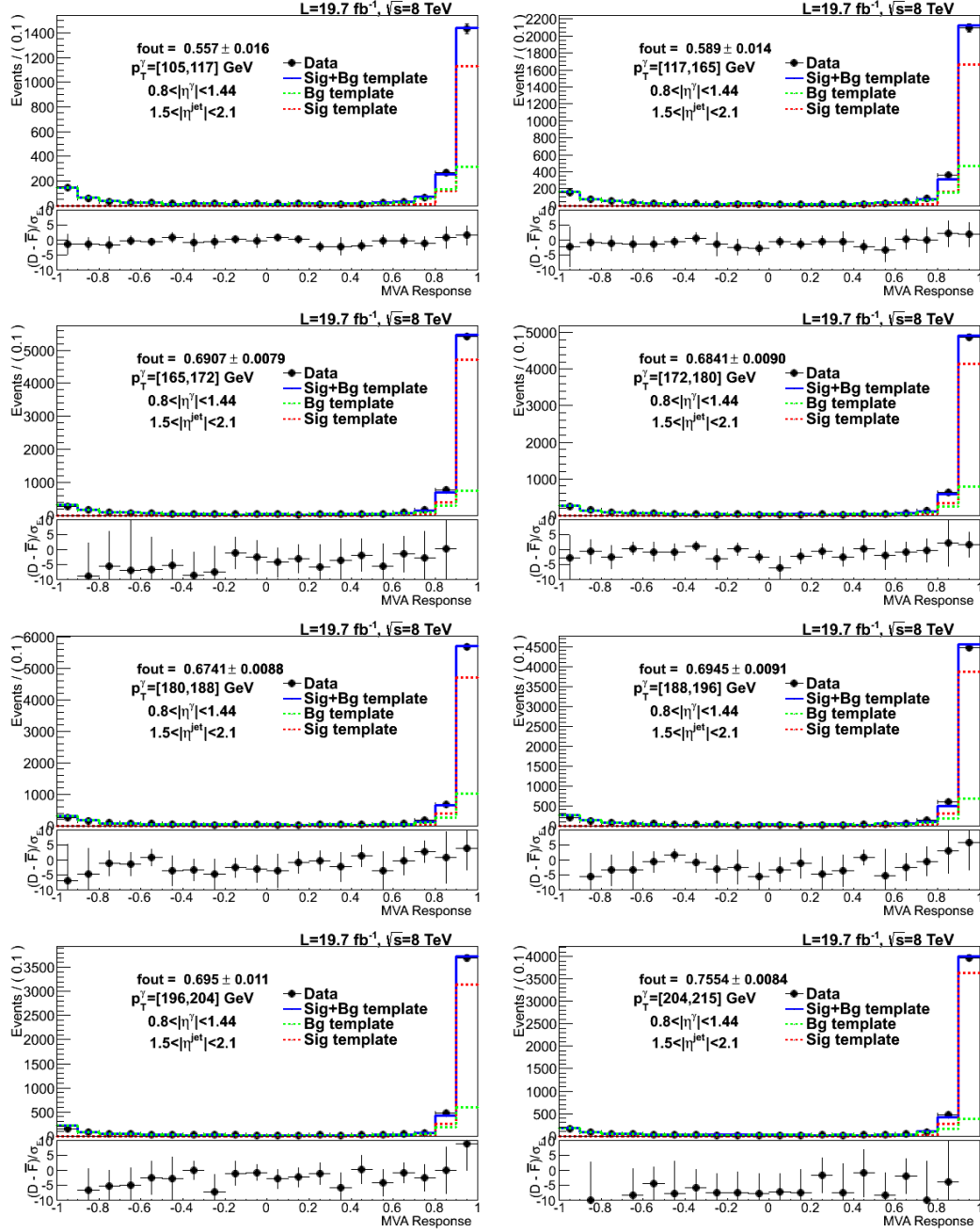


Figure C.17: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

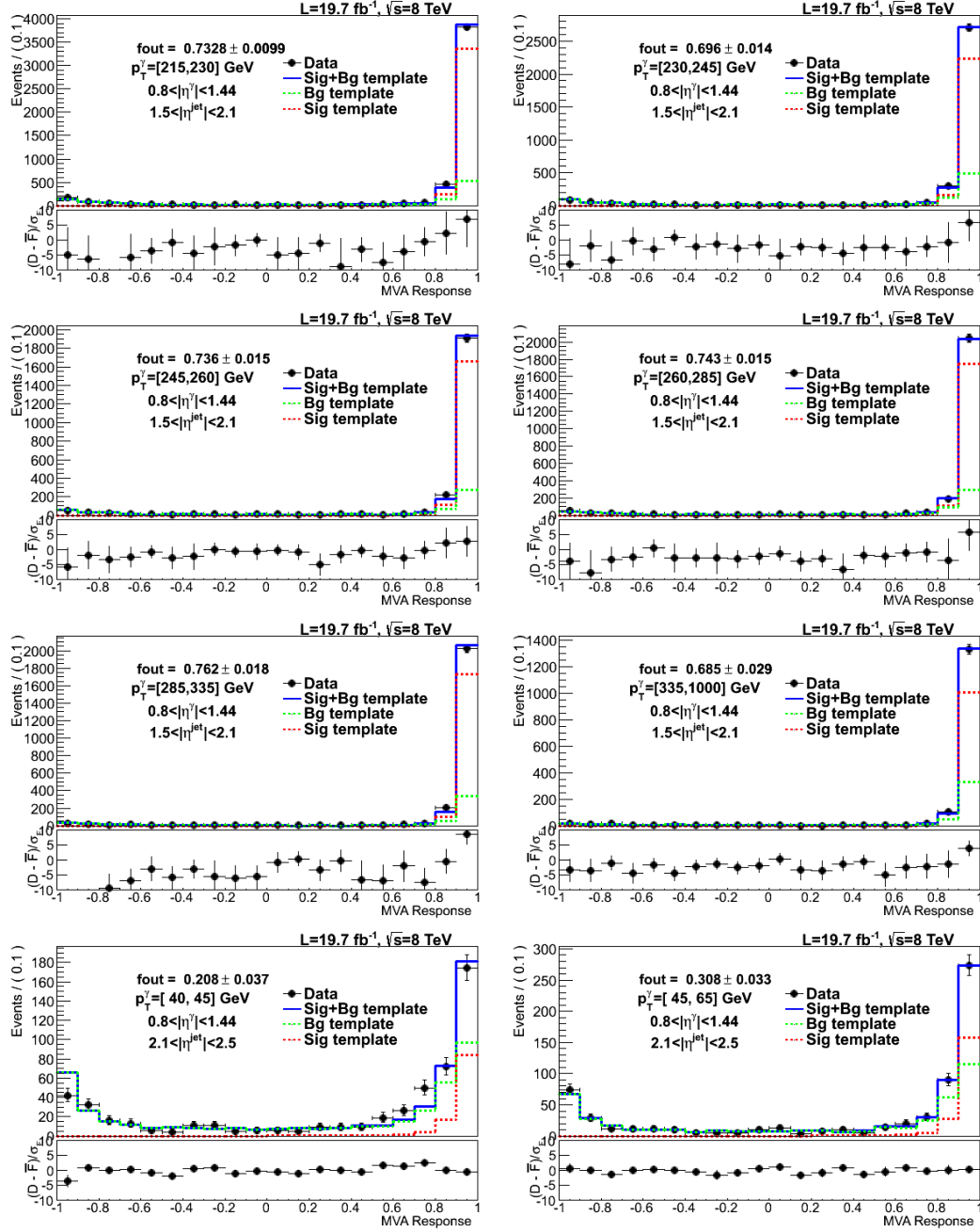


Figure C.18: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

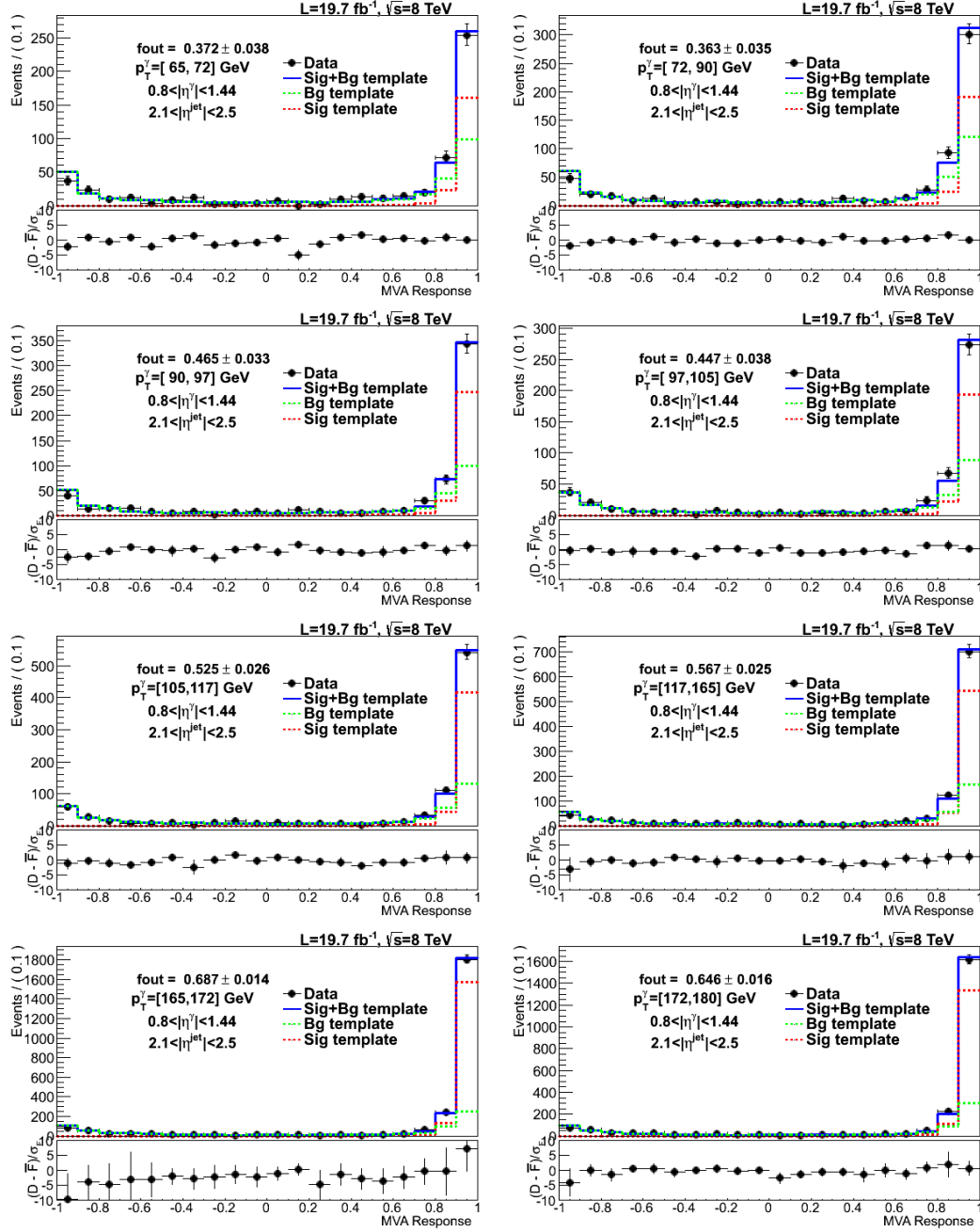


Figure C.19: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

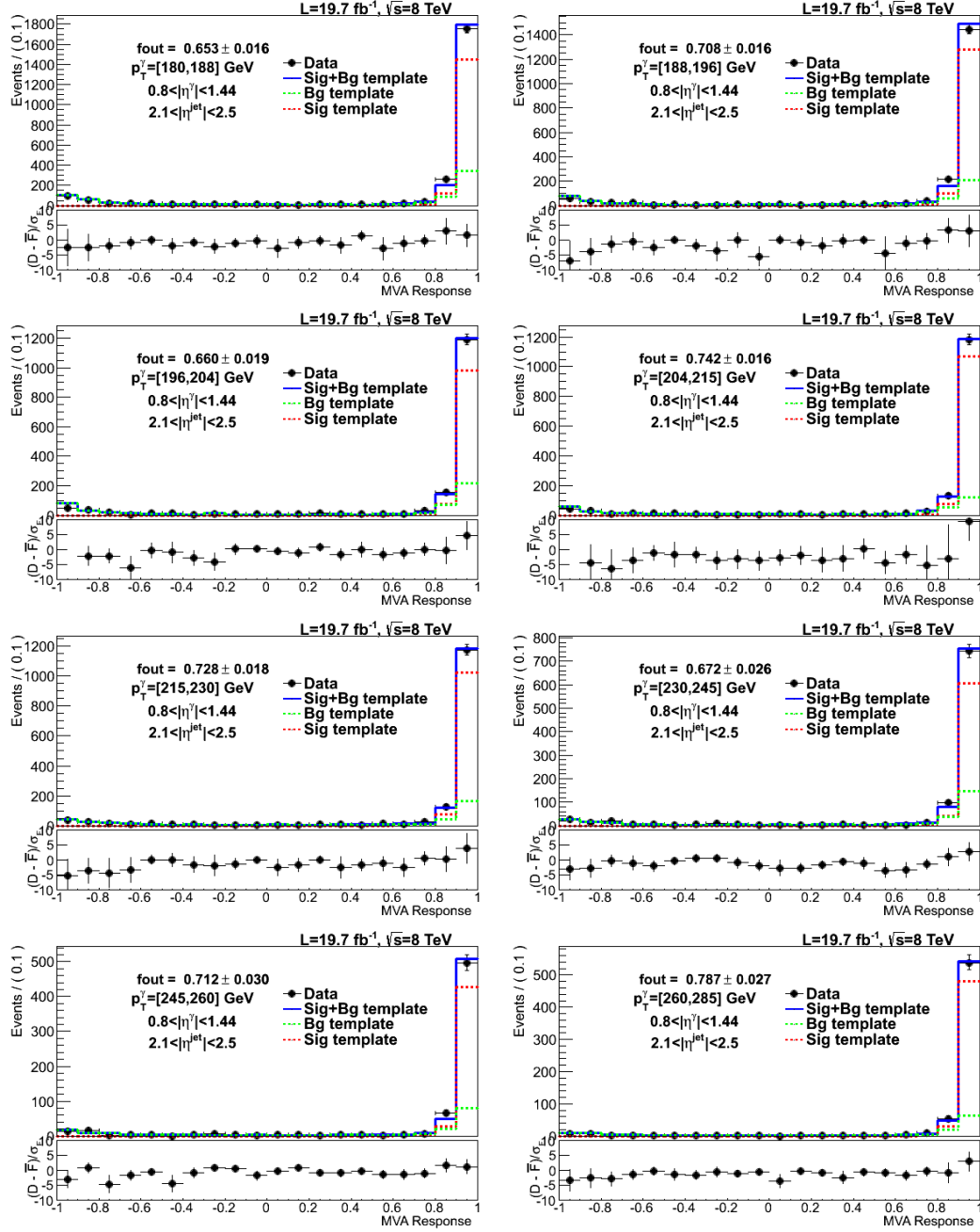


Figure C.20: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

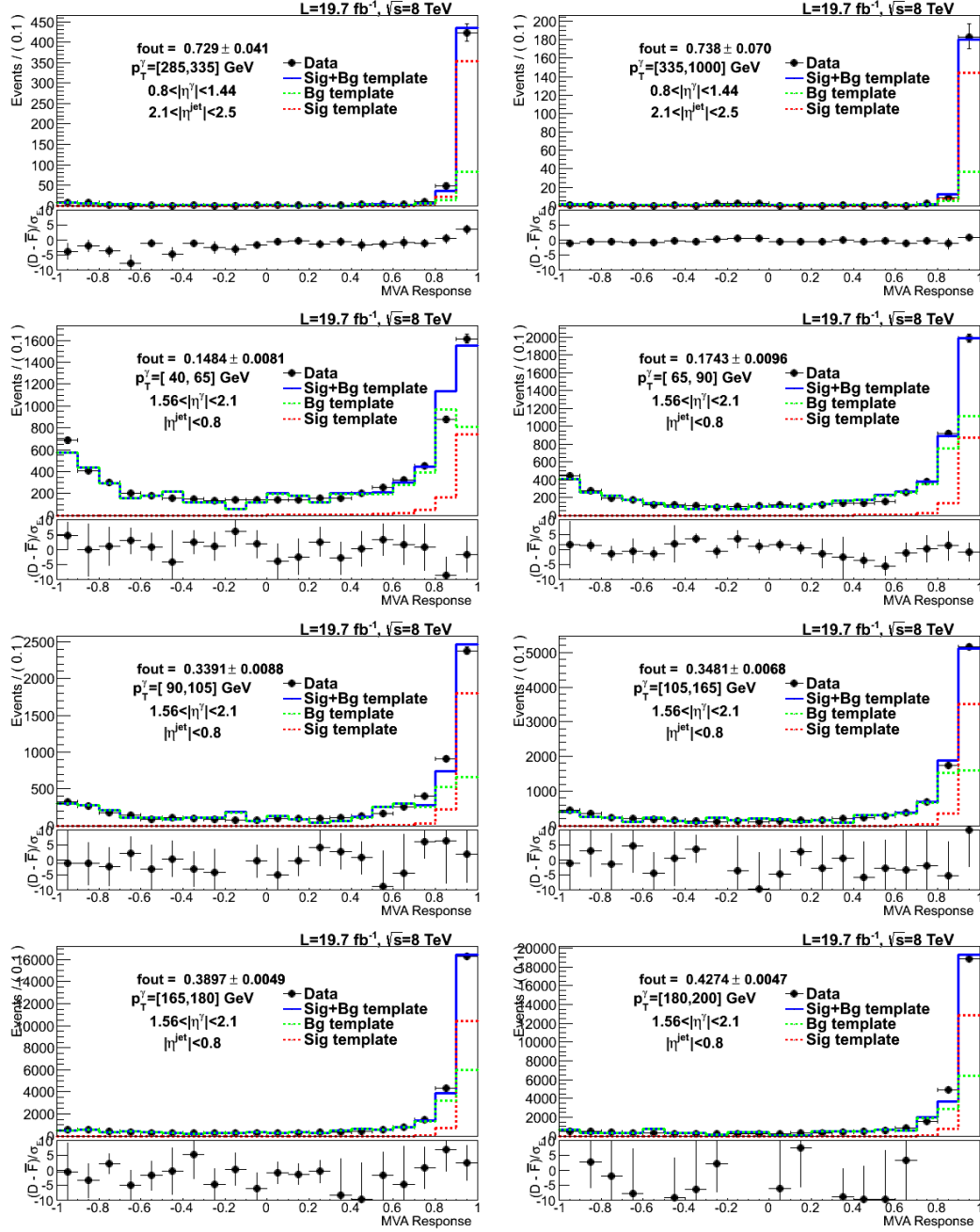


Figure C.21: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

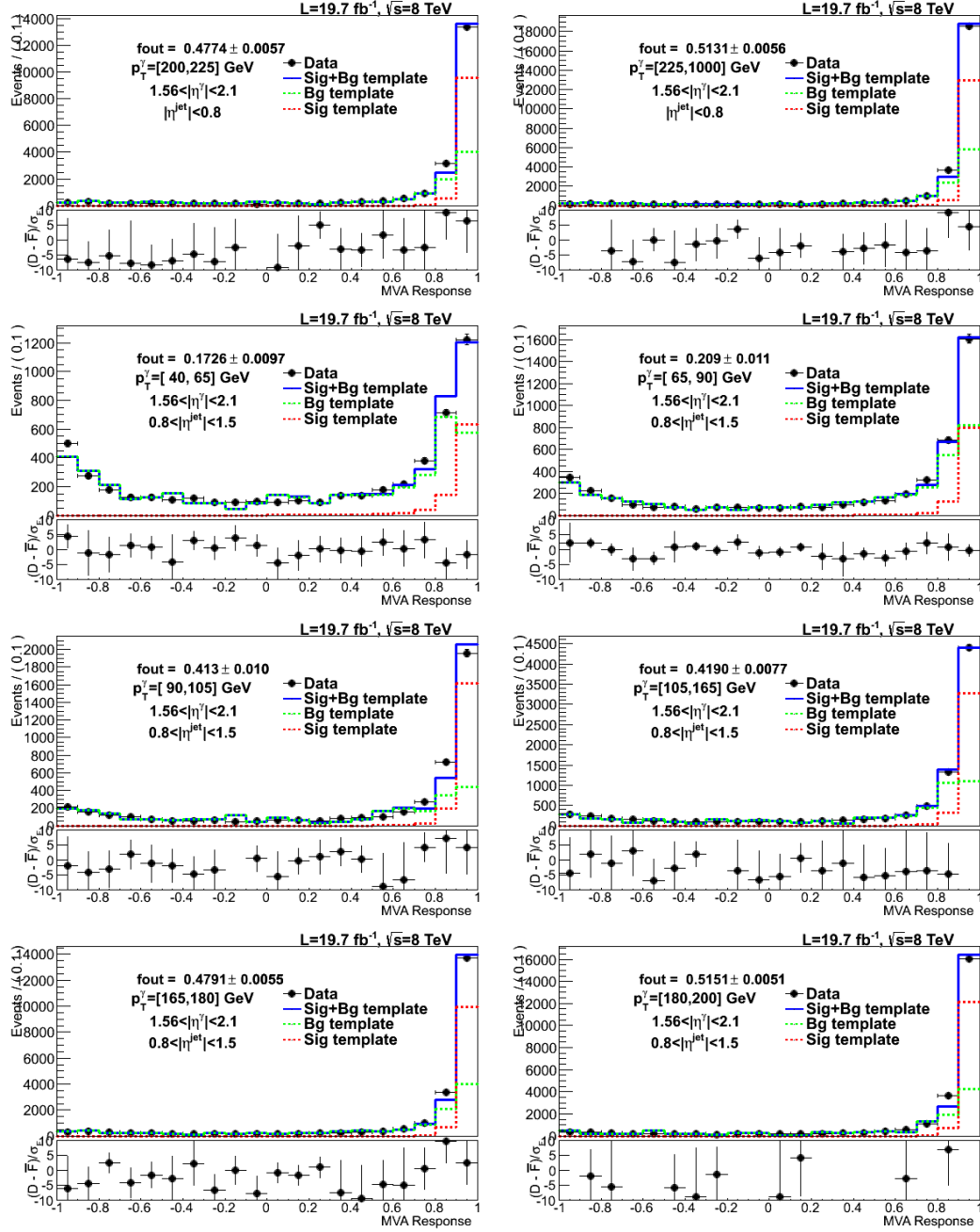


Figure C.22: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

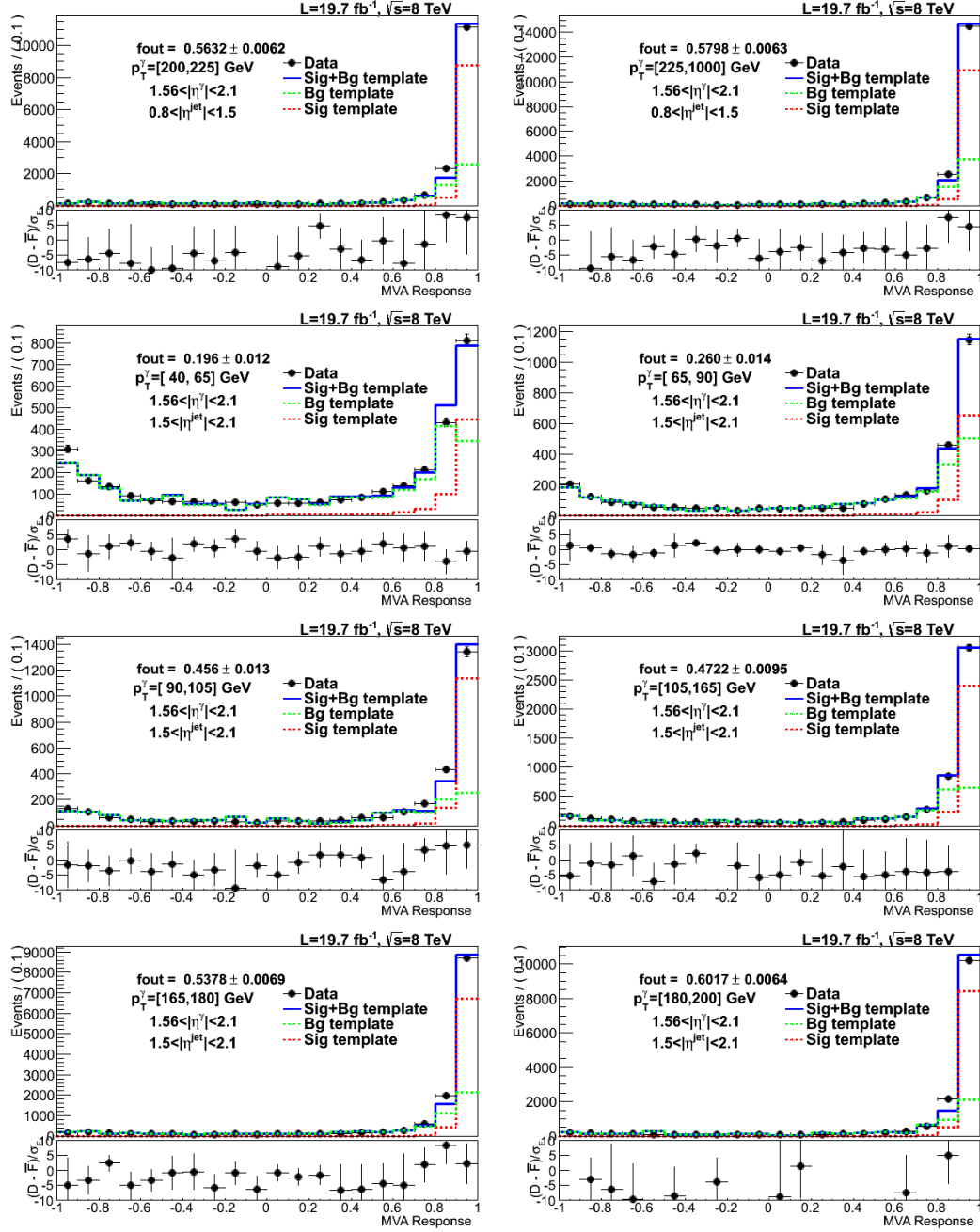


Figure C.23: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

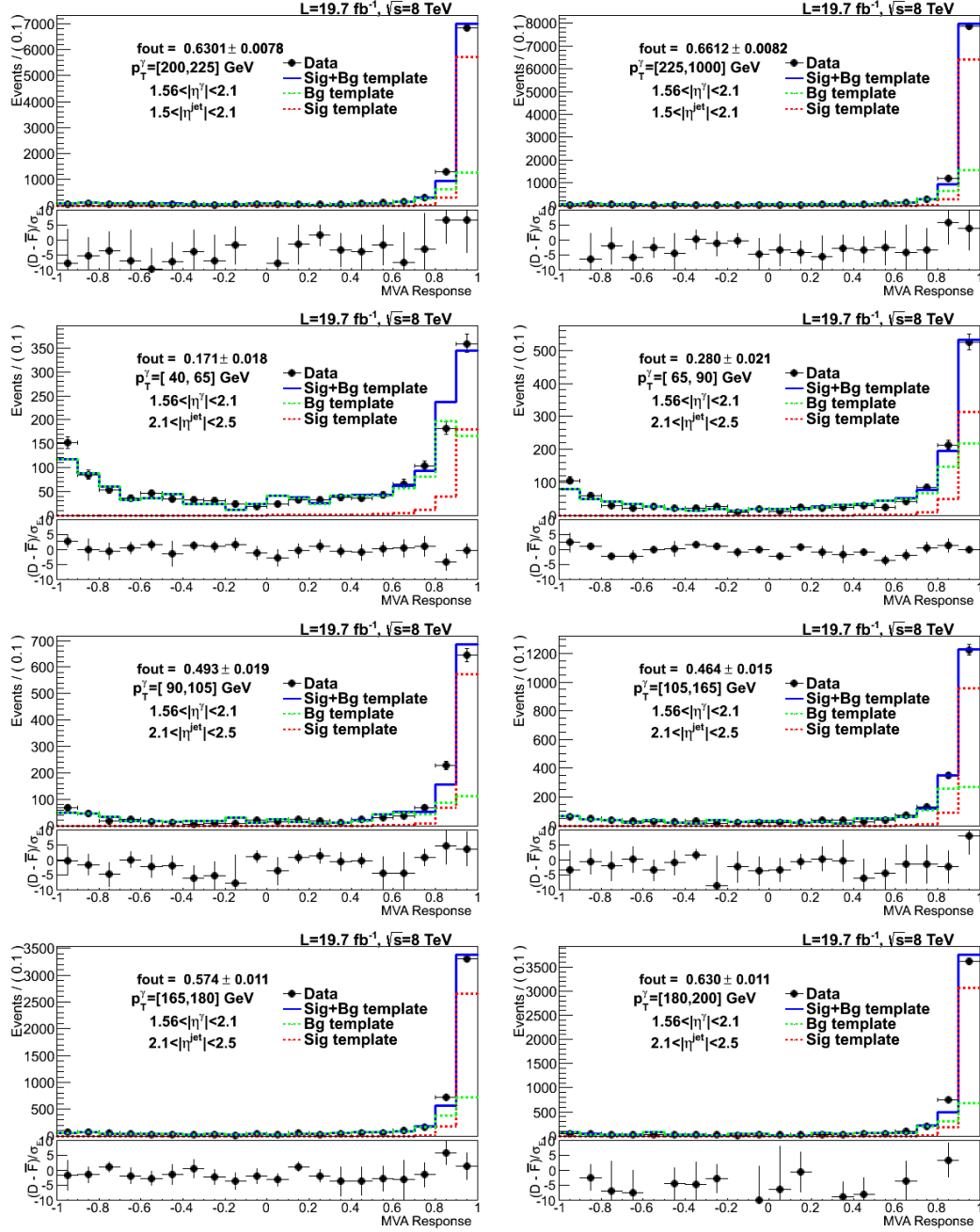


Figure C.24: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

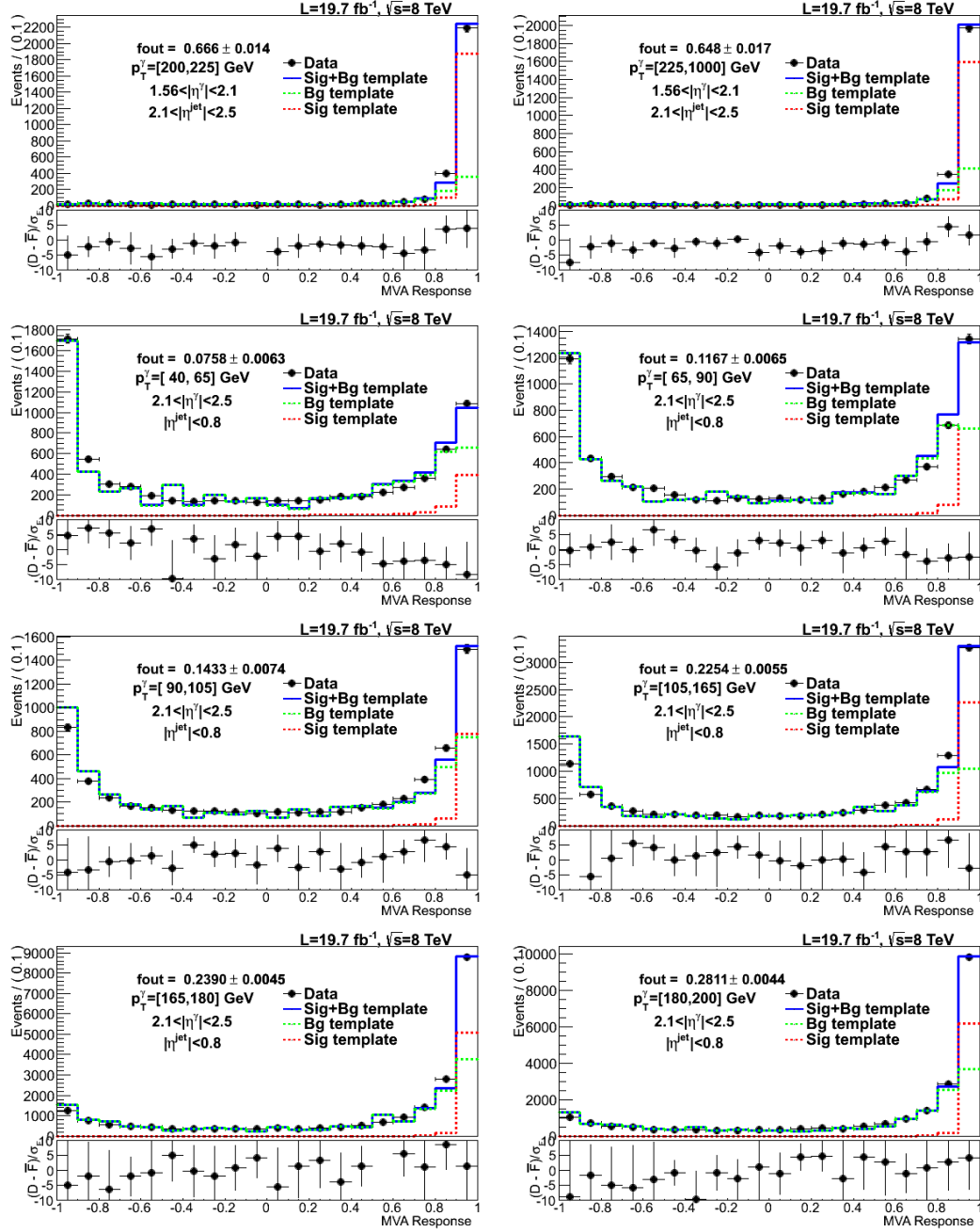


Figure C.25: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

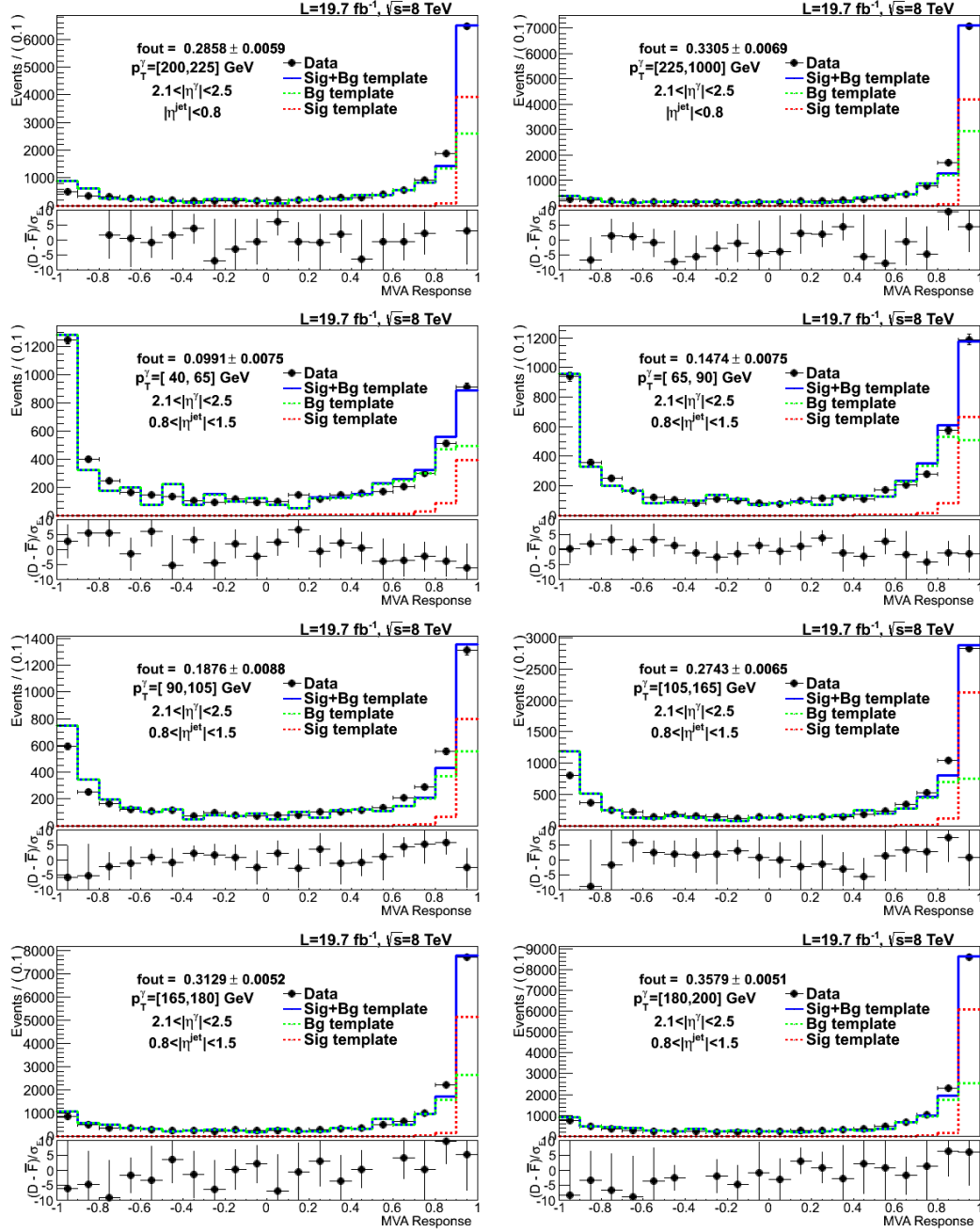


Figure C.26: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

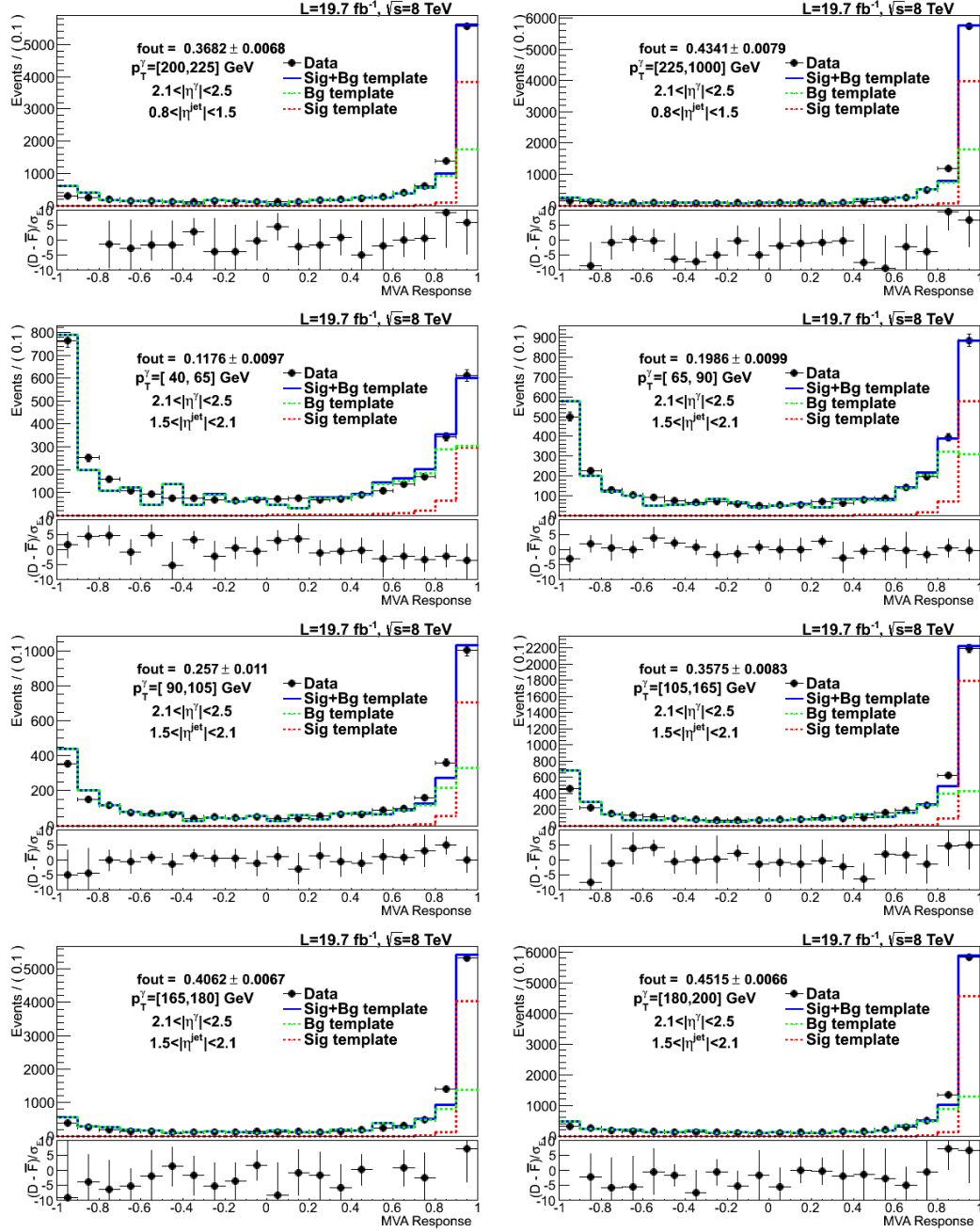


Figure C.27: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

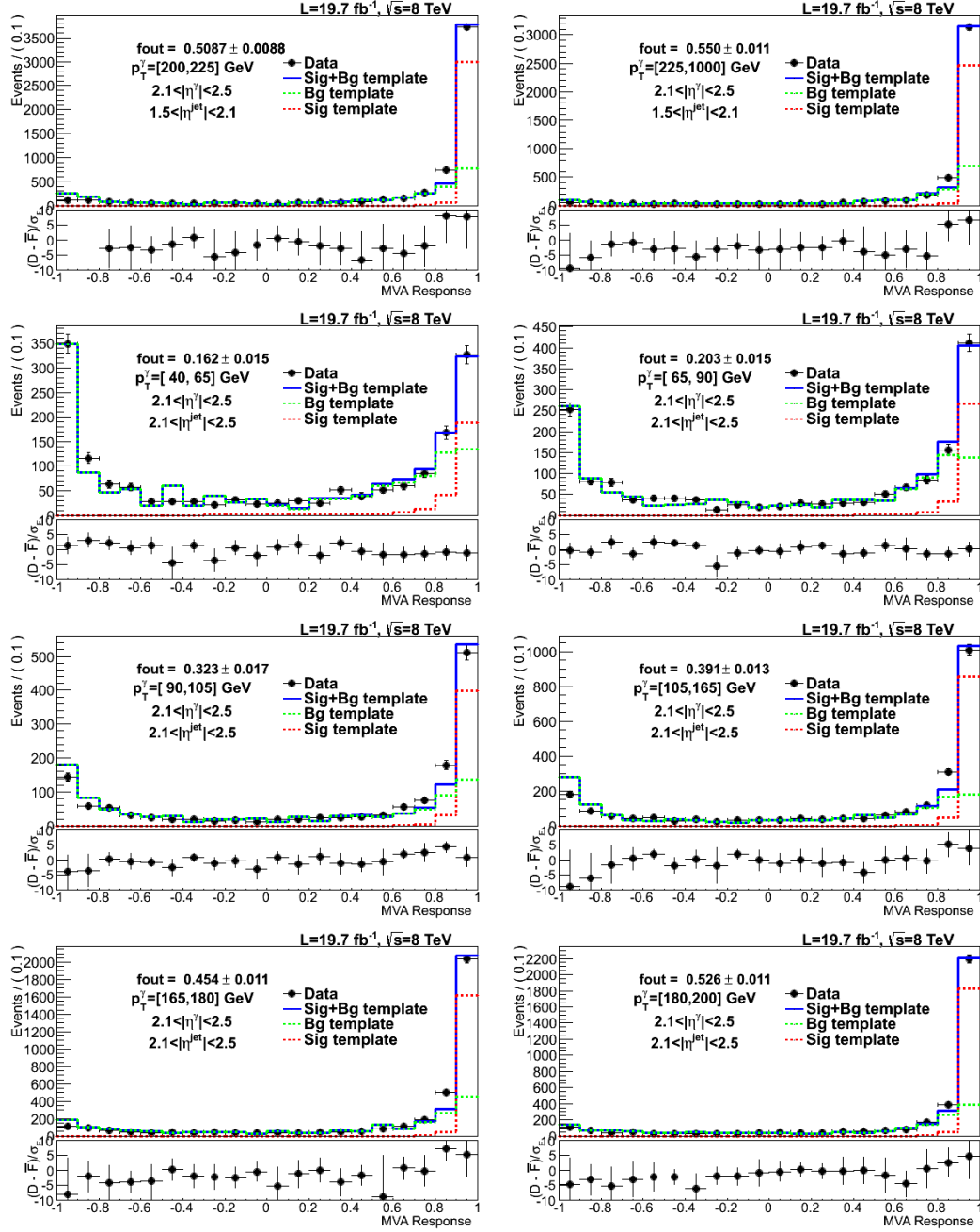


Figure C.28: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

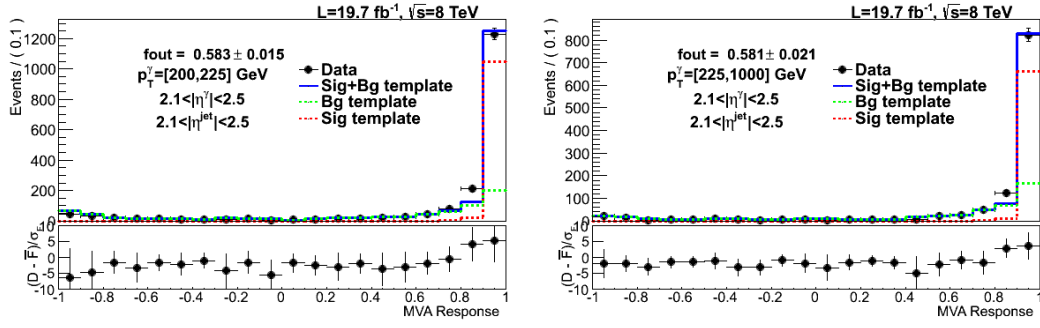


Figure C.29: The purity fits before applying bias correction. Red dotted lines represent the signal template, green dotted lines show the background template, blue dotted lines show the sum template, and the black markers represent the candidate data events. The residual plots are calculated bin-by-bin by taking the mean value of purity when fitting the candidate data with gaussian variation of signal and background templates. The uncertainty for residual plots is the bin-by-bin RMS.

APPENDIX D

f_{in} VS. f_{out}

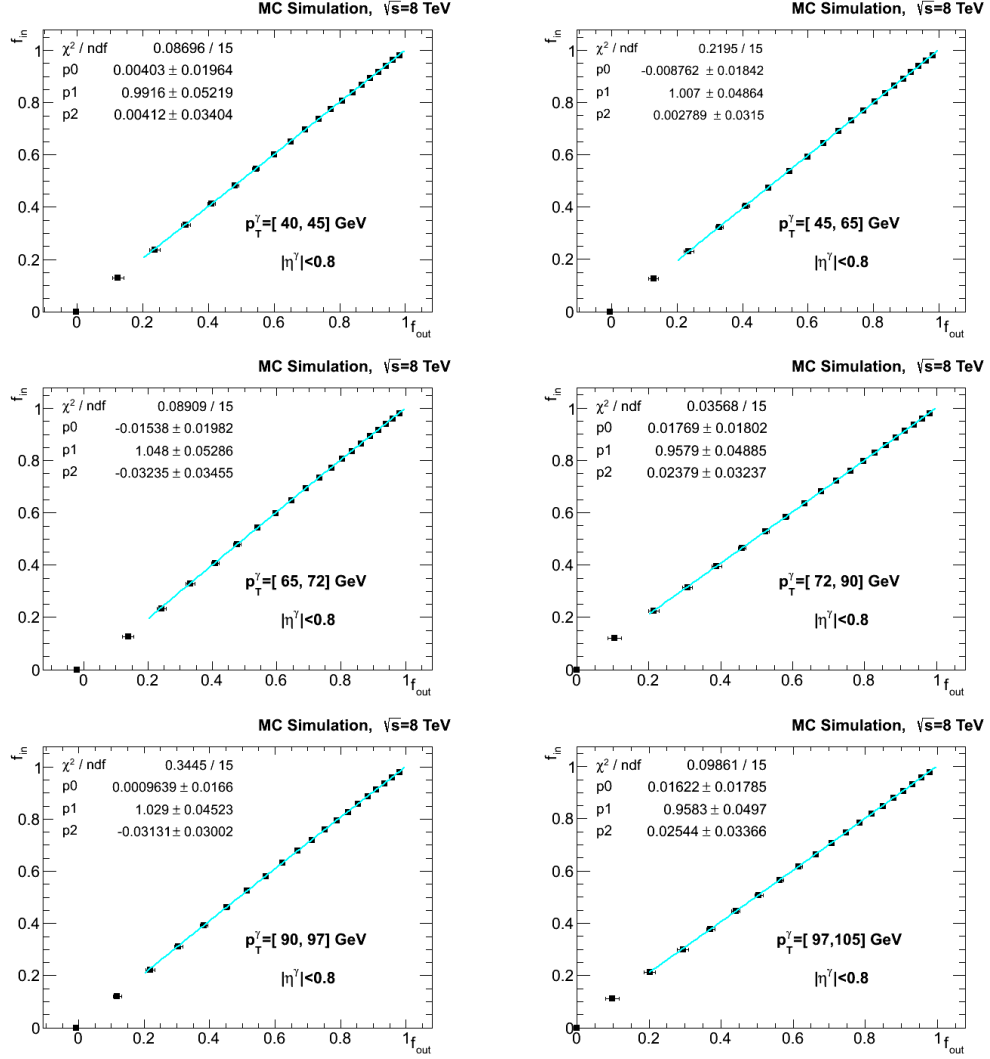


Figure D.1: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.

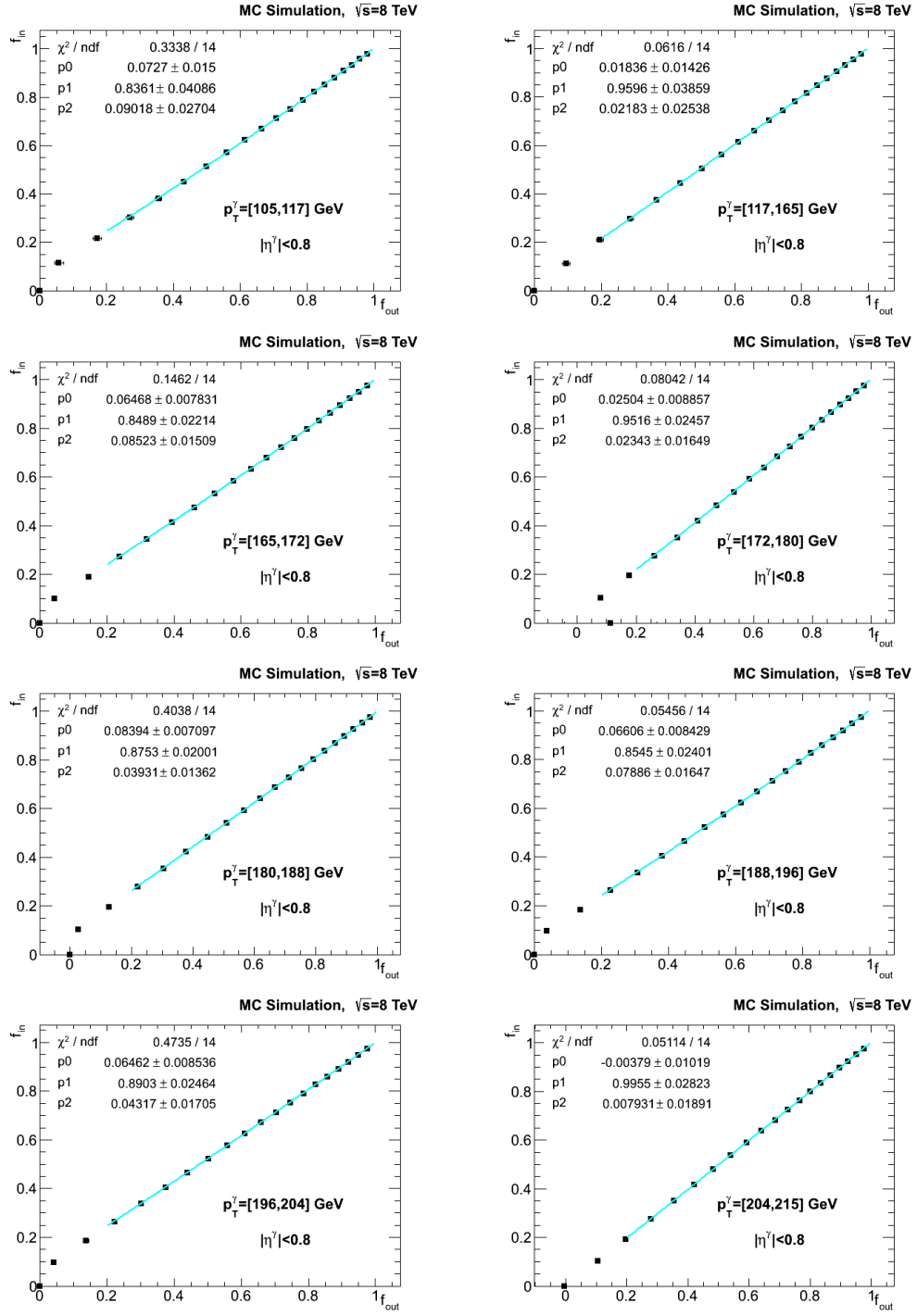


Figure D.2: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.

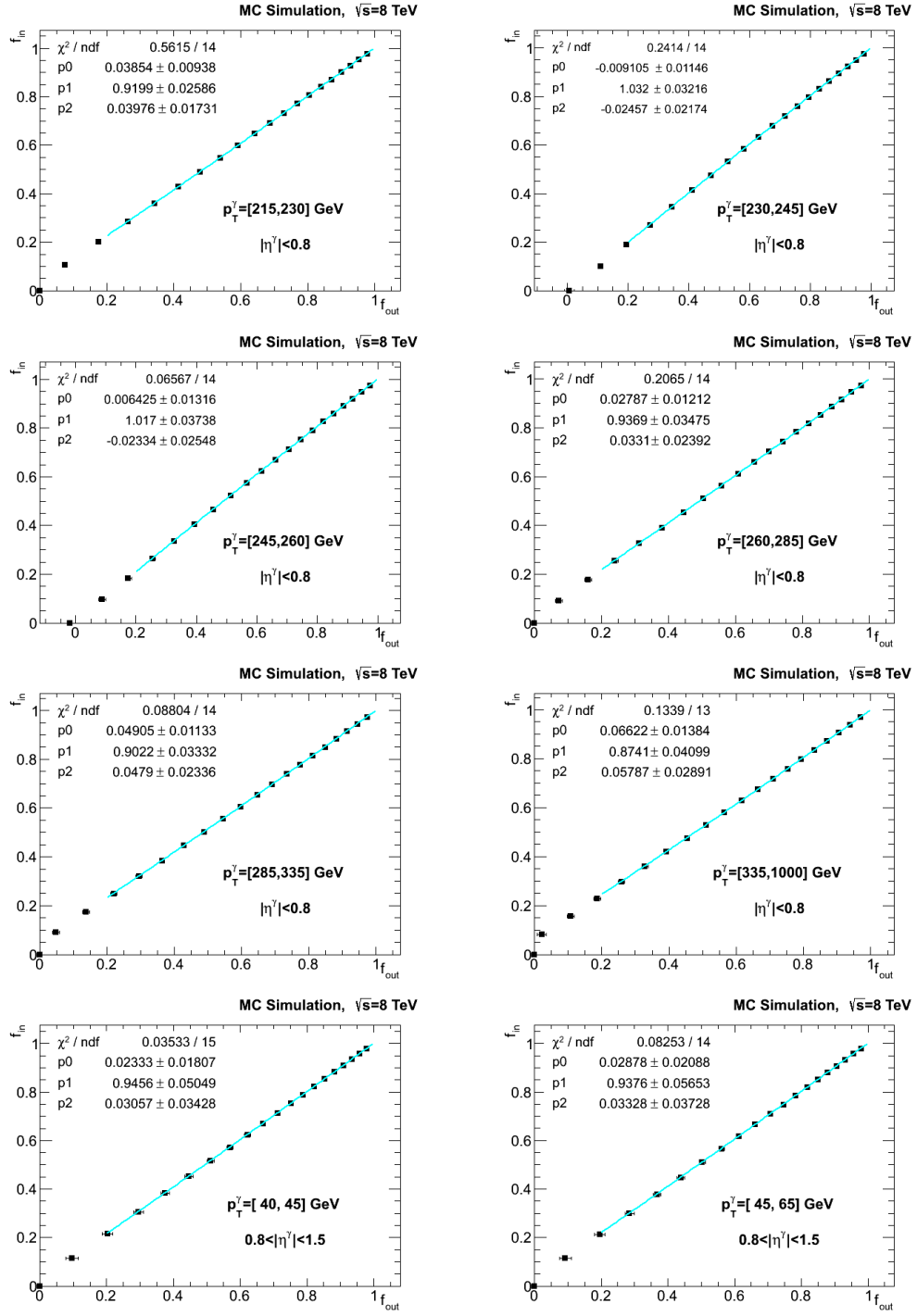


Figure D.3: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.

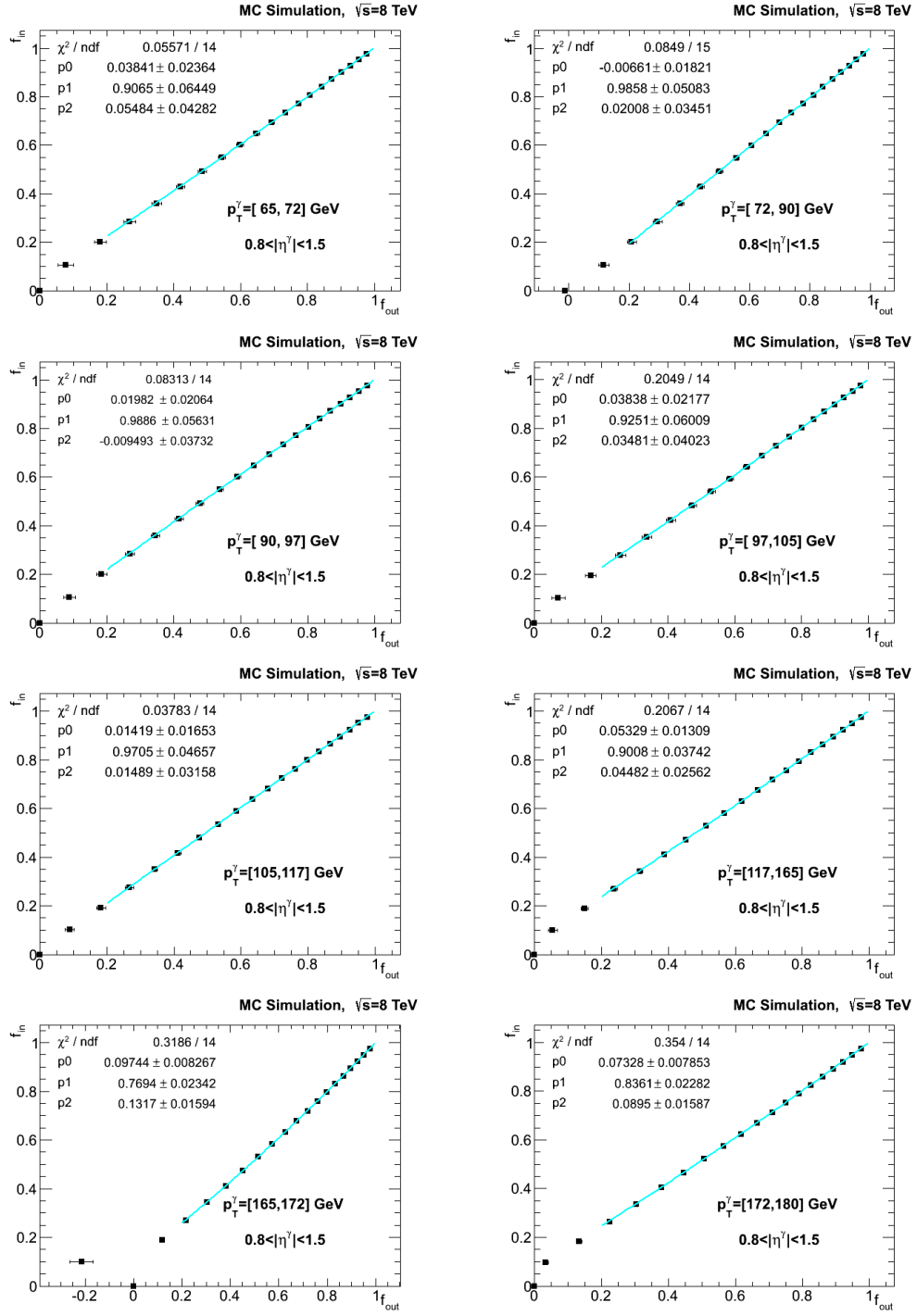


Figure D.4: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.

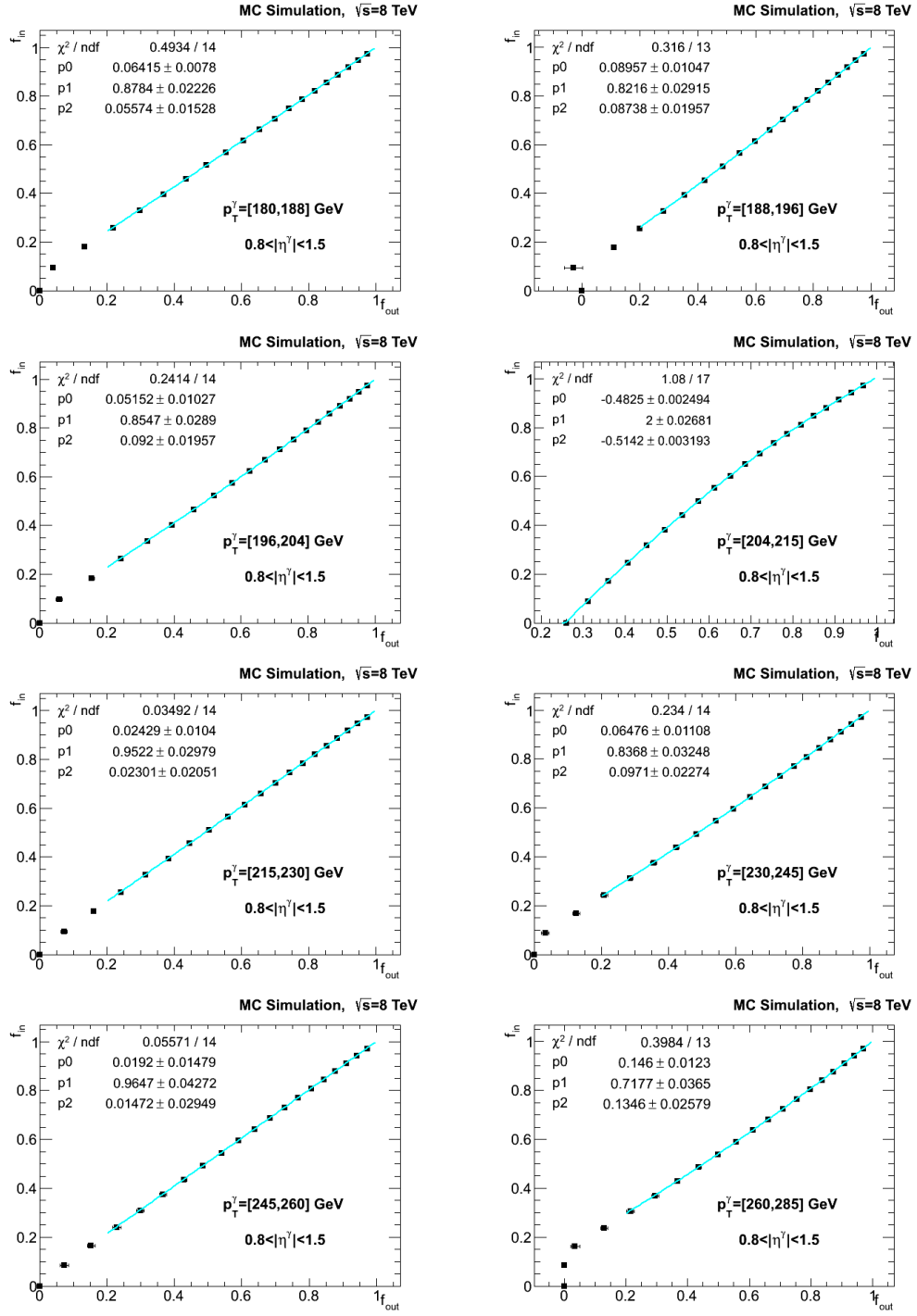


Figure D.5: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.

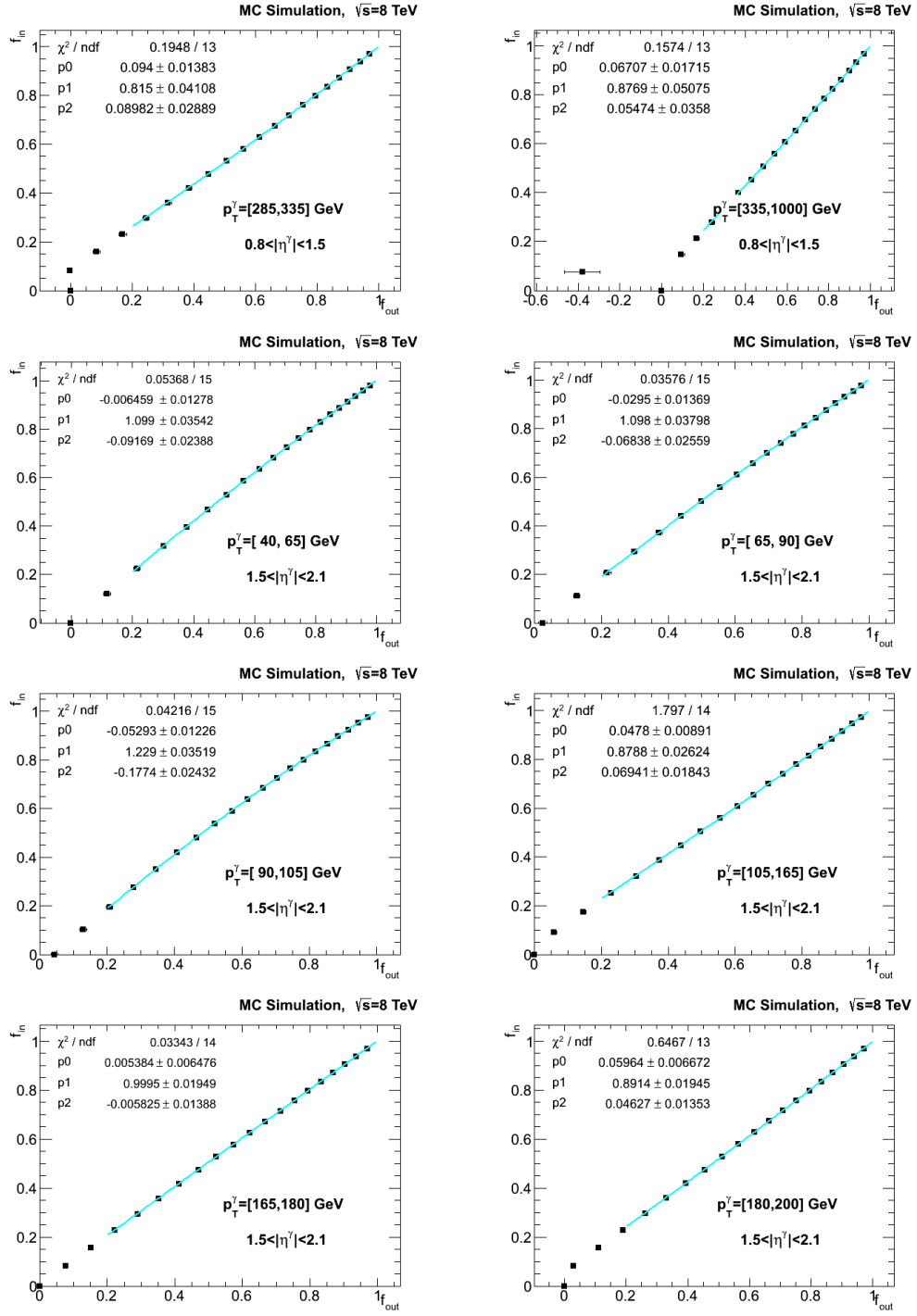


Figure D.6: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.

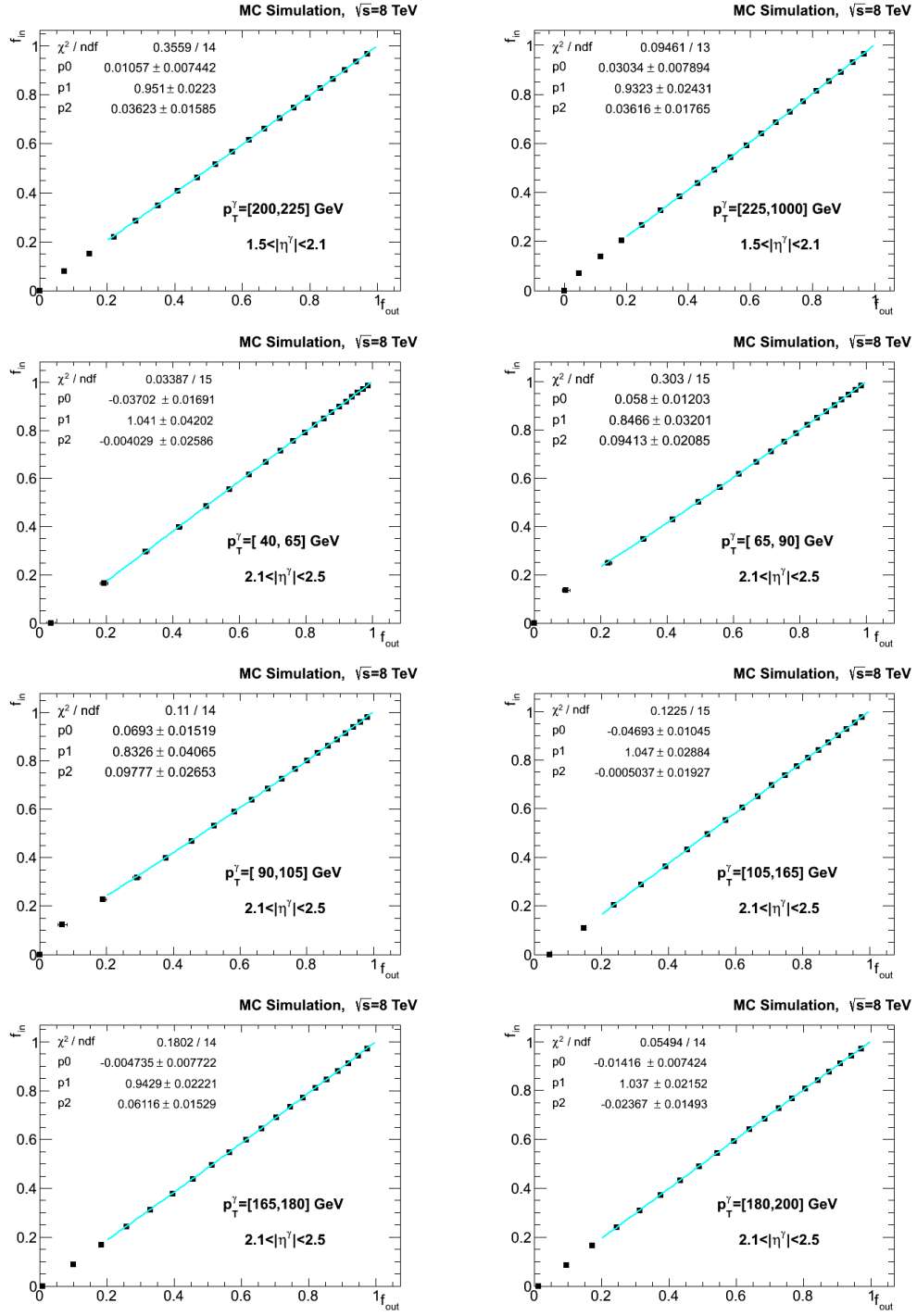


Figure D.7: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.

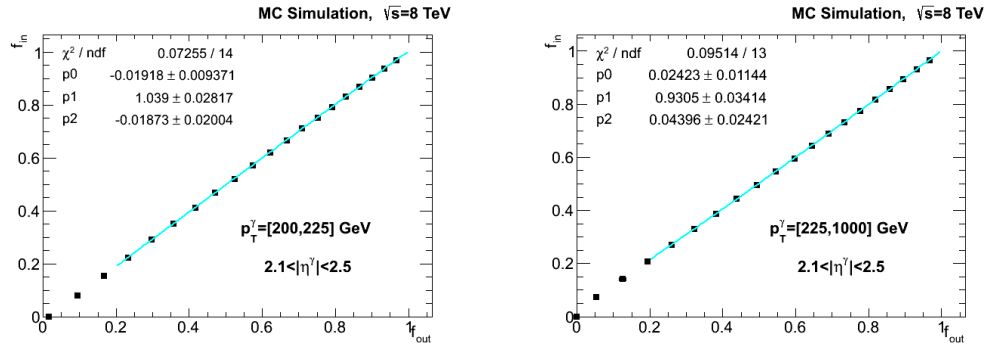


Figure D.8: MC toy experiment results to show the relationship between the input signal fraction and the output signal fraction for different bins.

BIBLIOGRAPHY

- [1] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.
- [2] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J.*, C63:189–285, 2009.
- [3] David d’Enterria and Juan Rojo. Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data. *Nuclear Physics B*, 860(3):311 – 338, 2012.
- [4] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008.
- [5] Tai Sakuma and Thomas McCauley. Detector and event visualization with sketchup at the cms experiment. *Journal of Physics: Conference Series*, 513(2):022032, 2014.
- [6] Danek Kotlinski, R Baur, R P Horisberger, R Schnyder, W Erdmann, and K Gabathuler. Readout of the CMS pixel detector. 2000.
- [7] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, 2008.
- [8] Min Suk Kim et al. CMS reconstruction improvement for the muon tracking by the RPC chambers. *PoS*, RPC2012:045, 2012. [JINST8,T03001(2013)].
- [9] S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems. 2000.
- [10] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.
- [11] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008.
- [12] A. Bettini. *Introduction to Elementary Particle Physics*. Cambridge University Press, 2008.
- [13] Georges 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.*, 114:191803, 2015.
- [14] D.H. Perkins. *Introduction to High Energy Physics*. Cambridge University Press, 2000.

- [15] George Sterman, John Smith, John C. Collins, James Whitmore, Raymond Brock, Joey Huston, Jon Pumplin, Wu-Ki Tung, Hendrik Weerts, Chien-Peng Yuan, Stephen Kuhlmann, Sanjib Mishra, Jorge G. Morfin, Fredrick Olness, Joseph Owens, Jianwei Qiu, and Davison E. Soper. Handbook of perturbative qcd. *Rev. Mod. Phys.*, 67:157–248, Jan 1995.
- [16] William B Kilgore. Introduction to quantum chromodynamics at hadron colliders. *PRAMANA - journal of physics*, 76, 5, May 2011.
- [17] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298–318, 1977.
- [18] Jet Energy Scale performance in 2011. May 2012.
- [19] CMS Collaboration. *CMS, tracker technical design report*. 1998.
- [20] O. Poeth. *The CMS Silicon Strip Tracker: Concept, Production and Commissioning*. Vieweg+Teubner research. Vieweg+Teubner Verlag, 2010.
- [21] Andrea Benaglia. The CMS ECAL performance with examples. Technical Report CMS-CR-2013-430, CERN, Geneva, Nov 2013.
- [22] CMS Collaboration. CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. 2013.
- [23] CMS Luminosity Measurement for the 2015 Data Taking Period. Technical Report CMS-PAS-LUM-15-001, CERN, Geneva, 2016.
- [24] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009, 2014.
- [25] E. Hossain, D. Niyato, and Z. Han. *Dynamic Spectrum Access and Management in Cognitive Radio Networks*. Cambridge University Press, 2009.
- [26] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.
- [27] R Frhwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Adaptive Vertex Fitting. Technical Report CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
- [28] Performance of electron reconstruction and selection with the cms detector in proton-proton collisions at $s = 8$ tev. *Journal of Instrumentation*, 10(06):P06005, 2015.
- [29] Vardan Khachatryan et al. Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10(08):P08010, 2015.

- [30] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. Fastjet user manual. *The European Physical Journal C*, 72(3):1–54, 2012.
- [31] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. 2009.
- [32] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.
- [33] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.
- [34] S. Chatrchyan, V. Khachatryan, et al. Measurement of the underlying event activity at the lhc with $\sqrt{s} = 7$ tev and comparison with $\sqrt{s} = 0.9$ tev. *Journal of High Energy Physics*, 2011(9):1–31, 2011.
- [35] Fabio Maltoni and Tim Stelzer. Madevent: automatic event generation with madgraph. *Journal of High Energy Physics*, 2003(02):027, 2003.
- [36] Pileup Jet Identification. Technical Report CMS-PAS-JME-13-005, CERN, Geneva, 2013.
- [37] Tim Adye. Unfolding algorithms and tests using RooUnfold. In *Proceedings of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland, January 2011, CERN-2011-006, pp 313-318*, pages 313–318, 2011.
- [38] G. D’Agostini. Improved iterative Bayesian unfolding. *ArXiv e-prints*, October 2010.
- [39] H. Baer, J. Ohnemus, and J.F. Owens. A calculation of the direct photon plus jet cross section in the next-to-leading-logarithm approximation. *Physics Letters B*, 234(1):127 – 131, 1990.
- [40] H. Baer, J. Ohnemus, and J. F. Owens. Next-to-leading-logarithm calculation of direct photon production. *Phys. Rev. D*, 42:61–71, Jul 1990.
- [41] B. W. Harris and J. F. Owens. The Two cutoff phase space slicing method. *Phys. Rev.*, D65:094032, 2002.
- [42] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato. Constraints on large- x parton distributions from new weak boson production and deep-inelastic scattering data. 2016.

BIOGRAPHICAL SKETCH

Ajeeta Khatiwada was born on July 20, 1987 in Kathmandu, Nepal. She attended Hermann Gmeiner School in Bhaktapur, Nepal from 1990 to 2003, and Nobel Academy in Kathmandu, Nepal from 2003 to 2005. She obtained her Bachelor of Science in Physics from Linfield College, McMinnville, Oregon in 2010. During her time as an undergraduate, she did physics research internships at Stanford Research Institute in Menlo Park, California; at Colorado Research Associates in Boulder, Colorado; and at Fermilab in Batavia, Illinois. She joined Florida State University for her graduate program in 2010. She worked on the DØ experiment at Fermilab during the initial years of her degree. In 2013, she joined the CMS experiment at CERN and began working on her Ph.D. dissertation using the CMS data to perform a standard model measurement. After the completion of her degree, she will continue working on CMS through Purdue University in West Lafayette, Indiana.